



Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

Fitzgerald, Anne M., [Pappalardo, Kylie M.](#), [Fitzgerald, Brian F.](#), [Austin, Anthony C.](#), Abbot, John W., Cosman, Brendan L., [O'Brien, Damien S.](#), & Singleton, Bill
(2007)

Building the Infrastructure for Data Access and Reuse in Collaborative Research : An Analysis of the Legal Context.

The OAK Law Project, Canberra, Australia.

This file was downloaded from: <https://eprints.qut.edu.au/215891/>

© Copyright 2007 The Authors

Licensed under an Australian Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License

License: Creative Commons: Attribution-Noncommercial-Share Alike 2.5

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

<http://www.oaklaw.qut.edu.au/>

BUILDING THE INFRASTRUCTURE FOR DATA ACCESS AND REUSE IN COLLABORATIVE RESEARCH

An Analysis of the Legal Context

Dr Anne Fitzgerald and Kylie Pappalardo

***Assisted by Professor Brian Fitzgerald, Anthony Austin, Dr John Abbot, Brendan
Cosman, Damien O'Brien, and Bill Singleton (Allens Arthur Robinson)***

Open Access to Knowledge (OAK) Law Project Legal Protocols for Copyright Management: Facilitating Open Access to Research at the National and International Levels	Legal Framework for e-Research Project
---	---

Funded by the Australian Government Department of Education Science
and Training (DEST)

Systemic Infrastructure Initiative (SII) funded projects and part of the
Commonwealth Government's *Backing Australia's Ability – An
Innovation Action Plan for the Future*

June 2007



This work is licensed under an Australian Creative Commons Attribution-NonCommercial-
ShareAlike 2.5 License

<http://creativecommons.org/licenses/by-nc-sa/2.5/au>

For information:

Professor Brian Fitzgerald
Project Leader
The OAK Law Project
Queensland University of Technology
GPO Box 2434 Brisbane Queensland 4001

The authors would like to extend their sincere gratitude to the following people for their valuable contribution in preparing this report for publication:

Libby Austen (cover design), Celeste Bennett, Elliot Bledsoe, Tanya Butkovsky, Jodi Proctor, Alex Cowley and the team at Elect Printing.

This work is licensed under an Australian Creative Commons Attribution Non-Commercial ShareAlike 2.5 Licence. To view a copy of this licence, visit <http://creativecommons.org/licences/by-nc-sa/2.5/au> or send a letter to creative Commons Australia, c/-Law faculty, Queensland University of technology, GPO Box 2434, Brisbane, Qld 4001, Australia.

ISBN 978 0 9802988 2 6

Printed by Elect printing, Canberra, Australia.

Preface

This Report has been undertaken as an initiative of the DEST funded Open Access to Knowledge (OAK) Law Project and the DEST funded Legal Framework for e-Research Project.

It aims to provide a framework for understanding how we can build legal and management infrastructure to support access to and reuse of research data within the Australian research system.

The Report was commenced in October 2006 as a response to the growing interest worldwide in data sharing as a driver of innovation and the emergence of the National Collaborative Research Infrastructure Strategy (NCRIS). It has expanded in its scope along the way in order to accommodate policy developments both in Australia and overseas.

The Report overviews and examines the fundamental legal issues surrounding data sharing in the context of case studies based on prominent data sharing networks. Most importantly, it also provides a strategy for further work in this area.

I am thankful to Dr Anne Fitzgerald and Ms Kylie Pappalardo for the enormous amount of work they have undertaken and to the team of people (Dr Abbot and Messrs Austin, Cosman, O'Brien and Singleton) who have worked with the project at various times over the last 6 months. They have all made a very worthwhile and productive contribution. I would also like to acknowledge the valuable support of QUT DVC Professor Tom Cochrane, DEST Director of Education, Innovation and Infrastructure Policy Ms Margot Bell and DEST Assistant Director of Education, Innovation and Infrastructure Policy Ms Clare McLaughlin, OAK Law Project Manager and Legal Framework for e-Research Project Manager Mr Scott Kiel-Chisholm, QUT research colleagues Ms Jessica Coates, Dr Amanda McBratney, Mr Nic Suzor and Ms Maree Heffernan, Queensland Government Senior Lawyer Mr Neale Hooper and Mr Peter James, Partner of Allens Arthur Robinson (who provided the opportunity for us to involve Mr Bill Singleton in the Project).

We hope you enjoy reading this Report and we encourage your feedback and insights.

Brian Fitzgerald

**Professor of Intellectual Property and Innovation (QUT) <www.ip.qut.edu.au>
Project Leader for OAK Law and Legal Framework for e-Research**

June 2007

Executive Summary

This Report examines the broad legal framework within which research data is generated, managed, disseminated and used. The background to the Report is the growing support for systems that enable research data generated in publicly-funded research projects to be made available for access and use by others in the research community.

The Report provides an overview of the operation of copyright law, contract and confidentiality laws, as well as a range of legislation - privacy, public records and freedom of information legislation, etc – that is of relevance to research data. The Report considers how these legal rules apply to define rights in research data and regulate the generation, management and sharing of data. In any given research project there will be a multitude of different parties with varying interests – legal and otherwise – in the data produced. These parties include researchers, research funders, licensees and other users, for example members of the general public who access the data online. The Report examines the relationships between these parties and the legal arrangements that must be implemented to ensure that research data is properly and effectively managed, so that it can be accessed and used by other researchers.

Important in the context of collaborative research and open access, the Report describes and explains current practices and attitudes towards data sharing. A wide array of databases is analysed to ascertain the arrangements currently in place to manage and provide access to research data. Often these practices are informed by international and national policies on access and use, formulated by international organisations and conferences, research funders and research bodies. The Report considers these policies at length and canvasses the development of the open access to research data movement.

Finally, the Report encourages researchers and research organisations to adopt proper management and legal frameworks for research data outputs. It provides practical guidance on the development and implementation of legal frameworks for data management with the objective of ensuring that research data can be accessed and used by other researchers. The Report describes best practice strategies and mechanisms for organising, preserving and enabling access to and reuse of research data, including data management policies and principles, data management plans and data management toolkits. Proposals are made for further work to be undertaken on data access policies, frameworks, strategies and mechanisms.

Aim of this Report

This Report aims to provide practical guidance for individual researchers, research institutions and funding bodies in developing and implementing legal frameworks for the management of data generated in research projects. The adoption of appropriate legal protocols will enhance the management of research data and ensure that data outputs can more readily be accessed and used by others in the research community. Legal frameworks for data management should cover all stages of a research project, from the collection or generation of research data through to its management, dissemination and use.

This Report examines and considers:

- current themes and developments at national and international levels relating to access to and reuse of research data, particularly data generated in publicly funded research projects;
- how key concepts and terminology relating to data and data management are understood in the research sector;
- the different roles played, and the level of control exercised by, persons with interests in or rights in relation to research data;
- the legislative and administrative instruments that regulate data storage, public availability (in the case of government data) and privacy requirements (in the case of health and genetic data and personally identifying information);
- key examples of database practices and arrangements for data access and sharing in Australia and overseas;
- attitudes of researchers towards data access and sharing, as measured through surveys conducted by QUT and other research bodies;
- the application of copyright law (in Australia and key overseas jurisdictions) and open content licences (such as Creative Commons licences and Science Commons licences) to data and databases;
- the imposition and enforcement of confidentiality obligations on data recipients and the use of contractual arrangements to control and manage data (such as copyright assignment, licensing and general contracts); and
- Australian and international policies and principles on open access to research data.

Based on this examination of issues, the Report outlines practical steps that can be taken by researchers and research institutions to manage their research data so that it can be made available for access and use. The Report describes best practice strategies and mechanisms for organising, preserving and enabling access to and reuse of research data, including data management policies and principles, data management plans and data management toolkits. Proposals are made for further work to be undertaken on data access policies, frameworks, strategies and mechanisms.

CHAPTER 1 - THE DATA LANDSCAPE	1
1. "Data is the Next Intel Inside"	3
2. The Economic Benefits of Data Sharing.....	5
3. New Research Methodologies	6
4. e-Research and Cyberinfrastructure	7
5. Data Management as Infrastructure.....	8
6. Legal Framework as e-Research Infrastructure	9
7. Guidelines.....	10
8. Purpose of the Report	12
CHAPTER 2 - KEY CONCEPTS.....	17
1. Introduction.....	19
2. "Data" and "information"	19
Figure 1: Individual data items	21
3. "Datasets"	21
Figure 2: Dataset.....	21
4. "Database" (or "data archive" or "repository").....	22
Figure 3: Database.....	23
5. "e-Research"	23
6. "Ownership" and "Control"	24
7. "Access" and "Use"	26
KEY POINTS.....	28
CHAPTER 3 - THE REGULATORY CONTEXT OF DATA	29
1. Statutory Frameworks Governing Data	31
2. Health Information and Personally Identifying Information – Legal Frameworks for Information Privacy	33
De-identifying Personal Information	36
Responses to Privacy Act 1988 (Cth).....	37
Data Linkage	38
Recommendations	40
Draft National Health Privacy Code.....	40
National Health Act 1953 (Cth)	40
State Legislation	41
State and Territory Privacy Legislation:	41
New South Wales.....	41
Victoria	41
Australian Capital Territory	42
Northern Territory	42
Tasmania	42
South Australia	43
Western Australia	43
Genetic Health Information.....	44
3. Government Data and Information – Archiving, Accountability and Accessibility.....	45
Public Records (or Archives) Legislation and Freedom of Information Legislation	45
Electronic Transactions Legislation.....	47
Other Legislation	48
4. Data Quality and Control - Administrative Arrangements and Information Standards	49

5. Protocols Relating to Culturally Sensitive Data	50
6. Understanding the Legal Framework – Practical Application to Data Management	51
KEY POINTS.....	53

CHAPTER 4 - CURRENT PRACTICES AND ATTITUDES TO DATA

SHARING55

1. The Growth of Domestic and International Database Sharing Practices	57
2. Key Examples of Data Sharing Frameworks	59
Example 1 - The National Institutes of Health (NIH) Databases	59
Example 2 - NIH: GenBank	60
Ownership - GenBank	60
Control - GenBank	60
Access and Use - GenBank	60
Example 3 - NIH: Genetic Association Information Network (GAIN)	61
Ownership - GAIN	62
Control - GAIN	62
Access and Use - GAIN	63
Example 4 - NIH - National Human Genome Research Institute (NHGRI)	63
Ownership – NHGRI	63
Control – NHGRI	64
Access and Use – NHGRI	64
Example 5 - NHGRI: International Human Genome Sequencing Consortium (IHGSC).....	66
Ownership and Control – IHGSC.....	66
Access and Use – IHGSC.....	66
Example 6 - NHGRI: Mouse Genome Sequencing Consortium (MGSC) and the Jackson Laboratory Mouse Genome Informatics Database....	66
Ownership – MGI	67
Control – MGI.....	68
Access and Use – MGI.....	68
Example 7 - NHGRI: Rat Genome Sequencing Consortium (RGSC) ..	68
Ownership - RGSC	68
Control – RGSC.....	69
Access and Use – RGSC.....	69
Example 8 - NHGRI: The SNP Consortium (TSC)	69
Ownership -TSC	69
Control – TSC	70
Access and Use - TSC	70
Example 9 - NHGRI - Mammalian Gene Collection (MGC)	70
Ownership – MGC.....	71
Control – MGC.....	72
Access and Use – MGC.....	72
Example 10- NHGRI - Encyclopaedia of DNA Elements (ENCODE)....	72
Ownership – ENCODE	72
Control – ENCODE	73
Access and Use – ENCODE	73
Example 11 - The National Centre for Biotechnology Information (NCBI) Databases	74

Example 12 - NCBI: Database of Genotype and Phenotype (dbGaP)	74
Ownership - dbGaP	76
Control - dbGaP	76
Access and Use - dbGaP	76
Example 13 - NCBI: Entrez	78
Ownership - Entrez	78
Control - Entrez	78
Access and Use - Entrez	79
Example 14 - The California Stem Cell Initiative (CIRM)	79
Ownership - CIRM	80
Control - CIRM	80
Access and Use - CIRM	80
Example 15 - Malaria Genomic Epidemiology Network (MalariaGEN)	81
Ownership - MalariaGEN	81
Control - MalariaGEN	82
Access and Use - MalariaGEN	82
Example 16 - UniProtKB/Swiss-Prot Protein Knowledgebase	83
Ownership - UniProtKB	83
Control - UniProtKB	84
Access and Use - UniProtKB	84
Example 17 - BRENDA	85
Ownership - BRENDA	85
Control - BRENDA	86
Access and Use - BRENDA	86
Example 18 - Nanowerk	87
Ownership - Nanowerk	87
Control - Nanowerk	87
Access and Use - Nanowerk	87
Example 19 - Mars Climate Database (MCD)	88
Ownership - MCD	88
Control - MCD	89
Access and Use - MCD	89
Example 20 - Western Australian Genetic Epidemiology Resource (WAGER)	90
Ownership - WAGER	90
Control - WAGER	91
Access and Use - WAGER	91
Example 21 - The RettBase and the InterRett Databases	92
Ownership - IRSA	92
Control - IRSA	92
Access and Use - IRSA	93
Example 22 - Earth System Grid (ESG)	93
Ownership - ESG	93
Control - ESG	94
Access and Use - ESG	94
Example 23 - Australian Social Science Data Archive (ASSDA)	94
Ownership - ASSDA	95
Control - ASSDA	95
Access and Use - ASSDA	95
Example 24 - Marine Themes (MT)	96

Ownership – MT	96
Control – MT	97
Access and Use – MT	98
Example 25 - Australian Ocean Data Centre Joint Facility (AODCJF) and the Australian Ocean Data Centre (AODC)	98
Ownership –AODC.....	98
Control – AODC.....	99
Access and Use - AODC.....	100
Example 26 - The Australian Digital Thesis Program.....	100
Ownership – ADT	100
Control – ADT	102
Access and Use – ADT	102
Example 27 - Capture Wales Digital Storytelling Database - BBC.Co.UK	102
Ownership – Capture Wales.....	103
Control – Capture Wales.....	103
Access and Use – Capture Wales.....	103
Example 28 - Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC).....	104
Ownership – PARADISEC	104
Control – PARADISEC	105
Access and Use - PARADISEC	106
Example 29 - Sloan Digital Sky Survey (SDSS)	106
Ownership – SDSS	107
Control – SDSS	108
Access and Use - SDSS	109
Example 30 – The University of Melbourne – Experimental Particle Physics (EPP) – Collaboration with the ATLAS Experiment.....	109
Ownership – EPP	110
Control – EPP	111
Access and Use – EPP	111
Example 31 – Household Income and Labour Dynamics in Australia Survey (HILDA).....	111
Ownership – HILDA	111
Control – HILDA	112
Access and Use – HILDA	113
3. Observations on Data Sharing Practices	114
Clarity and Understanding	114
Consistency of Language	115
4. Patents and Open Licensing	116
What Are Patents?.....	116
What Are The Concerns Relating To Patents?	117
Researchers Intending To Patent.....	117
Researchers Not Intending To Patent.....	118
The Bermuda Principles Approach – Releasing Data to Pre-empt Patents	118
The HapMap Approach – Conditional Access Licensing.....	120
The CAMBIA Approach – Open Licensing.....	121
5. Researchers' Data Sharing Attitudes and Practices.....	124
Australian eResearch Sustainability Survey	126

NCRIS Platforms for Collaboration Data Management Survey	127
QUT e-Research survey	129
KEY POINTS.....	132
CHAPTER 5 - COPYRIGHT.....	133
1. Introduction.....	135
2. Materials Protected by Copyright.....	135
3. Requirement of Material Form	136
4. Copyright in Data and Databases.....	136
5. Duration of Copyright.....	138
6. Originality.....	138
7. Ownership	139
8. Assignment.....	140
9. Ownership in Computer-Generated Works.....	140
10. Rights	141
11. Infringement	142
12. Digital Rights Management (DRM) and Electronic Rights Management Information (ERMI)	143
13. Licences.....	144
Licences and Databases.....	145
Open Content Licensing.....	146
Creative Commons	146
Creative Commons - Baseline Features:	147
Creative Commons - Optional Features:.....	147
Science Commons	150
Statutory Licences.....	151
14. Other Jurisdictions	152
United States.....	152
United Kingdom	153
Copyright in Databases – United Kingdom	153
The European Database Right.....	155
Diagram: Categorisation of licences	157
KEY POINTS.....	158
CHAPTER 6 - CONFIDENTIAL INFORMATION	161
1. Introduction.....	163
2. Secret or Confidential Information	163
3. The Action for Breach of Confidence	164
Information Must be of a Confidential Nature	164
Obligation of Confidence.....	165
Unauthorised Use or Disclosure	165
Defences.....	166
Remedies.....	166
4. Information in the Employment Context	167
5. Government Information	168
6. Information May Lose Its Secrecy	169
7. Contracts	169
KEY POINTS.....	170
CHAPTER 7 - CONTRACT	171
1. Introduction.....	173
2. On-Line Contracts (“Click Wrap Agreements”).....	174
3. Contracts to Protect and Control Use of Information.....	174

Confidentiality Agreements.....	174
Contracts Controlling Access To and Use of Information.....	176
4. Contracts and Copyright	177
Assignment.....	177
Licence.....	177
5. Diagrammatic Examples	177
Example 1: Basic Confidentiality Agreement for Use of Research Data	178
Example 2: Internal Confidentiality Agreement	179
Example 3: Confidentiality Agreement – Conditional User Access.....	181
Example 4: Access Content Licence - Owner and User.....	183
Example 5: Open Access Content Sub-Licence - Database and User	185
Example 6: Open Access Content Licence – Commercial Restrictions	187
Example 8: dbGaP – Data Transfer Agreement – ‘Controlled Access to Data Licence’	191
KEY POINTS.....	193
CHAPTER 8 - DATA SHARING FRAMEWORKS	195
1. Introduction.....	197
2. International Organisations.....	198
Bermuda Principles (1996).....	198
Budapest Open Access Initiative (2002)	200
Bethesda Statement on Open Access Publishing (2003).....	200
Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003)	200
World Summit on the Information Society (WSIS) Declaration of Principles (2003).....	201
Organisation for Economic Co-operation and Development (OECD) Declaration on Access to Research Data from Public Funding (2004)	202
OECD Working Party on the Information Economy (31 May 2006)....	204
International Council for Science (ICSU) Scientific Data and Information Report (2004).....	205
3. Governments and Public Sector Research Funding Bodies	206
United States.....	207
NIH Data Sharing Policy (2003).....	207
Canada.....	209
Canadian Institutes of Health Research (CIHR)	209
GeoConnections Canada.....	210
The Dissemination of Government Geographic Data in Canada: Guide to Best Practices (2005)	210
CGDI Roadmap	212
GeoConnections Policy Advisory Node	213
United Kingdom and Europe	213
Research Councils UK	214
Research Information Network (RIN).....	214
European Research Council.....	215
European Union’s Directive on the Re-use of Public Sector Information (EU Directive).....	216

Australia	219
Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC)	219
Office of Spatial Data Management (OSDM) Spatial Data Access and Pricing Policy (2001)	221
OSDM Data Custodianship Guidelines.....	223
ANZLIC Model Data Access and Management Agreement.....	225
Australian eResearch Sustainability Survey	226
Australian Government Information Interoperability Framework	228
Productivity Commission Review of Access to and Pricing of Government Information 2002	229
Prime Minister's Science, Engineering and Innovation Council (PMSEIC) Report (2006).....	229
National Collaborative Research Infrastructure Strategy (NCRIS) - Platforms for Collaboration	230
4. Private Sector Organisations.....	231
Wellcome Trust Position Statements (2003)	231
Novartis.....	233
University and Industry Innovation Summit (2005): Open Collaboration Principles.....	234
KEY POINTS.....	236
CHAPTER 9 - DATA SHARING INFRASTRUCTURE.....	237
1. Introduction.....	239
2. Formulate a Data Access Policy and Principles.....	239
3. Identify Data to be Made Available for Access and Use	243
4. Ascertaining the Conditions of Access and Use	244
5. Creative Commons Licenses	246
6. Adopting Mechanisms for Practical Data Management.....	246
7. Data Management Plans.....	247
How Data is Generated or Collected.....	249
How Data is Integrated and Stored in the Database.....	249
Data Ownership.....	250
How Data will be Managed	250
Regulatory Mechanisms and Controls.....	251
Quality Assurance and Control Measures.....	252
Data Security Measures.....	252
Roles and Responsibilities of Researchers and Database Managers.....	254
Database Infrastructure.....	254
Sustainability of Data	255
Expenditures	255
8. Data Management Toolkit.....	256
KEY POINTS.....	258
CHAPTER 10 - CONCLUSIONS AND RECOMMENDATIONS.....	261
1. Conclusions.....	263
2. Key Points of the Report	264
3. Proposed Actions	264
Proposal One – Analysis of Data Access and Sharing Practices.....	265
Proposal Two – Analysis of Open Access Policies and Principles	266
Proposal Three – Development of Data Access Policies and Principles, and Model Data Management Plans, Toolkits and Templates	266
Index	269

Acronyms and Abbreviations

ACT	Australian Capital Territory
ADT	Australian Digital Thesis Program
AERES	Australian eResearch Sustainability Survey
AHMAC	Australian Health Minister's Advisory Council
ALRC	Australian Law Reform Commission
ANU	Australian National University
ANZLIC	The Spatial Information Council of Australia and New Zealand (formerly known as the Australia New Zealand Land Information Council)
AODC	Australian Ocean Data Centre
AODCJF	Australian Ocean Data Centre Joint Facility
APSR	Australian Partnership for Sustainable Repositories
ARC	Australian Research Council
AREDS	Age-Related Eye Diseases Study
ASSDA	Australian Social Science Data Archive
ASTRORC	Astrophysical Research Consortium
BBC	British Broadcasting Corporation
BCM-HGSC	Baylor College of Medicine Human Genome Sequencing Center
BiOS	Biological Open Source
BOAI	Budapest Open Access Initiative
CAL	Copyright Agency Limited
CAUL	Council of Australian University Librarians
CERN	European Organization for Nuclear Research (<i>Conseil Européen pour la Recherche Nucléaire</i>)
CGDI	Canadian Geospatial Data Infrastructure
CIHR	Canadian Institutes of Health Research
CIRM	California Stem Cell Research and Cures Initiative
CLRC	Copyright Law Review Committee
CSDMG	Commonwealth Spatial Data Management Group
CSIRO	Australian Commonwealth Scientific and Research Organisation
Cth	Commonwealth
DAC	Data Access Committee
dbGaP	Database of Genotype and Phenotype
DEST	Department of Education, Science and Training
DLU	Data Linkage Unit
DMP	Data Management Plan
DMT	Data Management Toolkit
DNA	Deoxyribonucleic acid
DRM	Digital rights management
DUC	Data Use Certificate
EBI	European Bioinformatics Institute
ECJ	European Court of Justice
EMRI	Electronic rights management information
ENCODE	Encyclopaedia of DNA Elements
EPP	Experimental Particle Physics
ESG	Earth System Grid

EU	European Union
FAQ	Frequently Asked Questions
FTP	File Transfer Protocol
GAIN	Genetic Association Information Network
GPL	General Public Licence
HapMap	International Haplotype Project
HILDA	Household Income and Labour Dynamics in Australia
ICSU	International Council for Science
IHGSC	International Human Genome Sequencing Consortium
I.M.A.G.E	Integrated Molecular Analysis of Genomes and their Expression
INSD	International Nucleotide Sequence Database Collaboration
IP	Intellectual Property
IPPNO	CIRM's Intellectual Property Policy for Non-Profit Organizations
IRSA	International Rett Syndrome Association
JISC	Joint Information Systems Committee
MalariaGen	Malaria Genomic Epidemiology Network
MCD	Mars Climate Database
MGC	Mammalian Gene Collection
MGI	Mouse Genome Informatics
MGSC	Mouse Genome Sequencing Consortium
MIT	Massachusetts Institute of Technology
MRC	Medical Research Council
MSP	Medical Sequencing Program
MT	Marine Themes
NCBI	National Center for Biotechnology Information
NCRIS	National Collaborative Research Infrastructure Strategy
NHGRI	National Human Genome Research Institute
NHMRC	National Health and Medical Research Council
NIH	National Institutes of Health
NINDS	National Institute of Neurological Disorders and Stroke
NPP	National Privacy Principle
NSF	National Science Foundation
NSIF	National Service Improvement Framework
NSW	New South Wales
OAK	Open Access to Knowledge
OECD	Organisation for Economic Cooperation and Development
OPC	Office of the Privacy Commissioner
OSDM	Office of Spatial Data Management
PARADISEC	Pacific And Regional Archive for Digital Sources in Endangered Cultures
PMSEIC	Prime Minister's Science, Engineering and Innovation Council
PSI	Public Sector Institutions
QLD	Queensland
QUT	Queensland University of Technology
R & D	Research and Development
RDF	Resource Description Framework
RIN	Research Information Network
RGSC	Rat Genome Sequencing Consortium

RS	Rett syndrome
SA	South Australia
SDSS	Sloan Digital Sky Survey
SNP	Single Nucleotide Polymorphism ¹
Tas	Tasmania
TSC	The SNP Consortium
UK	United Kingdom
US	United States
Vic	Victoria
WA	Western Australia
WADLS	Western Australian Data Linkage System
WAGER	Western Australian Genetic Epidemiology Resource
WSIS	World Summit of the Information Society

¹ Single point mutations in the human genome

CHAPTER 1 - THE DATA LANDSCAPE

“Today’s research community must assume responsibility for building a robust data and information infrastructure for the future.”²

² ICSU (International Council for Science), *Scientific Data and Information: A Report of the CSPR Assessment Panel*, December 2004 at p 7

BACKGROUND: THE DATA LANDSCAPE

Aims:

1. Explain the background to data access and reuse, with a particular focus on recent developments in the Australian research sector;
2. Introduce key umbrella terms such as “infrastructure”, cyberinfrastructure” and “e-Research”;
3. Explain the relevance of data management and legal frameworks and protocols to e-Research infrastructure; and
4. Provide an overview of the scope of this Report and summarise the issues discussed in each Chapter.

1. “Data is the Next Intel Inside”³

1.01 A focus of attention in recent years within the research community has been the development of systems to facilitate access to research data and promote the sharing of research outputs. There has been a revolution in the way research data is produced, stored, analysed and disseminated. Now, vast amounts of digital data generated through research, observational projects and instruments can be accessed through distributed networks online. Developments in information and communication technologies have made it possible to carry out research and address complex problems in ways that were not previously possible.⁴

1.02 Initiatives are underway in Australia and elsewhere to develop systems to facilitate access to research outputs in the form of data and publications. In May 2004, the Australian Prime Minister announced that the Federal Government was establishing quality and accessibility frameworks for the results of publicly funded research. The Accessibility Framework for Publicly Funded Research is designed to manage research information, outputs and infrastructure in order to enable them to be more readily discovered, accessed and shared.⁵ It aims to provide a regulatory environment that both enables and encourages the population of digital repositories in order to provide better access to information.⁶

³ Tim O'Reilly, 'What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software', 30 September 2005, at <<http://www.oreillynet.com/lpt/a/6228>> accessed on 17 May 2007

⁴ See, for example, the Alliance for Cellular Signaling's (AfCS) cell biology research project, at <http://www.signaling-gateway.org/>

⁵ See Australian Government, Department of Education, Science and Training (DEST), *Accessibility Framework*,

<http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/accessibility_framework/default.htm> at 16 January 2007

⁶ The projects funded to date under the Accessibility Framework include: (1) Australian Research Repositories Online to the World (ARROW) – led by Monash University, this project aims to identify and test software solutions to best support institutional digital repositories, see

1.03 In December 2006, the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) announced the introduction of open access guidelines for published papers and data resulting from funded research projects, effective 2008.⁷ Both policies encourage researchers to:

Consider the benefits of depositing their data and any publications arising from a research project in an appropriate subject and/or institutional repository [because in order to] maximise the benefits from research, findings need to be disseminated as broadly as possible to allow access by other researchers and the wider community.⁸

1.04 The Productivity Commission, in its 2007 report, *Public Funding for Science and Innovation*,⁹ also supported the introduction of requirements for open access to papers and data resulting from funded research projects. The Commission commended the steps taken by the ARC and NHMRC to promote open access to the results of the projects they fund but considered that, in light of experience in the United States, voluntary compliance was likely to be low. Consequently, the Productivity Commission considered that the aim of free and open access to publicly-funded research results would be better achieved by the progressive introduction of mandatory open access requirements.¹⁰

1.05 While there is now a heightened interest in enabling access to and reuse of research data, the benefits flowing from the sharing of research data among individual investigators and research groups have long been recognised.¹¹ In 1985, the United States Committee on National Statistics stated that the sharing of data reinforces open scientific inquiry and encourages a diversity of analyses, including analysis to verify or refute reported results, refine research results and check if the results are subject to

<http://www.arrow.edu.au/>; (2) Meta Access Management System (MAMS) – led by Macquarie University, MAMS helps to develop technical services (metadata searching) to enhance research dissemination, see <<http://www.melcoe.mq.edu.au/projects/MAMS>>; (3) Australian Partnership for Sustainable Repositories (APSR) – led by the Australia National University (ANU), this establishes a centre of excellence for the management of digital collections, see <<http://www.apsr.edu.au>>; and (4) Australian Digital Thesis Program Expansion and Redevelopment (ADT) – led by the University of New South Wales, this project creates a national collaborative database of digital theses, see <<http://www.anu.edu.au/aul/adtdt2006-2009businessplan.doc>>

⁷ Australian Research Council, *Discovery Projects Funding Rules for funding commencing in 2008* <http://www.arc.gov.au/pdf/DP08_FundingRules.pdf> National Health and Medical Research Council, *Project Grants Funding Policy for grants commencing in 2008*

<http://www.nhmrc.gov.au/publications/_files/profundingpol.pdf> See also the ARC's response to the Productivity Council's draft research report on Public Support for Science and Innovation (2006), recommending that consideration be given to the funding of institutional open access repositories: Australian Research Council, *Response to the Productivity Commission Draft Research Report – Public Support for Science and Innovation* (2006)

<http://www.arc.gov.au/pdf/response_PCdraftresearchreport_06.pdf>.

⁸ Australian Research Council, *Discovery Projects Funding Rules for funding commencing in 2008*, [1.4.5.1] <http://www.arc.gov.au/pdf/DP08_FundingRules.pdf>; National Health and Medical Research Council, *Project Grants Funding Policy for grants commencing in 2008*, [16.2]. <http://www.nhmrc.gov.au/publications/_files/profundingpol.pdf>.

⁹ Productivity Commission, *Public Support for Research and Innovation*, Research Report (2007) 240, 243 <<http://www.pc.gov.au/study/science/finalreport/index.html>> at 3 April 2007

¹⁰ Ibid.

¹¹ T Dedeurwaerdere, 'The Institutional Economics of Sharing Biological Information' (Paper presented at the 7th International Bioecon Conference, Cambridge, 20-21 September 2005); Responsible Conduct in Data Management, *Data Ownership* <http://ori.dhhs.gov/education/products/n_illinois_u/datamanagement/dotopic.html>.

varying assumptions.¹² The US National Institutes of Health (NIH) takes the view that data sharing benefits not only investigators and the scientific community but also funding agencies and the general public. From the viewpoint of the NIH as a major research funding agency,¹³ data sharing has the advantage of avoiding unnecessary duplication of data collection, allowing available funds to be used to support more researchers.

2. The Economic Benefits of Data Sharing

1.06 Identification of the economic benefits to be gained from improved access to publicly funded research results has been a focus of recent Australian studies. In the 2006 Australian Government Department of Education, Science and Training (DEST) commissioned report, *Research Communication Costs in Australia: Emerging Opportunities and Benefits*,¹⁴ Houghton, Steele and Sheehan concluded that new models for scholarly communications, such as open access, have the potential to increase the economic and social returns to public investment in research and development (R&D). They comment:

Perhaps the most important potential benefit of open access is enhanced access to, and greater use of, research findings, which would, in turn, increase the efficiency of R&D as it builds upon previous research. There is also significant potential for open access to expand the use and application of research findings to a much wider range of users, well beyond the core research institutions that have had access to the subscription-based literature...

...Estimating the benefits of a one-off increase in accessibility and efficiency (*e.g.* because of a move to open access), we find that if accessibility and efficiency are constant over the estimation period but then show a one-off increase, then, to a close approximation, the return to R&D will increase by the same percentage increase as that in the accessibility and efficiency parameters. Assuming that the increase in both parameters is the same, that the change to open access has no *net* impact on the rates of accumulation and obsolescence of the stock of knowledge, and that the information are discoverable, we find that:

- With *public sector R&D expenditure* at AUD 5,912 million and a 25% rate of social return to R&D, a 5% increase in accessibility and efficiency would be worth AUD 150 million a year;
- With *higher education R&D expenditure* at AUD 3,430 million and a 25% rate of social return to R&D, a 5% increase in accessibility and efficiency would be worth AUD 88 million a year; and
- With *ARC administered funding* (competitive grants) at AUD 480 million and a 25% rate of social return to R&D, a 5% increase in accessibility and efficiency would be worth AUD 12 million a year (Table A2). Note that these are recurring annual gains from the effect on one year's R&D. Assuming that the change is permanent they can be converted to growth rate effects.¹⁵

¹² S E Fienberg, M E Martin and L M Straf, *Sharing Research Data* (1985).

¹³ NIH is the largest funder of basic biomedical research in the world, with an expenditure of US\$27 billion in FY 2005: source Claire Driscoll, Director, Technology Transfer Office, National Human Genome Research Office.

¹⁴ J Houghton, C Steele and P Sheehan, *Research Communication Costs in Australia: Emerging Opportunities and Benefits* (2006) <http://www.dest.gov.au/NR/rdonlyres/0ACB271F-EA7D-4FAF-B3F7-0381F441B175/13935/DEST_Research_Communications_Cost_Report_Sept2006.pdf>; see also J Houghton and P Sheehan, 'The Economic Impact of Enhanced Access to Research Findings' (Working Paper No 23, Centre for Strategic Economic Studies, Victoria University 2006) <<http://www.cfes.com/documents/wp23.pdf>>.

¹⁵ Ibid vi, ix, 44 – 46, 56

1.07 It is only through the technological advances that have occurred in recent years that the benefits that stand to be gained by making research outputs (whether in the form of published papers or data) more readily available to the research community have begun to be fully realised.¹⁶ Various factors including grid enabled technologies, high performance computing, greatly increased computing power, high speed broadband networks, the development of the world wide web and reductions in the cost of computing and collaboration have had a major impact not only on the way research data is generated but also on how it is disseminated and used.

3. New Research Methodologies

1.08 Technological developments have fundamentally changed the way research is carried out. Until quite recently, a large proportion of research data was produced by small groups of scientists and individual researchers who kept their data secret and maintained proprietary control over access to the raw data and their analyses of it. Increasingly however, scientists are involved in data-intensive research projects which cut across geographic and disciplinary borders.¹⁷ Quality research now often involves virtual communities of researchers participating in large-scale web-based collaborations, opening their early-stage research to the research community in order to encourage broader participation and accelerate discoveries. An example is the Earth System Grid (ESG), an experimental data network that integrates supercomputing power with large-scale data and analysis servers for scientists collaborating on climate studies. The grid is expected to speed up the execution of climate models 100-fold and allow scientists to use the community's distributed data systems to perform high-resolution, long-duration simulations.¹⁸

1.09 Advances in information and communications technologies will continue, ensuring an ongoing revolution in the way data is produced, managed and used. As Burk comments:

The advent and proliferation of global computer networks have altered the practice of science and additional changes seem sure to come. Scientists already routinely collaborate and access informational resources by way of the internet and associated technologies. Further advances in this direction are contemplated, via so-called "Grid" technologies to enable collaborative sharing of both information and resources on an international scale. Such distributed computing architectures promise to make available processing power, data storage, and related large-scale computing resources independent of geographic location. Researchers participating in such technological collaborations are increasingly drawn into distributed communities and far-flung alliances that might previously have been impossible.¹⁹

1.10 Accompanying these developments has been a growing recognition that if the benefits of enhanced access are to be realised through e-Research, it will be necessary to develop the systems and services that enable data to be managed and

¹⁶ See Productivity Commission, *Public Support for Science and Innovation*, Research Report (2007) [5.7] <<http://www.pc.gov.au/study/science/finalreport/index.html>> at 3 April 2007.

¹⁷ ICSU, *Scientific Data and Information: A report of the CSPR Assessment Panel* (2004) 7.

¹⁸ Don Tapscott and Anthony Williams, 'The New Science of Sharing', *BusinessWeek.com* 2 March 2007, <http://www.businessweek.com/innovate/content/mar2007/id20070302_219704.htm?chan=technology_technology+index+page_more+of+today's+top+stories>.

¹⁹ Dan Burk, *Intellectual Property in the Context of E-Science*, Minnesota Legal Studies Research Paper No. 06-47 (2006) 2 <<http://ssrn.com/abstract+929479>> at 18 August 2006.

secured.²⁰ In the Australian context, the National Collaborative Research Infrastructure Strategy (NCRIS) refers to the infrastructure for e-Research as “platforms for collaboration”.

4. e-Research and Cyberinfrastructure

1.11 Key recommendations relating to the development of e-Research collaborative platforms were made by the Prime Minister’s Science, Engineering and Innovation Council’s Working Group on Data for Science (PMSEIC Data Working Group) in its report, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006).²¹ It recommended that:

Australia’s government, science, research and business communities establish a nationally supported long-term strategic framework for scientific data management, including guiding principles, policies, best practices and infrastructure.²²

1.12 Several other countries are developing national research infrastructures, although there are differences in terminology: the term “cyberinfrastructure” is used in the United States (US),²³ “e-infrastructure”²⁴ in the United Kingdom (UK)²⁵ and Europe²⁶ and “GRID” in Canada.

1.13 In the US, the term “cyberinfrastructure” was coined by a National Science Foundation (NSF) Blue Ribbon Committee Report:

²⁰ For a complete definition of ‘e-Research’, see Chapter 2.

²¹ Prime Minister’s Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006) <http://www.dest.gov.au/sectors/science_innovation/publications_resources/profiles/Presentation_Data_for_Science.htm> at 26 March 2007.

²² Ibid 11- Recommendation 1. Note also Recommendation 2: “That a high-level expert committee be established to provide the leadership role in progressing the formation of the long-term strategic framework for scientific data management.”

²³ See the conference ‘Designing Cyberinfrastructure for Collaboration and Innovation’ (National Academies, Washington DC January 2007) <<http://www.si.umich.edu/cyber-infrastructure>>. See also, the National Science Foundation website on cyberinfrastructure <<http://www.nsf.gov/crssprgm/ci-team/>>.

²⁴ Older references in UK research literature use the term “e-Science”. The National e-Science Centre (NeSC) web site at <<http://www.nesc.ac.uk/index.html>> at 17 May 2007 states:

What is meant by e-Science? In the future, e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists.

²⁵ UK Office of Science and Innovation (OSI) e-Infrastructure Working Group, *Developing the UK’s e-Infrastructure for Science and Innovation* (2007) <<http://www.nesc.ac.uk/documents/OSI/report.pdf>>. The e-Infrastructure Working Group comprised senior representatives from JISC (Joint Information Systems Committee), the Research Councils, RIN (Research Information Network) and the British Library, the Working Group was formed in response to the *Science and Innovation Investment Framework 2004-2014*, published by the Treasury, the DTI and the DfES in 2004, to explore the current provision of the UK’s e-infrastructure and help define its future development. See <http://www.jisc.ac.uk/publications/publications/pub_einfrastructurebp.aspx> at 23 April 2007.

²⁶ To give impetus to the development and use of digital repositories for science data, the European Commission’s Information Society and Media DG in 2007 commissioned a study entitled ‘Towards a European e-Infrastructure for e-Science Digital Repositories’ (‘e-SciDR’). See information at the conference site <<http://www.e-scidr.eu/>> at 23 April 2007.

The newer term cyberinfrastructure refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy.²⁷

1.14 Cyberinfrastructure is a non discipline-specific term that refers to computing systems, data storage systems, advanced instruments and data repositories, all linked together by high speed communications networks.²⁸

5. Data Management as Infrastructure

1.15 In this report, the term "infrastructure" is used as an umbrella term to refer to the systems and frameworks that operate to control and manage how data is collected, stored and disseminated. "Systems" are the technological systems and mechanisms, including hardware, software and digital repositories, which contain data and make data viewable and accessible to a wider audience. "Frameworks" (and "protocols") are the legal frameworks that govern how data can be managed and shared in accordance with contractual obligations, copyright and any other rights granted or restrictions imposed by law.

1.16 While these new collaborative research platforms have been made possible by advances in information and communications technologies, they also require active and professional management of the processes by which data is generated, organised, evaluated and disseminated. The importance of professional management of research data and information has consistently been identified as central to data and information infrastructures. The International Council for Science (ICSU) commented in its 2004 report, *Scientific Data and Information: A report of the CSPR Assessment Panel*:

Scientific data and information management can no longer be viewed as a task for untrained amateurs or as part of the routine "clean up" conducted hurriedly by scientists at the completion of a research project. It remains a responsibility of all scientists and should be valued accordingly, but it is also an increasingly important professional activity, one that is essential to the scientific enterprise. Because data centres and permanent archives are now among the most critical components in the infrastructure of science and constitute the legacy that the current generation of scientists will leave to its successors, working partnerships between scientists and data managers will increasingly be required. All scientists need to have some data management awareness, but the use of advanced information technology in scientific data management and dissemination makes it essential that data management be the responsibility of professionals. There is a need for improved management of data in research projects and for dedicated individuals and institutions to disseminate, manage, and archive scientific data and information. Within research projects, data management must be recognised as an essential component in overall project management that takes place in parallel with other research activities. ... [B]y providing widespread access to well-documented and managed research data, improved data management practices will provide economies of scale for the scientific enterprise as a whole, now and in the future. In particular, good data and information management should significantly improve access to useable data among scientists in developing and transitional countries.²⁹

²⁷ National Science Foundation (NSF) Blue Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyberinfrastructure* (2003) 5 <<http://www.nsf.gov/od/oci/reports/atkins.pdf>> at 21 March 2007.

²⁸ Teragrid Knowledge Base, *What is cyberinfrastructure?* <<http://www.teragrid.org/cgi-bin/kb.cgi?docid=auhfh>> at 21 March 2007.

²⁹ ICSU, *Scientific Data and Information: A Report of the CSPR Assessment Panel*, (2004) 20.

1.17 DEST's *National Collaborative Research Infrastructure Framework – Strategic Roadmap* (NCRIS Strategic Roadmap)³⁰ identified five key inter-related components of collaborative platforms, one of which comprises the activities of “data storage management, access, discovery and curation”.³¹

1.18 The PMSEIC Data Working Group in its report, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006),³² recommended that:

Data management expertise become a core skill for researchers, including graduate and postgraduate science students across all disciplines, and that they receive data management training as part of their education.³³

6. Legal Framework as e-Research Infrastructure

1.19 It has become increasingly apparent that to achieve seamless access to data it is necessary not only to adopt appropriate technical standards, practices and architecture, but also to develop legal frameworks that facilitate access to and use of research data, whether on an inter-organisational basis or across national borders.³⁴ The benefits that may potentially be gained through advances in information and communications technologies will not be achieved solely by engineering but will result from a combination of social, legal and technical factors. Professor Paul David has observed that:

Engineering breakthroughs alone will not be enough to achieve the outcomes envisaged for [e-Science].Success in realizing the potential of e-Science—and other global collaborative activities supported by the “cyberinfrastructure” - if it is to be achieved, will more likely be the resultant of a nexus of interrelated social, legal *and* technical transformations. The socio-institutional elements of a new infrastructure supporting research collaborations - that is to say, its supposedly “softer” (non-engineering) parts—are every bit as complicated as the hardware and computer software, and, indeed, may prove much harder to devise and implement. The roots of this latter class of challenges facing “e-Science” lie in the micro- and meso-level incentive structures created by the existing legal and administrative regimes.³⁵

³⁰ Australian Government, Department of Education, Science and Training (DEST), *National Collaborative Research Infrastructure Strategy (NCRIS)* <http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/ncris/> at 28 February 2006.

³¹ Other key elements of platforms for collaboration are: high-performance computing; grid-enabled technologies and infrastructure; network access through high capacity bandwidth; and support skills to assist researchers in developing and using infrastructure.

³² Prime Minister's Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science*, December 2006, <<http://www.dest.gov.au/pmseic>>

³³ Ibid 13 - Recommendation 10. The PMSEIC Data Working Group also recommended the Australian government to give early consideration to the findings of the e-Research Coordinating Committee regarding changing research behaviour, practices and skills: Recommendation 11.

³⁴ D Greenbaum and M Gerstein, 'A universal legal framework as a prerequisite for database interoperability' (2003) 21 *Nature Biotechnology* 979.

³⁵ Paul A David, “Towards a cyberinfrastructure for enhanced scientific collaboration: providing “soft” foundations may be the hardest part”, Oxford Internet Institute Research Report No. 4 (2004 revised May 2005) <<http://www.oii.ox.ac.uk/research/publications.cfm>> at 30 April 2007.

1.20 The NCRIS Strategic Roadmap (2006)³⁶ acknowledged that the management of research outputs requires the coordination of many elements, including the appropriate hardware and software, supporting workflows, policy and regulatory frameworks, administrative arrangements and resources. Importantly, the NCRIS Strategic Roadmap makes the point that while much of the work on data access to date has focused on the use of technical mechanisms to overcome barriers to access,³⁷ it is also necessary to ensure that the legal context is understood and that IP interests (notably copyright) are effectively managed.

Seamless access to information and other resources can be impeded...particularly in a networked environment, if researchers are not mindful of intellectual property law. In many cases, there is no certainty. A key challenge for the future is to establish legal protocols that can allow access to, or downloading of, research to be clarified and simplified.³⁸

7. Guidelines

1.21 Growing awareness of the need to develop procedures for the guidance of researchers on ownership and use rights in relation to research data and materials is found in the 2006 consultation draft of the *Joint NHMRC/AVCC Statement and Guidelines on Research Practice – Australian code for the responsible conduct of research*.³⁹ Several paragraphs of the code explicitly address the responsibilities of research organisations in actively addressing the management of IP and other legal issues relating to research data and records. These include the following paragraphs:⁴⁰

3.3 Identify ownership of research data and records

Institutions must have policies on the ownership of research materials, research data, databases and other material retained at the end of a research project, where this is not established by the funding arrangements for the project. As a general rule, the most satisfactory arrangement will be that the materials and data retained at the end of a project are the property of the institution that hosted the project, another institution with an interest in the research, or a central repository.

As far as possible, the data should be available for use by other researchers.

3.4 Ensure security and confidentiality of research data and records

It is important that institutions have policies and procedures for the establishment and ownership of, and access to, databases and archives containing confidential information, and that these procedures are consistent with relevant legislation or other guidelines, including privacy guidelines. To achieve this, institutions have the following responsibilities:

³⁶ Australian Government, Department of Education, Science and Training (DEST), *National Collaborative Research Infrastructure Strategy (NCRIS)* <http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/ncris/> at 28 February 2006.

³⁷ Such as hardware and software

³⁸ Australian Government, Department of Education, Science and Training (DEST), *National Collaborative Research Infrastructure Strategy (NCRIS)* <http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/ncris/> at 28 February 2006 [5.16.1].

³⁹ NHMRC/AVCC, *Australian code for the responsible conduct of research*, Second Consultation Draft, <http://www.nhmrc.gov.au/publications/_files/acrcr.pdf>.

⁴⁰ Other relevant paragraphs are 2.5 (Provide clear contractual arrangements) and 3.2 (Provide secure research data storage and record-keeping facilities).

3.4.1 Develop procedures to guide researchers on matters concerning the ownership and use of research data and original material, and the confidentiality of these.

5.2 Protect confidentiality and manage intellectual property

Institutions must have policies for protecting confidentiality and intellectual property. Therefore, institutions have the following responsibilities:

5.2.1 Ensure that there are processes to make all parties to the research aware of the nature of confidentiality provisions where they apply (see also paragraph 3.7).

5.2.2 Follow policies that protect the intellectual property rights of the institution, the researcher and sponsors of the research.

5.2.3 Ensure that researchers are aware of contractual arrangements that limit publication.

5.2.4 Ensure that external sponsors of research understand the importance of publication in research and that sponsors do not discourage publication or dissemination of research findings for longer than the minimum time required for the necessary protection of intellectual property or other relevant interests (usually 6 to 12 months).

9.1 Have written agreements for each collaboration

Institutions involved in a multi-institutional research project, or in a collaborative project between public and private research organisations and sponsors, need to have written agreements to manage all aspects of the research. The written agreement must cover the following:

- intellectual property
- copyright issues
- sharing commercial returns
- reporting to appropriate agencies (see also paragraph 2.9 about contractual arrangements).

1.22 Institutions have a responsibility to ensure that researchers understand the policies and written agreements of multi-institutional research collaborations. The need for guidelines for digital data management was also highlighted in the Australian Partnership for Sustainable Repositories' (APSR) report, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne*, published in August 2006.⁴¹ The report, based on an audit of the data management practices of eleven research communities from diverse disciplines at the University of Melbourne during 2006, made eight recommendations, including⁴²:

⁴¹ Anna Shadbolt, Dirk van der Kniff, Eve Young and Lyle Winton, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories*, August 21 2006, <http://www.apsr.edu.au/aeres/sustainable_paths.pdf> at 18 June 2007, p2-4; 38-42

⁴² Ibid p 2 - 4

- Recommendation 1: That the University develops a strategy that broadly addresses the policy, infrastructure, support and training needs of eResearch;
- Recommendation 3: That Information Services initiate a consultative process for the development of appropriate guidelines and, where relevant, policy statements, to support researchers with the management of their research data and records;
- Recommendation 7: To establish and registry of e-research expertise; and
- Recommendation 8: To review the implications of project findings for researcher education and training.

1.23 The APSR's findings demonstrated varying capabilities among the University of Melbourne's research communities to comply with the University's Policy on the Management of Research Data and Records and the draft NHMRC/AVCC *Australian code for the responsible conduct of research*. In general, the APSR found a lack of best practice guidelines and policy statements to support researchers in making decisions about data management. Consequently, APSR recommended (recommendation 3) that consultations be initiated "for the development of appropriate guidelines and, where relevant, policy statements, to support researchers with the management of their research data and records". The APSR identified a need to for researcher education and training, to assist them to acquire and develop skills in data management, including "an understanding of research data policies, responsibilities, collections, curation, preservation, copyright/IP, metadata and standards" (recommendation 8).

1.24 Data management policies and principles, data management plans and data management toolkits are further discussed in Chapter 9.

8. Purpose of the Report

1.25 The focus of this report is the development and implementation of legal frameworks for the generation, management, use and dissemination of research data and information.

1.26 **Chapter 2** introduces key concepts and terms used throughout this report. It sets out working definitions of the terms "data", "dataset" and "database" and explains what the concepts of "ownership", "control" and "use" mean when applied in relation to research data and databases. The chapter identifies the various persons who may claim to have rights in relation to data, including creators, consumers, compilers and research funding bodies, and demonstrates that the ability to exercise control over data or a database is not necessarily concomitant with ownership. A brief overview is provided of the principal areas of law that have a direct bearing on the existence and exercise of rights to own, control and use data, namely, copyright, patent, contract, confidential information and privacy law.

1.27 **Chapter 3** examines the regulatory context in which data is generated, managed and used. Particular focus is given to legislative and administrative provisions applying to the collection, storage, maintenance and archiving of data and information in the public sector and measures regulating access to materials held by public sector entities. For example, public records (or archives) legislation provides for the archiving of and access to public records, while freedom of information

legislation regulates access to documents of government agencies. Some governments have established administrative policies and guidelines with which public authorities must comply in the responsible collection and management of data and information (for example, Queensland's Information Standards). This chapter also considers the operation of legislative protection of personal information ("data privacy") which is of direct relevance to the activities of research organisations involved in the collection or use of personally identifying information and health information about human subjects. The *Privacy Act 1988* (Cth) establishes Information Privacy Principles and National Privacy Principles, which make it unlawful for certain bodies to collect and use personal information unless it is done for a permitted purpose. The relevance of cultural protocols which may affect how some data, including social science and anthropological data and information about biological resources derived from traditional knowledge, is collected, managed and made accessible to the wider public is also considered.

1.28 **Chapter 4** surveys and describes arrangements that have been developed to provide access to and facilitate the sharing of research data in Australia and other jurisdictions. It examines current practices as exemplified in general frameworks for data access and sharing as well as in subject-specific databases, especially collections of medical and genetic research data. Arrangements for ownership, control and use of data in a sample of web-accessible United States and European databases are described, as well as in several Australian databases, and observations are drawn from examining these arrangements. Evidence of the attitudes of researchers towards data sharing is considered, in order to identify factors that need to be taken into account in further developing technical and legal systems to facilitate data access and sharing. The findings of surveys of researchers' attitudes carried out by the Australian eResearch Sustainability Survey undertaken by APSR and the NCRIS Platforms for Collaboration of Data Management Survey are considered. A survey of attitudes to data sharing recently carried out by Queensland University of Technology (QUT) is also described, as well as observations drawn from a preliminary analysis of the survey findings.

1.29 **Chapter 5** provides an overview of the basic principles of copyright, focusing specifically on how copyright applies to data, information, datasets and databases. The position in relation to the protection of data compilations under Australian copyright law following the decision of the Federal Court in *Desktop Marketing Systems Pty Ltd v Telstra Corporation Ltd* [2002] FCAFC 112 is explained. A brief outline is provided of the expanded range of rights that can be exercised by copyright owners following recent amendments to the *Copyright Act* to strengthen the protection available for digital materials which are made available in the online networked environment. Digital rights management (DRM) and electronic rights management information (ERMI), the technological mechanisms used by copyright owners to protect their copyright material, are explained. The chapter considers the principal means used by copyright owners to grant permissions to others to make use of their copyright materials, that is, contractual licences and non-contractual permissions. It explains the use of open content licensing, including Creative Commons and Science Commons, to grant a wide range of permissions to make use of copyright materials, while reserving some rights (such as the right to be attributed) to the copyright owner. The operation of copyright law in key overseas jurisdictions, notably the United States and the United Kingdom, is examined, and the *sui generis*

(unique) database right introduced by members of the European Union in accordance with the Database Directive is briefly considered.

1.30 **Chapter 6** explains the different circumstances in which data or information can be considered confidential so as to attract legal protection. An overview is provided of the action for breach of confidence, to illustrate how this action can be used to control and limit access to data that has not yet been publicly disseminated. The special categories of government information and information generated or obtained during employment and the issues of ownership rights in relation to the information and rights to access the information are addressed. The chapter also explains how the quality of confidentiality can be lost and how contract can be used to control access to data notwithstanding that confidentiality may have been lost.

1.31 **Chapter 7** considers various contractual arrangements relevant to the protection and sharing of data and information. Contracts may be used to control access to data to protect confidentiality or commercial interests, or may impose conditions on the use of data to which access has been granted. Conditions are commonly imposed restricting the use of data to specific purposes or projects. The interplay between contract and copyright law is explained, and how a copyright owner may, by means of a contractual licence, grant permissions to licensees to exercise rights in copyright-protected data, datasets and databases.

1.32 **Chapter 8** provides an overview of open access policies and principles, specifically as they relate to data generated through scientific research. The increasing range and volume of research data raises questions about whether, and on what basis, research data is to be made accessible to others. As can be seen from the databases surveyed in Chapter 4, different approaches have emerged for managing information outputs. There is a need for clear policies and procedures that address key issues including identification and selection of data to be made available for access and reuse, funding of data management systems, intellectual property rights, and the role of repositories. Chapter 8 aims to present a wide-ranging overview of policy statements and guidelines relating to open access and data sharing, to provide a basis for understanding the principles and frameworks that have been developed and are rapidly emerging in Australia and other jurisdictions. Summaries are provided of key international statements on open access, including the Bermuda Principles (1996), the Budapest Open Access Initiative (2002), the Bethesda Statement on Open Access Publishing (2003) and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). The chapter also considers the open access policies and statements of public sector research funding organisations including the ARC and NHMRC in Australia, the National Institutes of Health (NIH) in the United States and the various United Kingdom Research Councils. Open access policies and position statements developed by private sector research organisations, such as the Wellcome Trust in the United Kingdom, are also examined.

1.33 **Chapter 9** considers the steps to be taken in developing legal frameworks for data management for use by the Australian research community. In developing effective frameworks that enable research data to be made available for access and use by other researchers, whether located in Australia or overseas, it is proposed that the following steps be followed:

- formulation of a data access policy and principles;
- identification of the data to be made available for access and use;
- ascertainment of the conditions of access and use for specific kinds of data; and
- adoption of mechanisms for practical data management (e.g. a data management plan and a toolkit for practical management of data by researchers).

1.34 **Chapter 10** sets out the conclusions and recommendations of this report. It proposes steps to be taken on the development of data access policies, principles and practices for the guidance of individual researchers, research groups and institutions in addressing the legal issues integral to data management. Proposals are made for further work on analysing the data access and sharing practices described in Chapter 4 and the open access policies described in Chapter 8. Proposals are also made for the development of templates for data management plans and data management toolkits, as discussed in Chapter 9.

CHAPTER 2 - KEY CONCEPTS

“Data is everything...”

DEFINING SOME KEY CONCEPTS – “DATA”, “OWNERSHIP” AND “USE”

Aims:

1. Define and discuss key terms and concepts relevant to data, including “data”, “dataset” and “database”, “e-Research”, “ownership”, “control”, “access” and “use”;
2. Identify the different parties who may own or exercise rights in relation to data or who may be able to control access to and use of research data; and
3. Provide an initial overview of the principal areas of law relevant to the management of research data, particularly copyright, patent, confidentiality and contract law.

1. Introduction

2.01 It is important to establish key terms and concepts before proceeding to consider the main components of the legal framework for management of research data and information and proposing structures designed to enable access to and use of research data and information.

2.02 Key terms and concepts that are used throughout this Report and which require a common understanding are:

- “data”, “information”, “datasets” and “databases”;
- “e-Research”;
- “ownership” and “control”; and
- “access” and “use”.

2. “Data” and “information”

2.03 The terms “data” and “information” can be interchangeable. “Data” refers to research results, facts, and statistical or survey information, including text, numbers, images, audio and video recordings, software, animations, metadata and model simulations. In the digital context, “data” refers to any information that can be stored in digital form.

2.04 The PMSEIC Data Working Group adopts a broad definition of “data”, which includes:

Data from the social sciences and humanities as well as other scientific disciplines such as astrophysics, mathematics and humanities and biology, and information collected not just by

scientists or researchers but by agencies for administrative purposes such as health, welfare, population, education, employment and crime.⁴³

2.05 The International Council for Science (ICSU) in its report, *Scientific Data and Information: A report of the CSPR Assessment Panel* (2004) explains “data” and “information” as follows:

“Data” includes, at minimum, digital observations, scientific monitoring, data from sensors, metadata, model output and scenarios, qualitative or observed behavioural data, visualisations, and statistical data collected for administrative or commercial purposes. Data are generally viewed as input to the research process.

“Information” generally refers to conclusions obtained from analysis of data and the results of research. But the distinction between them is flexible and will vary according to the situation. Increasingly, the output of research (traditionally viewed as “information”) includes data and has become input into other research, rendering the output-input distinction between data and information meaningless.

In this report, both “data” and “data and information” are used interchangeably to refer to the entire continuum because the continuum as a whole has been affected by changes in information technology and is subject to many of the same issues. ... [A distinction can be made] between “data” and “scientific publications”, which are a specific sub-set of scientific information that raise particular issues.⁴⁴

2.06 The term “data” is used in a global sense to refer to all kinds of data in digital form, whether it is raw data (gathered by any method, including simulation and modelling) or data that has been derived from raw data by processes of consolidation, selection, calculation, statistical analysis or other methods.

2.07 In the Organisation for Economic Cooperation and Development’s (OECD) *Principles and Guidelines for Access to Research Data from Public Funding*,⁴⁵ “research data” is defined as:

Factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.

2.08 However, the following are excluded from the scope of “research data”:

Laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g., laboratory samples, strains of bacteria and test animals such as mice).

2.09 What follows are some diagrammatic examples of what constitutes data, datasets and databases.

⁴³ Prime Minister’s Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006) 15 <http://www.dest.gov.au/sectors/science_innovation/publications_resources/profiles/Presentation_Data_for_Science.htm> at 26 March 2007.

⁴⁴ ICSU, *Scientific Data and Information: A Report of the CSPR Assessment Panel* (2004) 14.

⁴⁵ OECD, *Recommendation of the Council Concerning Access to Research Data from Public Funding* (2006)184 <<http://webdomino1.oecd.org/horizontal/oecdacts.nsf/Display/3A5FB1397B5ADFB7C12572980053C9D3?OpenDocument>> at 23 April 2007.

Figure 1 provides a visual representation of “data”:

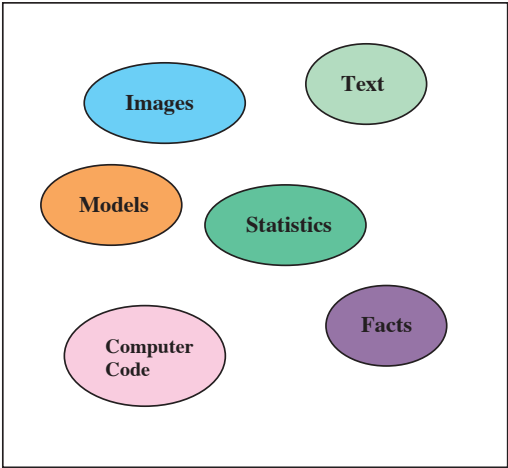


Figure 1: Individual data items

3. “Datasets”

2.10 “Datasets” is used to refer to compiled collections of data, which may be collected from diverse sources and be in different formats. Figure 2 provides a visual representation of “dataset”:

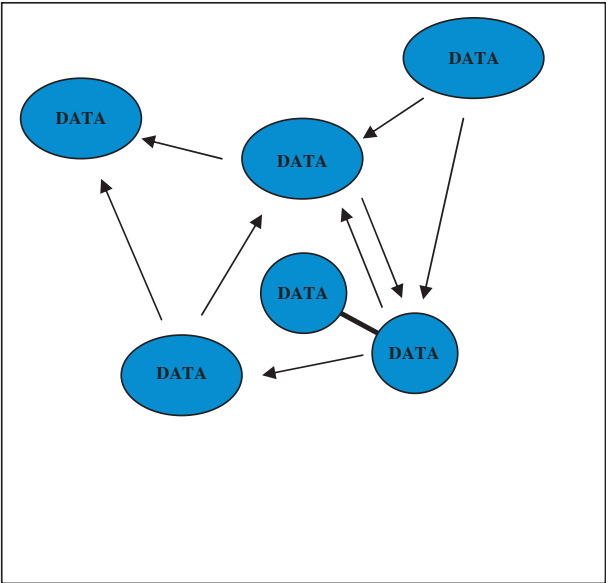


Figure 2: Dataset

4. “Database” (or “data archive” or “repository”)

2.11 “Database” refers to a collection of data and datasets, often compiled from a range of sources and usually organised to permit data to be readily retrieved, managed and updated. Typically databases involve software programs which enable the data to be collected, copied, stored, retrieved and distributed.

2.12 Other terms used to refer to data collections are “data archive” and “repository”. NIH uses the term “data archive”, which it defines as:

A place where machine-readable data are acquired, manipulated, documented, and finally distributed to the scientific community for further analysis.⁴⁶

2.13 The PMSEIC Data Working Group uses the term “data archiving” to refer to the “curation activity which ensures that data are properly selected and stored”, but refers to the location where archived data is kept as a “repository”, which it defines as follows:

A central place where data are stored and maintained. A repository can be a place where multiple databases or files are located for distribution over a network, or can be a location that is directly accessible to the user without having to travel across a network.

A digital repository is either a local, institutional, or central (e.g. subject-based or discipline-based) digital archive for depositing and providing access to digital contents.⁴⁷

2.14 A database (or data archive or repository) may contain data that cannot be accessed by the general public because of, for example, confidentiality obligations, third-party licensing agreements or national security obligations. A controlled, secure environment in which only eligible researchers are permitted to obtain access to data is sometimes referred to as a “data enclave”.⁴⁸

⁴⁶ NIH Office of Extramural Research, *NIH Data Sharing Policy and Implementation Guidance* (updated 5 March 2003)

<http://grants2.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#archive>.

⁴⁷ Prime Minister’s Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006)17, <<http://www.dest.gov.au/pmseic>>.

⁴⁸ NIH defines ‘**data enclave**’ as ‘a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources’; ‘**restricted data**’ is defined as ‘datasets that cannot be distributed to the general public, because of, for example, participant confidentiality concerns, third-party licensing or use agreements, or national security considerations’. See

<http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave>.

Figure 3 provides a visual representation of “Database”:

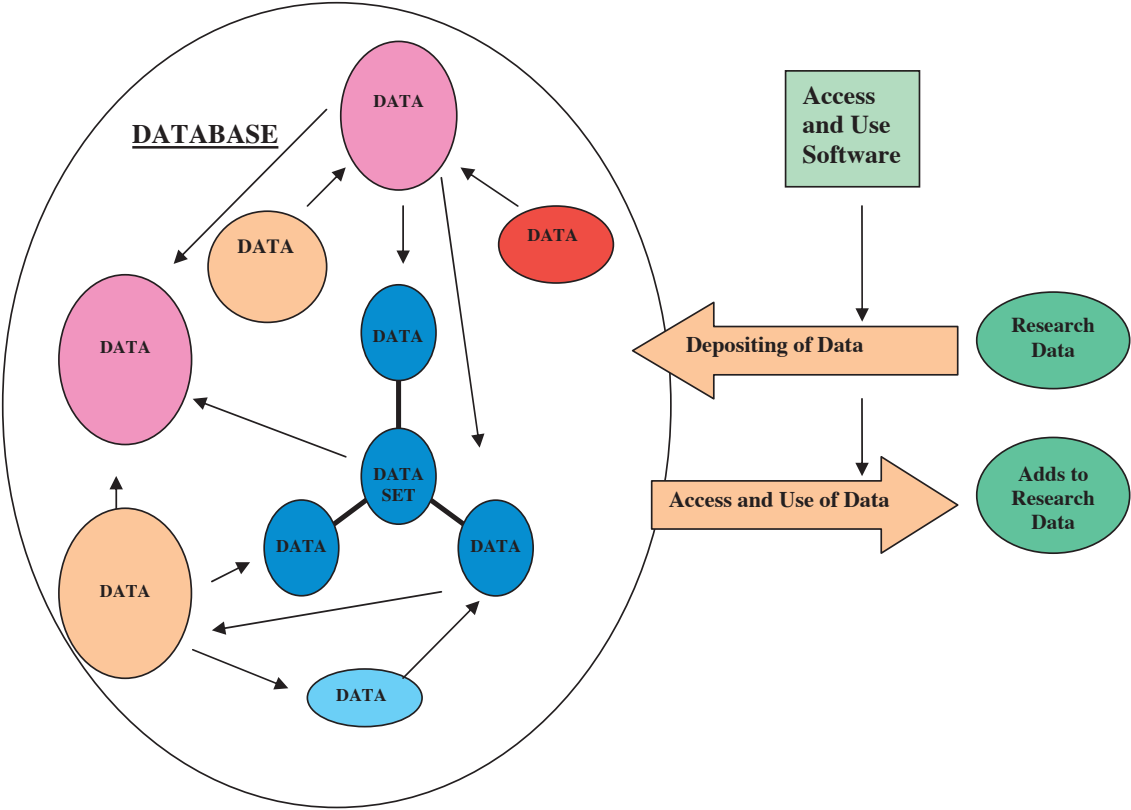


Figure 3: Database

5. “e-Research”

2.15 The term “e-Research” is used in this report in the context of research in a digital environment.

The term ‘e-Research’ encapsulates research activities that use a spectrum of advanced information and communications technologies (ICT) capabilities and embraces new research methodologies emerging from increasing access to:

- broadband communications networks, research instruments and facilities, sensor networks, data repositories and their associated data standards, management and curation tools, and high performance computing resources;
- software and infrastructure services that enable secure connectivity and interoperability between researchers and the wide variety of data repositories, computers, systems and networks on which they depend;

- application and discipline-specific tools such as graphics intensive visualisation and simulation software, and interaction tools that provide the human interface allowing researchers to interact with each other and with their instruments, computational tools and data resources.⁴⁹

6. “Ownership” and “Control”

2.16 A fundamental concept in the context of data management systems for research data is that of “ownership”. When applied to data and databases, the concept of “ownership” is potentially complex.⁵⁰ Reference to the “owner” of data will usually be a reference to:

- the person who owns legal rights in relation to the data;
- the person who has physical possession of the data;
- the person who controls access to and use of the data; or
- a combination of any of the above.

2.17 Thus, the “owner” may be the person who has generated or collected the data, the custodian of the data or the manager of the database. A simple view tends to equate ownership purely with data custodianship or curation. However, while a party who is in possession of data or a database and is responsible for maintaining its integrity (i.e. the “custodian”) may be in a position to exercise control over it, they may not be the owner in a legal sense.⁵¹

2.18 The term “control” is used to indicate the rights of a person or entity over research data which enable them to determine how research data may be used. The controller of research data may be the owner, or some other person or entity to whom the owner has granted the rights to control the use of the data.

2.19 The question of what is meant by “ownership” and what it entails becomes much more important in a context where it is proposed that data is gathered from different sources and compiled in one or more repositories to permit other parties to obtain access to and use of the data. It is possible to identify numerous parties who may claim rights in data, based on the kind and extent of their contribution to the research endeavour. For example:⁵²

- the creator – the party who creates or generates the data;
- the consumer – the party who uses the data;

⁴⁹ B Fitzgerald and J Abbot, *Legal Framework for e-Research* (2006) <<http://eprints.qut.edu.au/archive/00005311/01/5311.pdf>> at 26 April 2007, citing ‘An E-Research Strategic Framework: Interim Report of the E-research Coordinating Committee’ (2005) <<http://www.dest.gov.au/NR/rdonlyres/B6F765A7-DD2C-432B-9064-2F9CD4E17E66/10518/InterimReport2.doc>>.

⁵⁰ Responsible Conduct in Data Management, *Data Ownership* <http://ori.dhhs.gov/education/products/n_illinois_u/datamanagement/dotopic.html>; E A Fishbein, ‘Ownership of research data’ (1991) 66(3) *Academic Medicine* 129-33; C Shores, ‘Ownership of data’ (Paper presented at AUTM Conference, March 2002).

⁵¹ For more on custodianship, see Chapter 8, OSDM’s Custodianship Guidelines

⁵² D Loshin, *Knowledge Integrity: Data Ownership* (2002)

<<http://www.datawarehouse.com/article/?articleid=3052>> at 8 June 2004.

- the compiler – the party who selects and compiles information from different information sources;
- the funder – the party who commissions the data to be generated;
- the decoder – where informed is protected by encoded formats (e.g. encryption), the party who can unlock the information;
- the packager – the party who collects information for a particular use and adds value through formatting it for a particular market or set of consumers;
- the reader – the person who reads data added to an information repository;
- the subject of the data – the person from whom the data is derived or who the data is about; and
- the purchaser or licensee – the party who buys or licences the data.

2.20 Identification of the party or parties who have ownership rights in relation to data is not necessarily a straightforward exercise. Nevertheless, since the question of ownership of rights in relation to research data has implications for the dissemination of data, it is an issue that needs to be addressed and resolved before metadata about rights is assigned to each data file. An illustration of the difficulties that may arise in ascertaining ownership is found in the Joint Information Systems Committee's (JISC) *Project Spectra Final Report* (2007) which comments:

The project encountered uncertainties and inconsistencies among members of the chemistry communities at Cambridge [University] and Imperial [College] in the interpretation and application of their ownership rights to research data. For example, both the chemist who creates a substance and the crystallographer who analyses it may wish to assert rights over the outputs.⁵³

2.21 From the legal perspective, an “owner” in relation to research data or information generally means the person or entity who may exercise rights in relation to the data or information. A more accurate expression may therefore be the “rights-owner” or “rights-holder”. Rights of ownership may arise on different legal grounds, the most immediately relevant being copyright, the common law (through the protection afforded to confidential information) and contract. As the “ownership” rights that may arise through the operation of each of these bodies of law differ, it is necessary to understand what they entail, in order to reliably identify the owner or owners and the rights they can exercise. For example, a person may be in possession of data and therefore be able to control who may access and use the data. They can be said to own this right of access and control of the data. In certain circumstances, this right may be exercised by someone else with the permission of the rights-owner. A custodian may be given the responsibility to control who may access the data or a user may be given permission to reuse the data. Unless otherwise stipulated, allowing others to exercise such rights will be a mere permission and will not transfer ownership of the rights from the rights-owner.

2.22 Research data and information are typically included under the heading of “Intellectual Property” (referred to throughout this report as “IP”) both in legal theory and in practice. Although data and information are commonly included within the

⁵³ Alan Tonge and Peter Morgan, *Project SPECTRA: Submission, Preservation and Exposure of Chemistry Teaching and Research Data* JISC Final Report (2007) 16–17
<http://www.lib.cam.ac.uk/spectra/documents/SPECTRA_Final_Report_v10.doc> at 30 April 2007.

broad scope of IP and referred to as being “owned” by a person or organisation, it is important to understand that data and information are not, in themselves, property in the usual sense of the word. This is the case even though data and information may be valuable, whether in terms of actual financial value, its potential to contribute to the scientific discovery process or otherwise. For the purposes of this paper, the person or entity that owns IP rights in relation to research data is described as the “owner”, whether such IP rights are based on contract, copyright or common law.

2.23 Copyright: Copyright is the form of IP that is of most direct relevance to research data, datasets and databases, particularly in the context of e-Research. Australian law does not provide any specific form of protection for databases, although data, datasets and databases may all qualify for copyright protection if the statutory criteria are satisfied. Copyright is intangible personal property which can be dealt with in the same ways as other forms of private property: it can be bought, sold, left in a will, used as a security for a loan or licensed. If data is protected by copyright, the owner(s) automatically have a very extensive range of rights that can be exercised to prevent others from dealing with the data. Copyright is dealt with in detail in Chapter 5.

2.24 Confidential information: The common law can prevent unauthorised disclosure of data which is not part of the public domain, where the data is relatively secret and efforts are made to maintain its secrecy. The common law doctrine of confidentiality is examined in Chapter 6.

2.25 Contract: Individuals and research organisations may obtain rights in relation to data under a contract. The terms of the contract may address issues such as rights to control, access, re-use or distribute the data created. Increasingly, the contract under which rights to use data are granted will take the form of a click-wrap agreement formed through a website, with one party clicking an on-screen icon to signify assent to the terms and conditions of access and use displayed on the screen, after which the data can be downloaded or the database accessed. Contractual issues are addressed in Chapter 7.

2.26 Patents: Where a person has used data to create a process or product that is novel, useful and involves an inventive step, they may seek to protect their process or product by filing a patent. A patent is a form of IP protection that imposes a period of time during which only a patent owner or licensee may legally use or exploit the protected invention. Patents are not dealt with extensively in this Report, although some of the issues relating to the patenting of research outputs are discussed in Chapter 4.

7. “Access” and “Use”

2.27 As well as needing to understand ownership interests in the databases they develop or to which they contribute, researchers need to know how they can exercise control over the use of their database by other parties and the basis on which they are permitted to use databases developed by other researchers.

2.28 “Use” in relation to data and databases can have the following meanings:

- “use” in the sense of exercising copyright rights, such as reproduction, adaptation or electronic communication;
- “use” of the data or information to develop a new product, for example where raw data is analysed or combined with data or information from another source to carry out a more complex analysis; and
- “use” of the data or information to carry out a service, such as to produce a consultant’s report based on an interpretation or analysis of the data or information.

2.29 Usage rights in data that exist by the operation of copyright or common law may be subject to contractual provisions or licensing conditions about how they will be exercised. This raises the question of the conditions of use that database owners can impose on researchers who have access to the database, subject to any exceptions under law (for example, the fair dealing for research or study exception to copyright infringement). For example, a copyrighted dataset may be made available under a contractual copyright licence to another party for inclusion in a database being constructed by that party, or rights to access a database containing confidential information may be granted to another party by means of a contract imposing confidentiality obligations on the party to whom access and use is granted.

KEY POINTS

In developing legal frameworks for data management it is necessary to understand:

- *the sense in which terms such as “data”, “datasets”, “databases” and “repositories”, “e-Research”, “ownership”, “control”, “access”, and “use” are used in this context;*
- *the principles governing the subsistence and ownership of intellectual property (IP) rights in data, datasets and databases under Australian law (in particular, copyright, patent, confidentiality and contract law); and*
- *how legal and other controls may be exercised to regulate access to and reuse of research data.*

A broad meaning of “data” should be adopted, encompassing not only pure scientific information but also other materials in digital form, including text, software, numbers, graphs, audio and video recordings, and metadata.

A range of persons may claim to be entitled to rights in relation to research data, including the person who has created, collected or generated the data, the person who has funded its production or commissioned it and the consumer who uses the data. Whether these parties can be said to own or control the data will depend upon various factors including the circumstances in which the data has been generated, obligations relating to maintenance and management of the data and the operation of laws (such as copyright and confidentiality) which confer legally enforceable rights exercisable in respect of the data.

There are various legal rights and restrictions that may attach to data and data compilations, including:

- *copyright, in both datasets and databases as compilations, and in data itself where it possesses at least a threshold level of originality;*
- *rights to apply for and obtain a patent, where a patentable invention is derived from the data;*
- *rights to maintain the confidentiality of data which is not in the public domain; and*
- *contractual obligations relating to access to and use of data.*

CHAPTER 3 - THE REGULATORY CONTEXT OF DATA

“On the one hand information wants to be expensive, because it's so valuable. On the other hand, information wants to be free, because the cost of getting it out is getting lower and lower all the time.”⁵⁴

⁵⁴ The statement is generally attributed to Stewart Brand at the first Hackers' Conference in 1984. It was quoted by John Perry Barlow, *The Economy of Ideas: A framework for patents and copyrights in the Digital Age. (Everything you know about intellectual property is wrong)* (1994) Issue 2.03 *Wired* <http://www.wired.com/wired/archive/2.03/economy.ideas_pr.html>. For an historical account of the use of the statement, see Roger Clarke, (2000) *Information Wants to be Free...* <<http://www.anu.edu.au/people/Roger.Clarke/II/IWtbF.html>> at 30 April 2007. See also <http://en.wikipedia.org/wiki/Information_wants_to_be_free#_note-clarke> at 30 April 2007.

THE REGULATORY CONTEXT OF DATA

Aims:

1. Provide an overview of the regulatory context in which data is generated, stored, managed and used, including a focus on data generated by the public sector and through publicly-funded research projects;
2. Examine the legal framework governing the collection, access to and use of personally identifying information and health information in the course of research; and
3. Explain and analyse the operation in practice of these legal provisions on the development of systems for management of research data.

1. Statutory Frameworks Governing Data

3.01 Stewart Brand's remarks at the Hackers' Conference in 1984, taken up as the mantra of the open content movement, highlighted the tensions that arise from the natural propensity of information to transition from secrecy and control into the public domain where it can be freely used by recipients for their own purposes.⁵⁵ The ephemeral nature of information means that rights in relation to data and controls over it exist primarily through the operation of legal rules. In other words, if rights are to be asserted in data or if controls are to be exercised, it is necessary for such rights or controls to be founded in one or more of the legal regimes that affect the generation, dissemination or use of data. Rules defining the generation, ownership and control of data may be based on legislative provisions, administrative requirements, common law principles or contractual obligations. For example, statutory provisions may permit or require data to be generated and the resulting datasets may give rise to intangible property rights in the form of copyright or patent. Rights to access and use the data may arise from generally applicable legislative provisions or be granted pursuant to a contractual arrangement between a data custodian or rights-holder and a user.

3.02 The importance of understanding the broader legal framework within which research is conducted and of ensuring that controls on data do not create regulatory impediments to data access and sharing was acknowledged by the PMSEIC Data Working Group in its report, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006).⁵⁶ The PMSEIC Data Working Group stated, in recommendation 9, that:

⁵⁵ See the opening quote for Chapter 3, p43

⁵⁶ Prime Minister's Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006) 15 <http://www.dest.gov.au/sectors/science_innovation/publications/resources/profiles/Presentation_Data_for_Science.htm> at 26 March 2007.

In the context of developing the strategic framework for scientific data management, Australia's intellectual property approaches [need to] be checked to ensure [that] they do not impede the sharing of data.⁵⁷

3.03 In Australia, controls on research data and information are most commonly imposed by government. A very high proportion of research funding is provided by the Commonwealth, State and Territory governments. It has recently been estimated that the Federal Government alone provides more than \$5 billion annually in funding for science and innovation.⁵⁸ In some fields (e.g. human health-related biotechnology), virtually all research carried out in Australia (whether in universities, research institutes, or government departments or agencies) is funded by the Australian Government. The State and Territory governments also invest heavily in research activities, either through their own departments or, as is increasingly the case, in conjunction with University researchers or institutes or private sector entities. It is estimated that in 2002–03, State and local government expenditure on research and development exceeded \$800 million.⁵⁹ Some of the largest data holders in Australia are public sector research agencies within the Commonwealth, State and Territory governments, including the CSIRO, the Bureau of Meteorology (BOM), and Geosciences Australia, as well as the numerous departments responsible for natural resources, primary industries, marine resources, health and the environment. As a result, government entities will often be involved as collaborators in research projects (as is typically the case with Cooperative Research Centres) or conditions will be attached to funding grants which require project participants to account in various ways (including financial) to the government funding body. Consequently, in developing an e-Research infrastructure for Australia which includes the research output of these significant contributors, it is necessary to appreciate the broader regulatory framework in which they operate, rather than focus on research which takes place entirely in an academic context.

3.04 This chapter provides an overview of the kinds of regulatory provisions that are relevant to the generation and handling of research data. Particular attention is given to legislative and administrative provisions applying to the collection, storage, maintenance and archiving of data and information in the public sector and measures regulating access to materials held by public sector entities. For example, public records legislation provides for the archiving of and access to public records such as Ministerial records, while freedom of information legislation regulates access to documents of government agencies and “official” documents. Some governments have established administrative policies and guidelines with which public authorities must comply in the responsible collection and management of data and information (for example, Queensland's Information Standards). This chapter also considers the operation of legislative protection of personal “data protection” privacy, which is of particular relevance to the activities of research organisations that collect or use personally identifying information and health information from and about human subjects. The relevance of cultural protocols which may affect how some data,

⁵⁷ Ibid.

⁵⁸ Productivity Commission, *Terms of Reference for Economic, Social and Environmental Returns on Public Support for Science and Innovation in Australia* (2006) <<http://www.pc.gov.au/study/science/tor.html>>.

⁵⁹ Productivity Commission, *Public Support for Science and Innovation*, Draft Research Report Overview, (2006) xx <<http://www.pc.gov.au/study/science/draftreport/overview.pdf>>.

including social science and anthropological data and information about biological resources derived from traditional knowledge, is collected, managed and made accessible to the wider public is also considered. Clearly, not all of these regulatory controls applying to data collection and use will be relevant to every research project, although it is not uncommon for a research project to be affected by a range of legislative and administrative requirements.

2. Health Information and Personally Identifying Information – Legal Frameworks for Information Privacy

3.05 Much of the discussion in Australia and internationally about legal issues surrounding the development of strategic frameworks for access to and reuse of research data evidences a wide-ranging appreciation of the relevance of laws governing the collection, handling and use of personally identifying and health information.⁶⁰ In the report, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006), the PMSEIC Data Working Group acknowledged the need for funding agencies such as ARC and NHMRC to develop, and ensure compliance with, best practice guidelines and policies so that bona fide researchers are able to:

Access individual population data, including the integration and linking of data from multiple sources, whilst protecting privacy, and ensuring that ethics committees fully understand these policies and their rationale.⁶¹

Similarly, Roger Magnusson writes:

In the medical sphere...subject to maintaining patient anonymity, it seems reasonable for the law to foster the *utility* of information, including the use of patient data in medical research.⁶²

The Australian Privacy Commissioner, Karen Curtis, in a submission to the Australian Law Reform Commission's current review of privacy law stated:

The Office acknowledges the clear benefits that health research delivers to the community.⁶³

⁶⁰ See, for example, Australian Government Office of the Privacy Commissioner, Submission to the Australian Law Reform Commission's Review of Privacy – Issues Paper 31, February 2007; Charles Oppenheim, 'Data Protection' in *The Legal and Regulatory Environment for Electronic Information*, (4th ed, 2001) 161 – 184; Prime Minister's Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006) 15

<http://www.dest.gov.au/sectors/science_innovation/publications_resources/profiles/Presentation_Data_for_Science.htm> at 26 March 2007.

⁶¹ Prime Minister's Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (2006) 12 – Recommendation 8

<http://www.dest.gov.au/sectors/science_innovation/publications_resources/profiles/Presentation_Data_for_Science.htm> at 26 March 2007.

⁶² R. Magnusson, 'Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia's Health Information System' (2002) 24 *Sydney Law Review* 5, 21.

⁶³ Australian Government Office of the Privacy Commissioner Submission to the Australian Law Reform Commission's Review of Privacy – Issues Paper 31, February 2007, 349

Privacy Act 1988

3.06 In Australia, the principal legislation governing privacy of personal information is the *Privacy Act 1988* (Cth). The *Privacy Act 1988* originally applied only to acts done by Commonwealth government departments and agencies. It was amended in 2000 to extend to the private sector. The *Privacy Act 1988* applies to Commonwealth public sector entities; it does not apply to state public sector entities. Therefore, its operation may not extend to public teaching hospitals and associated research bodies, where such bodies are established for a public purpose under a law of the state.⁶⁴

3.07 The *Privacy Act 1988* contains Information Privacy Principles, which operate to prevent the collection of personal information by a government agency except where the collection is for a lawful purpose directly related to a function or activity of the agency.⁶⁵ “Personal information” is defined in the *Privacy Act 1988* as:

Information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion.⁶⁶

3.08 Where personal information is collected, it must be stored so that it is protected by reasonable security safeguards.⁶⁷ The Information Privacy Principles then limit the extent to which the personal information can be accessed, used and disclosed.⁶⁸

3.09 In the 2000 amendments to the *Privacy Act 1988*, which came into effect in December 2001, an additional set of privacy principles – the National Privacy Principles – applying to private sector organisations were inserted into the Act. The *Privacy Act 1988* defines “organisation” as:

- (a) an individual; or
- (b) a body corporate; or
- (c) a partnership; or
- (d) any other unincorporated association; or
- (e) a trust; that is not a small business operator, a registered political party, an agency, a State or Territory authority or a prescribed instrumentality of a State or Territory.⁶⁹

3.10 The National Privacy Principles are based on the Information Privacy Principles, and provide that personal information cannot be used except for the lawful purpose for which it was collected (subject to some exceptions).⁷⁰ In particular, the National Privacy Principles set out how private sector organisations should:

collect, use and disclose personal information, maintain data quality, keep personal information secure, maintain openness, allow for access and correction of person information,

⁶⁴ Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.220]

⁶⁵ *Privacy Act 1988* (Cth) s 14, Information Privacy Principle 1.

⁶⁶ *Privacy Act 1988* (Cth) s 6.

⁶⁷ *Privacy Act 1988* (Cth) s 14, Information Privacy Principle 4.

⁶⁸ See *Privacy Act 1988* (Cth) s 14, Information Privacy Principles 6, 9, 10 and 11.

⁶⁹ *Privacy Act 1988* (Cth) s 6C.

⁷⁰ See *Privacy Act 1988* (Cth) sch 3.

use identifiers, allow anonymity, conduct trans-border data flows and collect sensitive information.⁷¹

3.11 The National Privacy Principles draw a distinction between “personal information” and “sensitive information”. The latter is given a higher level of protection and is defined to include health information about an individual. As such, the *Privacy Act 1988* may have particular relevance to data collections for health and medical purposes.⁷²

3.12 National Privacy Principle 10 provides that generally an organisation must not collect sensitive information (including genetic and health information) about an individual unless the individual has consented. Obtaining consent for research purposes should involve explaining to participants the purpose, methods and risks of the research process, the possible outcomes of the research and the likelihood of publication of research results.⁷³ It is also helpful to remember that under the National Privacy Principles, personal information can be used for the purpose for which it is collected, and under the Information Privacy Principles, personal information can be used for a lawful purpose directly related to a function or activity of the government agency. Where the information is collected for the purpose of inclusion in a database or for use in a health or medical research project, the individuals who participate in the project will be held to have consented to the use and disclosure of the information for those purposes.

3.13 There are, however, some exceptions to the requirement that consent be obtained. These exceptions apply both to the collection and to the disclosure of personal, sensitive and health information.

3.14 Under National Privacy Principle 10.3, health information can be collected without consent in limited circumstances for research relevant to public health or public safety, and for compilation or analysis of statistics relevant to public health or public safety.

3.15 An agency will not need to obtain consent for disclosure of an individual’s personal information under the Information Privacy Principles if the individual was reasonably likely to have been aware that such disclosures are usually made.⁷⁴ An organisation will not need to obtain consent for disclosure under the National Privacy Principles if the information is disclosed for the purpose it was collected or a directly related purpose and the individual would reasonably expect the organisation to disclose health information in that way.⁷⁵

⁷¹ Australian Law Reform Commission, *Protection of Human Genetic Information* (2001) Issues Paper 26, Ch 4 Privacy of genetic information
<http://www.austlii.edu.au/au/other/alrc/publications/issues/26/CH04_Privacy.html> at 12 March 2007.

⁷² For further discussion on Australian privacy legislation as it relates to health data and information, see R Magnusson, ‘Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia’s Health Information System’ (2002) 24 *Sydney Law Review* 5, 7, 23-4, 30, 35-7.

⁷³ See generally Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.212].

⁷⁴ Australian Law Reform Commission *Review of Privacy* Issues Paper 31 (2006) [8.124].

⁷⁵ *Ibid.*

3.16 Consent may also not need to be obtained where it is impractical to do so and where the collection, use and disclosure of the personal information is carried out in accordance with guidelines made under s 95 or s 95A *Privacy Act 1988* (Cth). Section 95 applies to public sector agencies and s 95A applies to private sector organisations.

Guidelines under Section 95A

3.17 Section 95A *Privacy Act 1988* provides:

(1) This section allows the Commissioner to approve for the purposes of the National Privacy Principles (the NPPs) guidelines that are issued by the CEO of the National Health and Medical Research Council or a prescribed authority.

(2) For the purposes of subparagraph 2.1(d)(ii) of the NPPs, the Commissioner may, by notice in the *Gazette*, approve guidelines that relate to the use and disclosure of health information for the purpose of research, or the compilation or analysis of statistics, relevant to public health or public safety.

(3) The Commissioner may give an approval under subsection (2) only if satisfied that the public interest in the use and disclosure of health information for the purposes mentioned in that subsection in accordance with the guidelines substantially outweighs the public interest in maintaining the level of privacy protection afforded by the NPPs (other than paragraph 2.1(d)).

3.18 In 2001, the Privacy Commissioner approved the *Guidelines on Privacy in the Private Health Sector*⁷⁶ (the Guidelines) under s95A. The Guidelines recommend that where consent is required from individuals for the collection and inclusion of data in a health or genetic register or database, participants should be given the opportunity to opt out of inclusion in the database.⁷⁷ National Privacy Principle 10 also provides that where health information is collected for research purposes (such as inclusion in a health or genetic database), an organisation must take reasonable steps to permanently de-identify the information before the organisation discloses it.

3.19 In 2007, the Privacy Commissioner considered:

In regard to health and medical research, the Office submits that the existing regulatory framework affords individuals with an appropriate degree of assurance that their personal health information will not be misused, particularly where it is handled without their consent. The Office draws attention to provisions where regulatory complexity could be reduced, particularly by harmonizing the enabling provisions for the section 95 and 95A mechanisms.⁷⁸

De-identifying Personal Information

3.20 With regard to identifying and de-identifying personal information, the Australian Law Reform Commission (ALRC), in its *Issues Paper 31 - Review of Privacy* (2006), considers that data or personal information can take three forms:

⁷⁶ Office of the Federal Privacy Commissioner (2001) *Guidelines on Privacy in the Private Health Sector*.

⁷⁷ Australian Law Reform Commission, *Protection of Human Genetic Information*, Issues Paper 26 (2001) Ch 4 Privacy of genetic information

<http://www.austlii.edu.au/au/other/alrc/publications/issues/26/CH04_Privacy.html> at 12 March 2007

⁷⁸ Australian Government Office of the Privacy Commissioner, Submission to the Australian Law Reform Commission's Review of Privacy – Issues Paper 31, February 2007, 278

- individually identifiable: data from which the identity of a specific individual can reasonably be ascertained. Examples of identifiers may include the individual's name, image, date of birth or address;
- re-identifiable or potentially re-identifiable: data from which identifiers have been removed and replaced by a code, but from which it remains possible to re-identify a specific individual, for example, by using the code or by linking different data sets; and
- non-identifiable: data that have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. This includes a subset - anonymous: data which can be linked with other data so it can be known that they are about the same data subject, while the identity of that specific individual remains unknown.⁷⁹

3.21 Research undertaken by the Australian Government Department of Health and Ageing suggests that although individuals express strong reservations about identifiable personal information being disclosed, they are generally very accepting of researchers sharing de-identified health information. Individuals also support the idea that researchers should be able to access health information from databases as long as the health information is identified by a unique number rather than a name.⁸⁰

3.22 There is a requirement in the Guidelines that health information must not be used or disclosed for research purposes unless the public interest in the use and disclosure of the information substantially outweighs the public interest in maintaining privacy protection. Presumably, this requirement will primarily apply to identified, identifiable or re-identifiable information (and not non-identifiable data). In determining which public interest carries the greater weight in any given situation, it will be necessary to consider both the type of information and the purpose and value of the research. Some factors to consider would include:

- the extent of the breach of privacy and the risk of harm to the individuals concerned;
- whether the research design can be modified and the financial implications of this;
- the value and public importance of the research and the likely benefits to participants; and
- any detrimental effects of not proceeding with the research.⁸¹

Responses to Privacy Act 1988 (Cth)

3.23 Some researchers have expressed frustration with the limits imposed by existing privacy legislation. For example, Dr Richie Gun of the Department of Public Health, University of Adelaide has commented:

⁷⁹ Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31(2006) [8.254]. (For more on this, see Australian Government Office of the Privacy Commissioner, Submission to the Australian Law Reform Commission's Review of Privacy – Issues Paper 31, February 2007, 346-349)

⁸⁰ Ibid [8.237].

⁸¹ See Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.229].

In Australia we are now in a uniquely advantageous position to carry out such [cancer] research, as we have mandatory registration of cancers in every State and Territory. We therefore have almost complete enumeration of all invasive cancers occurring in Australia, with the potential to carry out epidemiological studies on cancer incidence equal to or better than anywhere else in the world. Unfortunately privacy laws are impeding access to cancer registry data, so that it is becoming increasingly hard to carry out the linkage of cancer registrations with exposure data.

Rulings such as this suggest that we researchers are not to be trusted to protect privacy; that names will be released to outside parties; or that publications will identify individuals. This might be justified if there were some evidence that researchers have actually misused such data. Yet where is such evidence? The fact that there is no evidence of misuse is easily explained: researchers have nothing to gain by providing information and everything to lose. I know that if it became known that confidential information had been given out from my research team, it would be the end of my research and my career.⁸²

3.24 Similarly, the NHMRC has acknowledged the difficulties created by the *Privacy Act 1988* (Cth). It notes that the use or disclosure of health information in a register or database without consent is unlikely to comply with the National Privacy Principles and that attempts to obtain the consent of all individuals for the inclusion of their personal information in a register is likely to be impracticable in terms of both time and expenditure.⁸³ These difficulties will continue notwithstanding that:

[t]he distinction between clinical records and data registers is likely to diminish as health records gradually *become* databases.⁸⁴

3.25 For health information to be included in a register or database, the approval of the relevant Human Research Ethics Committee (HREC) will be required. The ALRC comments that:

[i]t would be extremely difficult, however, for an HREC to decide where the balance of interests lay in relation to an individual register, in the absence of specific information about the proposed future use of the register.⁸⁵

Data Linkage

3.26 Data registers and repositories raise the issue of data linkage. Data linkage - also called record linkage - can be described as:

[t]he process of linking or matching the records contained in two (or more) databases or data collections which were not originally designed to be combined. In the absence of unique identifiers (such as a number) which are common to both data collections, a series of non-unique partial identifiers such as name or initials, date of birth, sex, residential address or postcode or country of birth are usually used to link the records, using probabilistic or 'fuzzy' matching techniques. The purpose of doing this is to yield more information than is available

⁸² University of Adelaide, *Submission to the Office of the Privacy Commissioner Review of the Private Sector Provisions of the Privacy Act 1988* (2004), cited in Australian Law Reform Commission *Review of Privacy*, Issues Paper 31 (2006) [8.230]-[8.231].

⁸³ Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.282]-[8.283].

⁸⁴ R. Magnusson, 'Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia's Health Information System' (2002) 24 *Sydney Law Review* 5, 15.

⁸⁵ Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.283].

from the two databases by themselves, often at far less cost than would be required to establish a new data collection to collect the same information.⁸⁶

3.27 The ALRC has also commented that:

[i]dentifying and investigating the relationships between risk factors and disease frequently require researchers to match accurately data relating to the same individual.⁸⁷

3.28 It is the “fuzzy” matching techniques in data linkage, using partially identifying or identifiable information, which raise privacy concerns. The NHMRC has observed that some HRECs appear to automatically reject research proposals involving data linkage of health information without consent, because they mistakenly believe that this is not ethically or legally acceptable.⁸⁸ In reality, however, privacy concerns are to some degree unwarranted:

To researchers, the identity of the individual is irrelevant, except as a means of linking one set of information with another in order to identify factors influencing health outcomes.⁸⁹

3.29 The NHMRC’s *National Statement on Ethical Conduct in Research Involving Humans* recommends that HRECs authorise the transitory use of patient identifiers for the purposes of data linkage, even without patient consent, provided that:

- the personal information enabling linkage is not retained after the linkage;
- the identifying information is used with sufficient security; and
- the research for which the data is being linked has public benefit.⁹⁰

3.30 The Australian Privacy Commissioner has recognised the potential benefits of data linkage, shared electronic health record systems and health data registers, especially for medical research.⁹¹ In a submission to the ALRC’s review of privacy laws in 2007, the Privacy Commissioner suggested that consideration be given to introducing specific legislative provisions for establishing health data registers so that

⁸⁶ Australian Institute of Health and Welfare and the National Public Health Information Working Group, *National Public Health Information Development Plan – Directions and Recommendations 1999* 15 <<http://www.aihw.gov.au/publications/health/nphidp99/nphidp99.pdf>> at 9 May 2007.

⁸⁷ Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.285].

⁸⁸ Ibid.

⁸⁹ Beverly Sibthorpe, Erich Kliever and Len Smith, ‘Record Linkage in Australian Epidemiological Research: Health Benefits, Privacy Safeguards and Future Potential’ (1995) 19 *Australian Journal of Public Health* 250, 253, cited in R Magnusson, ‘Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia’s Health Information System’ (2002) 24 *Sydney Law Review* 5, 46.

⁹⁰ National Health and Medical Research Council (NHMRC), *National Statement on Ethical Conduct in Research Involving Humans* (1999) <http://www.nhmrc.gov.au/publications/synopses/_files/e35.pdf> cited in R Magnusson, ‘Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia’s Health Information System’ (2002) 24 *Sydney Law Review* 5, 27.

⁹¹ See Australian Government Office of the Privacy Commissioner, Submission to the Australian Law Reform Commission’s Review of Privacy – Issues Paper 31, February 2007, 286, 360-366. For more on consent see p349-356.

they are brought within the “required or authorised by law” exceptions contained in NPP 10.⁹²

Recommendations

3.31 The Privacy Commissioner has made various recommendations about the *Privacy Act*. Recommendation 60 was most relevant, being:

As part of a broader inquiry into the Privacy Act, the Australian Government should consider:

- How to achieve greater consistency in regulating research activities under the Privacy Act
- Whether regulatory reform is needed to address the issue of de-identification in the context of research and the handling of health information
- Where the balance lies between the public interest in comprehensive research that provides overall benefits to the community, and the public interest in protecting individuals’ privacy (including individuals having choices about the use of their information for such research purposes)
- Whether there is a need to amend [National Privacy Principle] 2 to permit the use and disclosure of personal information for research that does not involve health information
- Undertaking further research and education work with the broader community to ensure that the balance between research and privacy accords with what the community expects and understands.

The Office will work with the NHMRC to simplify the reporting process for human research ethics committees under the section 95A guidelines.⁹³

Draft National Health Privacy Code

3.32 In 2003, the Australian Health Ministers’ Advisory Council (AHMAC) National Health Privacy Working Group developed the draft *National Health Privacy Code*.⁹⁴ The draft Code is expressed to apply to “every organisation that is a health service provider or collects, holds or uses health information.” It contains 11 National Health Privacy Principles, which regulate the collection and handling of health information so that a balance is achieved between the public interest in protecting privacy and the public interest in legitimate use of relevant information. The Privacy Commissioner has recommended that the draft Code be adopted as a schedule to the *Privacy Act 1988* (Cth), but to date this recommendation has not been implemented.

National Health Act 1953 (Cth)

3.33 Section 135AA *National Health Act 1953* (Cth) deals specifically with personal information held in the Medicare and Pharmaceutical Benefits databases. It requires the Privacy Commissioner to issue written guidelines covering the storage, disclosure and use of the information. In November 2004, the guidelines issued under this section by the Privacy Commissioner in 1993 and amended in 2000 were reviewed. The review considered the use of health information for research and concluded that the guidelines should be amended to allow an individual to permit their

⁹² Ibid, 366

⁹³ Ibid 780.

⁹⁴ National Health Privacy Working Group of the Australian Health Ministers’ Advisory Council, *Draft National Health Privacy Code* (2003); Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.38]-[8.43], [8.223]-[8.225].

claims information held in the Medicare database to be linked to for research purposes.⁹⁵

State Legislation

3.34 Some Australian States and Territories have independently enacted legislation to protect the privacy of health information specifically – the Australian Capital Territory has enacted the *Health Records (Privacy and Access) Act 1997*, Victoria has enacted the *Health Records Act 2001*, which provides a separate set of Health Privacy Principles, and New South Wales has enacted the *Health Records and Information Privacy Act 2002*, which has some overlap with the *Privacy and Personal Information Protection Act 1998*. In Queensland, there are administrative rather than legislative protocols, such as Information Standards 42 and 42A (see below).

State and Territory Privacy Legislation:⁹⁶

New South Wales

- *Privacy and Personal Information Act 1998* (NSW) – contains “Information Protection Principles” which apply to the collection and use of personal information in the public sector. The Act establishes the Office of the NSW Privacy Commissioner to investigate privacy complaints relating to the public sector handling of personal information.⁹⁷
- *Health and Records and Information Privacy Act 2002* (NSW) – contains “Health Privacy Principles” which state how health information is to be collected, stored, accessed, used and disclosed. The Act also covers issues of identification and anonymity, data transfers and linkage.⁹⁸

Victoria

- *Information Privacy Act 2000* (Vic) – contains “Information Privacy Principles” which apply to the handling of information by the State public sector or other bodies declared as organisations under the Act (not private sector organisations). The Act also establishes the Office of the Victorian Privacy Commissioner for the investigation of privacy complaints.⁹⁹
- *Health Records Act 2001* (Vic) – contains “Health Privacy Principles” for the collection, access and use of health information held by public sector health service providers.¹⁰⁰

⁹⁵ See Australian Law Reform Commission, *Review of Privacy*, Issues Paper 31 (2006) [8.72]-[8.80].

⁹⁶ For more information, see Australian Law Reform Commission Issues Paper 31 *Review of Privacy* (2006)[2.22] – [2.81]

⁹⁷ See for example, s 8 to 28, 29, 33 and 36 *Privacy and Personal Information Act 1998* (NSW)

⁹⁸ See for example, the Health Privacy Principles in Schedule 1 *Health and Records and Information Privacy Act 2002* (NSW)

⁹⁹ See for example, s 1, s9, s50 and Schedule 1 *Information Privacy Act 2000* (VIC)

¹⁰⁰ See for example, s1, s19 and Schedule 1 *Health Records Act 2001* (VIC)

Australian Capital Territory

- *Australian Capital Territory Government Service (Consequential Provisions) Act 1994* (Cth) – The Australian Capital Territory public sector complies with this amended version of the *Privacy Act* (Cth), administered by the Commonwealth Office of the Privacy Commissioner.
- *Health Records (Privacy and Access) Act 1997* (ACT) – contains “Privacy Principles” for the handling of and access to health information in the public sector. The Act is administered by the ACT Community and Health Services Complaints Commissioner.¹⁰¹

Northern Territory

- *Information Act 2002* (NT) – contains “Information Privacy Principles” based on the Commonwealth National Privacy Principles and allows for the development of approved codes of practice for the handling, protection, collection and use of personal information by public sector organisations. The Act empowers an Information Commissioner for the Northern Territory to investigate complaints relating to public sector handling of personal information.¹⁰²
- The Northern Territory has also enacted the *Code of Health Rights and Responsibilities* which contains principles relating to access, privacy, and confidentiality of health information. Complaints are heard by the Northern Territory Health and Community Services Complaints Commission under the *Health and Community Services Complaints Act 1998* (NT).¹⁰³

Tasmania

- *Personal Information Protection Act 2004* (Tas) – contains “Personal Information Protection Principles” for the collection, use and disclosure of personal information by government agencies known as “personal information custodians”. Under the Act, the Tasmanian Ombudsman investigates privacy complaints relating to public sector handling of personal information.¹⁰⁴
- *Health Complaints Act 1995* (Tas) – establishes a “Charter of Rights” which provides for confidentiality of health information.¹⁰⁵

¹⁰¹ See for example, s3, s5 and

¹⁰² See for example, s3, s10 to14, s65, s72, s85 and Schedule 2 *Information Act 2002* (NT)

¹⁰³ See for example, Principles 1 to 8 http://www.nt.gov.au/omb_hcscc/hcscc/code.htm (10May 2007) and s3 and s12 *Health and Community Services Complaints Act 1998* (NT)

¹⁰⁴ See for example, s 16 and Schedule 1 *Personal Information Protection Act 2004* (Tas)

¹⁰⁵ See for example s17 *Health Complaints Act 1995* (Tas) and

¹⁰⁶ See for example, Principles 4 to 8 PC012 - Information Privacy Principles Instruction at < PC012 - Information Privacy Principles Instruction> (9 May 2007)

¹⁰⁷ See for example, s 3, s 15, s 18 and Schedules 3 and 4 *Information Privacy Bill 2007* (WA) at < <http://www.parliament.wa.gov.au/web/newwebparl.nsf/iframewebpages/Bills+-+Current>> (9 May 2007).

¹⁰⁸ See for example, s 49 *State Records Act 2000* (WA)

¹⁰⁹ See for example, s 3 and Glossary *Freedom of Information Act 1992* (WA)

South Australia

Currently there is no legislation in South Australia that deals with the collection and use of private information. However, the State does have a Privacy Committee of South Australia which regulates the states administrative “PC012 - Information Privacy Principles Instruction” applying to public sector handling of personal information. Complaints regarding public sector handling of personal information are heard by the Committee with an avenue of further appeal to the State Ombudsman.¹⁰⁶

Western Australia

- *Information Privacy Bill 2007* (WA) – a Bill, introduced into parliament in March 2007, has reached the second reading stage before the State’s Legislative Assembly. The Bill seeks to establish “Information Privacy Principles” for the handling of personal information by the public sector and “Health Privacy Principles” for the handling of health information by the public and private sectors. It also seeks to establish an Office of the Western Australian Privacy Commissioner.¹⁰⁷
- *State Records Act 2000* (WA) – provides that access to medical or disability information in the State archives is prohibited unless consent is first obtained from the human subject or the information is provided in a way that does not disclose the identity of the subject.¹⁰⁸
- *Freedom of Information Act 1992* (WA) – allows for access to and amendment of personal information in an inaccurate document and extends to identification items such as fingerprints, retina prints or body samples.¹⁰⁹

3.35 The differences between privacy regulation at Commonwealth and State levels has caused some confusion for medical researchers. The Australian Government Productivity Commission in its report, *Public Support for Science and Innovation*, released March 2007, concluded that the complexity caused by the intersection of Australian government and State and Territory privacy laws had adverse effects on medical research.¹¹⁰ The NHMRC has expressed the same view.¹¹¹

3.36 The Productivity Commission considered that the adverse effects of privacy regulation on medical research included:

- substantially increasing the administrative burden on researchers who would have to decide whether data could be used or disclosed for research;

¹¹⁰ Productivity Commission, *Public Support for Science and Innovation*, Research Report, (2007) 189.

¹¹¹ Ibid 217.

- increasing the administrative costs associated with complying with privacy legislation, thereby reducing the amount of public funds directed to actual research; and
- restricting researchers from data linking or data matching with other research projects.¹¹²

3.37 In responding to the concerns expressed by the Productivity Commission, the NHMRC and other organisations, the Office of the Privacy Commissioner focused on studies showing that many individuals are sensitive about their health information being provided or used by organisations conducting research.¹¹³ The Privacy Commissioner considered that:

Strong privacy provisions are essential for sustaining the community confidence needed to make medical research viable.¹¹⁴

3.38 However, the Productivity Commission and the NHMRC still recommended that in order to reduce the complexity of existing privacy legislation, there should be national consistency in privacy regulation of health information.¹¹⁵

Genetic Health Information

3.39 The ALRC *Issues Paper 26: Protection of Human Genetic Information*, examined whether genetic information is more sensitive than other health information so as to require special privacy protection. The Issues Paper identified that this would be relevant to the creation and use of human genetic databases for clinical and research purposes. In particular:

Moves towards further regulation of health information privacy have been given momentum by moves towards the establishment of electronic health records systems. At federal level, the proposals include *HealthConnect*, a proposal for an Australia-wide network for exchanging health information online and the Better Medication Management System.¹¹⁶

3.40 The Issues Paper considered reforms proposed in 1998 by the Genetic Privacy and Non-Discrimination Bill 1998 (Cth) to tighten privacy controls over genetic information. It also considered the conclusions of the Senate Legal and Constitutional Legislation Committee that it was preferable to amend existing privacy and discrimination legislation to address issues raised by genetic technology, rather than create new legislation specific to genetics.¹¹⁷ The Paper also stated that there

¹¹² Ibid 217-18.

¹¹³ Ibid 219.

¹¹⁴ Ibid.

¹¹⁵ Ibid 189, see also 217, 219, 221.

¹¹⁶ Australian Law Reform Commission, *Protection of Human Genetic Information*, Issues Paper 26 (2001) Ch 4 Privacy of genetic information

<http://www.austlii.edu.au/au/other/alrc/publications/issues/26/CH04_Privacy.html>

at 12 March 2007; see also *HealthConnect*

<<http://www.health.gov.au/internet/hconnect/publishing.nsf/Content/home>> at 12 March 2007.

¹¹⁷ Australian Law Reform Commission, *Protection of Human Genetic Information*, Issues Paper 26 (2001) Ch 4 Privacy of genetic information

<http://www.austlii.edu.au/au/other/alrc/publications/issues/26/CH04_Privacy.html>

at 12 March 2007.

were no jurisdictions with legislation dealing specifically with protection of personal genetic information, except the Netherlands *Personal Data Protection Act 2000*.¹¹⁸ Currently in Australia, issues of privacy for genetic information are still dealt with under the *Privacy Act 1988* and the associated privacy principles.

3. Government Data and Information – Archiving, Accountability and Accessibility

Public Records (or Archives) Legislation and Freedom of Information Legislation

3.41 All Australian jurisdictions have enacted legislation to regulate the management of government records – in the form of public (or State) records or archives Acts – as well obtaining access to public records that are not otherwise openly available – in the form of freedom of information Acts. The public records legislation provides for public access to records as well as setting out restrictions on access to certain records while the freedom of information legislation enables the public to access information held by governments, subject to certain exceptions.¹¹⁹ The legislation currently in force is as follows:

- Commonwealth: *Archives Act 1983* (Cth)¹²⁰ and *Freedom of Information Act 1982* (Cth);
- Tasmania: *Archives Act 1983* (Tas) and *Freedom of Information Act 1991* (Tas);
- Victoria: *Public Records Act 1973* (Vic) and *Freedom of Information Act 1982* (Vic);
- South Australia: *State Records Act 1997* (SA) and *Freedom of Information Act 1991* (SA);
- Queensland: *Public Records Act 2002* (Qld) and *Freedom of Information Act 1992* (Qld);
- NSW: *State Records Act 1998* (NSW) and *Freedom of Information Act 1989* (NSW);
- WA: *State Records Act 2000* (WA) and *Freedom of Information Act 1992* (WA); and
- Northern Territory: *Information Act 2002* (NT); and

¹¹⁸ Ibid.

¹¹⁹ For example documents affecting personal privacy of third parties will usually be exempt from the access requirements or will only be released after a consultation process.

¹²⁰ *Archives Act 1983* (Cth) was reviewed by the Australian Law Reform Commission, *Australia's Federal Record: A Review of Archives Act 1983*, Report No 85 (1998)

<<http://www.austlii.edu.au/au/other/alrc/publications/reports/85/toc.html>> at 21 May 2007.

- ACT: *Territory Records Act 2002* (ACT) and *Freedom of Information Act 1989* (ACT).

3.42 It is necessary to consider the operation of this legislation in relation to research data primarily because of the fact that a significant proportion of research output in Australia is produced by government entities, for example State health, natural resources, environmental protection and primary industries departments. As well as conducting their own independent research, these public sector entities frequently engage in collaborative research projects with university and private sector partners through Centres of Excellence and Cooperative Research Centres. Materials produced by an extensive range of public sector entities, including government departments and statutory bodies, will be subject to the operation of this legislation. Consequently, rights to control and obtain access to documents and records containing the research data outputs of public sector entities will be determined by the provisions of the relevant records and freedom of information legislation.

3.43 To illustrate the operation of these laws, the operation of the legislation in force in Queensland is considered. However, it should be noted that where a project involves contributions of research data by public sector entities in more than one jurisdiction, it will also be necessary to take into account the operation of the equivalent legislation applying in each of the jurisdictions that contribute their research data.

3.44 The purpose of the *Public Records Act 2002* (Qld) is to ensure that the State's public records are "made, managed, kept and, if appropriate, preserved in usable form for the benefit of present and future generations" and that public access to such records is available in accordance with the principles of the *Freedom of Information Act 1992* (Qld).¹²¹ The purpose of the *Freedom of Information Act 1992* is to:

Extend as far as possible the right of the community to have access to information held by the Queensland government.¹²²

3.45 The *Public Records Act 2002* (Qld) provides for the making, keeping and archiving of public records as well as access to public records. "Public record" is defined in s 6 to mean:

- records made for use by, or a purpose of a public authority;
- records received or kept by a public authority in the exercise of its statutory, administrative or other public responsibilities or for a related purpose; and
- Ministerial records.

3.46 A public authority is required to make and keep full and accurate records of its activities, having regard to any relevant policy, standards or guidelines made by the archivist about the making and keeping of public records.¹²³ Upon request, any person must be allowed access to a public record, unless the record is subject to a restricted access period (usually 30 years) because it is a Ministerial record or because it relates

¹²¹ *Public Records Act 2002* (Qld) s 3.

¹²² *Freedom of Information Act 1992* (Qld) s 4.

¹²³ s 7 *Public Records Act 2002* (Qld)

to Cabinet or Executive Council matter.¹²⁴ Additionally, access to a public record may be restricted in accordance with s 44 of the *Freedom of Information Act 1992* (Qld) if the record contains matter affecting the personal affairs of a person.

3.47 Where a government document is not reasonably available for public inspection under the *Public Records Act 2002* for a reason other than the imposition of a restricted access period, access may be sought under the *Freedom of Information Act 1992* (Qld).¹²⁵ The *Freedom of Information Act 1992* provides that a person has a legally enforceable right to be given access to the documents of an agency and the official documents of a Minister.¹²⁶ “Agency” is defined to mean a government department, local government or public authority.¹²⁷ This enforceable right of access is subject to the exceptions set out in ss 36 to 50, for example there is no right of access to Cabinet matter (s 36), matter relating to national or State security (s 42A), or matter affecting an individual’s personal affairs (s 44).

3.48 Importantly, the broad range of entities to which the *Public Records Act 2002* applies means that the Act will be relevant to the records of bodies which would not usually be considered to be part of the government of the State. Schedule 2 of the *Public Records Act 2002* defines “public authority” to include “an entity, other than the parliamentary service, that...is established by an Act”. As Queensland universities are established under State legislation,¹²⁸ they would appear to fall within the definition of “public authority” and thus be subject to the public records legislation. Consequently, data and records of university research projects could be subject to the record-keeping and access requirements under the *Public Records Act 2002*, not only in research projects involving collaboration between university research academics and government departments but also in research projects carried out entirely by staff or student researchers within academic institutions.

Electronic Transactions Legislation

3.49 The *Electronic Transactions Act 1999* (Cth) and the corresponding Electronic Transactions Acts enacted by the States and Territories¹²⁹ provide that if a person is required by law to record information in writing or to retain a document in the form of a paper, an article or other material, that requirement can be met by recording the information or keeping the document in electronic form.¹³⁰ This provision is relevant where records that are required to be kept under the public records legislation¹³¹ are kept in electronic form in a digital database, rather than stored in physical archives.

¹²⁴ *Public Records Act 2002* (Qld) ss 16, 17, 18.

¹²⁵ *Freedom of Information Act 1992* (Qld) s 22(b).

¹²⁶ *Freedom of Information Act 1992* (Qld) s 21.

¹²⁷ *Freedom of Information Act 1992* (Qld) s 8.

¹²⁸ For example, QUT is established under the *Queensland University of Technology Act 1998* (Qld).

¹²⁹ The corresponding State and Territory Acts are: *Electronic Transactions (Queensland) Act 2001* (Qld); *Electronic Transactions Act 2001* (ACT); *Electronic Transactions Act 2000* (NSW); *Electronic Transactions (Northern Territory) Act 2000* (NT); *Electronic Transactions Act 2000* (SA); *Electronic Transactions Act 2000* (Tas); *Electronic Transactions (Victoria) Act 2000* (Vic); *Electronic Transactions Act 2003* (WA).

¹³⁰ *Electronic Transaction Act 1999* (Cth) s 12.

¹³¹ For example, under s 7 *Public Records Act 2002* (Qld)

3.50 However, the information or document can only be recorded or kept in electronic form if it is reasonable to expect that the information would be “readily accessible so as to be useable for subsequent reference.”¹³² This means that in making the electronic record or document, a person would need to use stable technology to ensure that the information is archived accurately and can be accessed in the future. The legislation does not give any guidance as to what would constitute stable technology or a suitable electronic form so as to satisfy the requirements of the *Electronic Transactions Act 1999*. Instead, it anticipates subsequent legislation or regulations that would clarify these technical standards. If regulations are made requiring a particular kind of data storage device to be used to record information electronically or retain an electronic form of a document, that requirement must be met.¹³³ However, to date, neither the Commonwealth nor any of the States or Territories has made regulations that require the use of a particular form of technology to record or store information or documents in electronic form.

Other Legislation

3.51 As well as general legislation which deals primarily with data collection, retention and control, there are numerous statutes and administrative directives affecting the collection, use and dissemination of data in specific areas of activity. Where there are governmental requirements for information to be generated and provided to a public sector body, the relevant legislation often contains provisions relating to how the information is to be collected and restrictions that may apply to the use and disclosure of such information. This will be particularly important where the information being collected is personal or private, such as health or financial information, and in areas of social service, especially where children are concerned.

3.52 Some examples of legislation dealing with information collection include:

- *Census and Statistics Act 1905* (Cth) - regulates how census information is to be collected, when the census information can and cannot be disclosed, and the publication of census results and statistics;¹³⁴
- *Taxation Administration Act 1953* (Cth) – provides that information about the affairs of a person can only be acquired for the purposes of the Act, determines how that information is to be collected, and prohibits the recording of tax file numbers except for specific purposes;¹³⁵
- *Statistical Returns Act 1896* (Qld) – governs the collection and dissemination of statistical information within the Queensland government;
- *Environmental Planning and Assessment Act 1979* (NSW) – deals with environmental protection, planning and development. It provides for a register of critical habitat to be kept by the Director-General of National

¹³² s 12(1) and (2) *Electronic Transactions Act 1999* (Cth)

¹³³ Ibid.

¹³⁴ See for example, *Census and Statistics Act 1905* (Cth) ss 10, 12, 19A. Note also the operation of ss 22B and 30A *Archives Act 1983* (Cth).

¹³⁵ See for example, *Taxation Administration Act 1953* (Cth) ss 3C, 8WB.

Parks and Wildlife and a register of applications for development consent to be kept by the council;¹³⁶

- *Fisheries Act 1988* (NT) – sets out the information to be included in a register of fishing licensees and the information to be included in a fisheries management plan;¹³⁷
- *Adoption Act 1984* (Vic) – requires that registers be kept recording the names of persons who have been declared fit and proper persons to adopt and the names of persons whose application for adoption have been refused, deferred or revoked.¹³⁸ It also requires records of adoptions to be kept and sets out the confidentiality and privacy restrictions applying to these records;¹³⁹ and
- *Disability Services Act 1993* (WA) – establishes a register of complaints made about the provision of disability services.¹⁴⁰

4. Data Quality and Control - Administrative Arrangements and Information Standards

3.53 To supplement legislative requirements relating to information collection and use, governments may establish information codes or standards which do not themselves carry the force of law but which offer further guidance into things such as the creation of public records by public authorities, and the responsible collection and management of personal information by government agencies. Taking Queensland as an example, there are Information Standards which contain principles mandated by legislative provisions in the *Public Records Act 2002* and the *Financial Management Standard 1997*.¹⁴¹ These Information Standards include:

- Information Standard 31: Retention and Disposal of Public Records – makes public authorities accountable for the creation, retention and appraisal of public records, and provides for the disposal of public records only in particular circumstances and with the requisite authority;
- Information Standard 33: Information Access and Pricing – provides that Government information must be made accessible, directly or indirectly, to citizens of Queensland and those doing business in Queensland at no more than the cost of provision, or where possible for free, unless statutory requirements vary the access and pricing arrangements;
- Information Standard 34: Metadata – aims to facilitate access to and interoperability of government information resources by implementing metadata schemes for the description and classification of information;

¹³⁶ See *Environmental Planning and Assessment Act 1979* (NSW) ss 5, 5B, 100.

¹³⁷ See for example, *Fisheries Act 1988* (NT) ss 9, 23, 25.

¹³⁸ See *Adoption Act 1984* (Vic) s 13A.

¹³⁹ See for example, *Adoption Act 1984* (Vic) ss 85, 88.

¹⁴⁰ See *Disability Services Act 1993* (WA) s 48.

¹⁴¹ See <http://www.qgcio.qld.gov.au/02_infostand/standards.htm>.

- Information Standard 40: Recordkeeping – helps public authorities comply with recordkeeping legislation and meet their obligations under the *Public Records Act 2002*;
- Information Standard 41: Managing Technology-Dependent Records – this complements the *Electronic Transactions (Queensland) Act 2001* by providing that technology-dependent records (usually electronic records, but also records stored on reel to reel film, microfiche etc) generated or received in the course of government business must be maintained and accessible for as long as they are required to meet legislative, accountability, business and cultural obligations;
- Information Standard 42: Information Privacy – establishes a framework for the responsible collection and handling of personal information in the Queensland Government public sector; and
- Information Standard 42A: Information Privacy for the Queensland Department of Health – guides the Queensland Department of Health in collecting and handling personal information in accordance with the National Privacy Principles set out in the *Privacy Act 1988*.

5. Protocols Relating to Culturally Sensitive Data

3.54 Cultural protocols may affect the way that data is managed and made accessible to the wider public. While this will be less relevant for purely mathematical or scientific data, it will be especially important in areas of anthropology and social sciences.¹⁴² For example, an anthropologist in the field might observe tales, songs, dances and rituals of a particular traditional group. These observations may be transposed into records including notes, articles, photographs, films and audiotapes. Some materials may in fact show secret rituals of the traditional community or other sensitive information that the community would not want openly available to the public. This becomes a problem particularly with film and audiotape where the identity of individuals or groups is so enmeshed in the data that it cannot be separated.¹⁴³ Whereas medical information and statistical information generated largely by the Government can be “anonymized” by removing personal information before it is released to the public, materials depicting cultural rituals will be almost impossible to anonymize.¹⁴⁴ Thus, care will need to be taken when archiving sensitive cultural data and restrictions may need to be imposed upon access to the data.¹⁴⁵

¹⁴² For more information, see S Carlson and B Anderson ‘What are data? The many kinds of data and their implications to data re-use’ (2007) 12(2) *Journal of Computer-Mediated Communication* article 15 <<http://jcmc.indiana.edu/vol12/issue2/carlson.html>> at 18 January 2007.

¹⁴³ S Carlson and B Anderson ‘What are data? The many kinds of data and their implications to data re-use’ (2007) 12(2) *Journal of Computer-Mediated Communication* article 15 <<http://jcmc.indiana.edu/vol12/issue2/carlson.html>> at 18 January 2007.

¹⁴⁴ S Carlson and B Anderson ‘What are data? The many kinds of data and their implications to data re-use’ (2007) 12(2) *Journal of Computer-Mediated Communication* article 15 <<http://jcmc.indiana.edu/vol12/issue2/carlson.html>> at 18 January 2007

¹⁴⁵ For an example of a database containing culturally sensitive information, see the PARADISEC database in Chapter 4

6. Understanding the Legal Framework – Practical Application to Data Management

3.55 The legislation and standards relating to data collection and management described in this chapter do not operate independently from each other. It is possible that a single dataset or database will be affected (in terms of how the contained data can be managed, controlled and made accessible) by several legislative and administrative frameworks, which may apply simultaneously or at different stages in the creation, use or dissemination of a dataset or database. Further complication is added by the fact that much of the legislation regulating the collection, use and dissemination of information is specific to the Commonwealth or to a particular State or Territory. Thus, it will be necessary to consider not only the type of information that will be collected or generated, but also where those activities occur, in order to determine which laws and administrative requirements apply.

3.56 The mix of Commonwealth, State and Territory laws may also apply differently to public sector entities than to private individuals or companies. For example, in some States and Territories there is currently no legislation covering the collection, storage and use of personal information by small private sector commercial entities (with a turnover of less than \$3 million annually), unless the information is health-related information (in which case, the Commonwealth *Privacy Act 1988* will apply). Obligations relating to the curation of data under the public records legislation will apply only to public sector entities.

3.57 In designing a system to manage or share research or academic data, it is necessary to understand this wider regulatory framework governing data as it forms the platform on which the access and sharing system is based. Technology has brought massive advantages through the ability to aggregate, access, sort through and use unprecedented amounts of data, but this has not occurred in an organisational or legal vacuum. Various pieces of legislation impose rights and obligations, and it is necessary to understand what they are, how they interact, and how they can be dealt with in order to achieve the desired outcomes.

3.58 The operation of the legal framework means that steps will have to be taken in the design of research project protocols to ensure that the data access and sharing system does not have to deal with data which cannot be included in the system because of statutory restrictions. In particular, personal or health information which is regulated by Commonwealth and State legislation would not normally be included unless it has been de-identified or has been collected in compliance with the statutory requirements so that it can be re-used.

3.59 In developing a data access and sharing system, the starting point will usually be that data inputs into the system have cleared all legislative and other controls on the use of the data. Apart from legislative controls, it is necessary to take into account the controls imposed by copyright, the common law and contract. For example, the owner of copyright in a dataset may have granted another party a licence to use the dataset in a specified location or for certain purposes, but imposed limitations on the extent to which the dataset can be used. In this case, it will be necessary to ascertain whether the contractual licence granted by the copyright owner is sufficiently broad to

permit the dataset to be re-used by incorporating it within a database. Copyright and contractual controls are considered in greater detail in Chapter 5 and Chapter 7.

3.60 A system of data management needs to be developed for ascertaining what data and datasets can be included in a database and what rights of access or use can be granted. Where data is subject to legal restrictions that impact on access or use, the data and corresponding restrictions need to be identified and, if necessary, the data should be excluded from the database. It is best practice to develop a qualifying process so that certain kinds or categories of data are not included in the database if their inclusion will breach legal obligations.

KEY POINTS

Frameworks for the management of research data are developed within the context of a complex regulatory environment.

In developing a data management framework and infrastructure, it is necessary to take into account all legislative and other legal requirements applying to the generation, management, dissemination and use of research data in that context.

Effective systems for the management of research data, that enable access to and reuse of the data by other researchers, must be developed against the background of all relevant regulatory mechanisms (legislative and administrative) impacting upon data generation, handling and use in that context.

In particular, it is necessary to understand, where applicable, the practical implications of:

- *privacy obligations, whether applying under legislation or administrative requirements (especially in relation to personal and health information);*
- *public records (or archives) legislation and freedom of information legislation;*
- *governmental administrative arrangements applying to public sector entities, such as the Queensland government's Information Standards relating to information access, privacy and metadata; and*
- *cultural protocols.*

These regulatory mechanisms do not operate independently of each other, and may overlap. Laws, regulations and administrative requirements may apply differently to private sector companies than to public sector entities.

CHAPTER 4 - CURRENT PRACTICES AND ATTITUDES TO DATA SHARING

*“Cyberinfrastructure enables the creation, use, reuse, combination, organisation, and sharing of knowledge within a virtually integrated environment of mixed resources. In contrast to the internet, an all-purpose globally accessible platform, cyberinfrastructure is constructed around specific projects, research communities, or unique resources. However, it is not limited by its origins. It can support technology transfer, inter-sector collaboration, public education, even commercial ventures. Full realisation and optimisation of cyberinfrastructure requires understanding its organisational, economic and legal context. ...”*¹⁴⁶

¹⁴⁶ “Designing Cyberinfrastructure for Collaboration and Innovation” conference, co-sponsored by the National Science Foundation, the University of Michigan, Science Commons, the Council on Competitiveness and the Committee for Economic Development, held at the National Academies, Washington DC, in January 2007, conference homepage at <<http://www.si.umich.edu/cyberinfrastructure/issues.htm>>, “Themes” pages at <<http://www.si.umich.edu/cyberinfrastructure/program.htm>>

CURRENT PRACTICES AND ATTITUDES TO DATA SHARING

Aims:

1. Examine and describe key examples of United States, European and Australian databases which have developed legal arrangements for providing access to research data, and analyse the various models of data ownership, control, access and use observed in these databases;
2. Make observations based upon the sample of databases examined, identifying legal issues requiring further attention and clarification;
3. Discuss the emergence of open licensing practices and their relevance to the patenting of inventions derived from research data; and
4. Consider the relevance of survey evidence of researchers' attitudes on accessibility and sharing of data to the development of systems for technical and legal management of research data.

1. The Growth of Domestic and International Database Sharing Practices

4.01 A major focus of this report is the development of a legal framework to facilitate access to and sharing of research data. The international nature of research collaboration and data sharing means that the development of frameworks for use in the Australian context should take into account and (as far as possible) be compatible with the technical and management practices that have been developed in other countries. Many of the issues relating to access and use of research data and databases that are currently being considered in Australia have been the subject of extensive review and discussion in recent years, particularly in the context of large scale collaborative programs based primarily in the United States and Europe.

4.02 The number of research databases is growing at an extremely rapid rate. A measure of this is found in the 2007 Nucleic Acids Research database which lists 968 molecular biology databases, representing an increase of 110 databases since the previous year.¹⁴⁷ There are numerous examples of research databases available on the internet as well as websites where collections of databases can be accessed.¹⁴⁸ The APSR, funded by DEST, provides a registry of databases and how to use them.¹⁴⁹

4.03 Large scale collaborative research projects that produce sequence data (such as mRNA, genomic DNA and protein sequences) have featured prominently in the

¹⁴⁷ Nodalpoint.org, January 2007 <http://www.nodalpoint.org/2007/01/05/nar_database_issue_2007>

¹⁴⁸ See for example the European Bioinformatics Institute (EBI) <<http://www.ebi.ac.uk/>> which provides access to several databases including: UniProt, UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, EMBL, GOA, InterPro, ChEBI Release and CluStr.

¹⁴⁹ See <<http://apsr.edu.au>>.

development of data sharing concepts and practices.¹⁵⁰ New communication tools for databases are developing, for example the Wiki for Professionals database which applies the Wikipedia philosophy to databases by allowing voluntary updates from the research community for interlinked data sharing.¹⁵¹

4.04 At a generalised level, it is possible to differentiate between data generated and used in what Reichmann and Uhler refer to as “big science” and “small science” research projects. “Big science” projects are publicly funded projects (for example, in physics, space and the earth sciences), the research outputs of which are often deposited into well-organised databases. “Small science” projects are non-public projects performed by autonomous researchers, the research outputs of which are exchanged on the basis of informal arrangements and exist in various states of accessibility. Reichmann and Uhler have observed that “big science” projects are more likely to have formal arrangements for data access than “small science” projects.¹⁵²

4.05 This chapter provides an overview and description of arrangements that have been developed to provide access to and facilitate the sharing of research data in Australia and other jurisdictions. It examines current practices as exemplified in general frameworks for data access and sharing as well as in subject-specific databases, especially collections of medical and genetic research data. Arrangements for ownership, control, access and use of data in a sample of web-accessible United States and European databases are described, as well those observable in a sample of the Australian databases listed in the APSR registry.

4.06 Evidence of the attitudes of researchers towards providing access to and sharing data is also considered, in order to identify factors that need to be taken into account in developing technical and legal systems to facilitate data access and sharing. The findings of surveys of researchers’ attitudes carried out by the Australian eResearch Sustainability Survey (undertaken by APSR) and the NCRIS Platforms for Collaboration of Data Management Survey are considered. A survey of attitudes to data sharing and current practices which is currently being undertaken by QUT is also described.

¹⁵⁰ For a history of developments relating to data sharing, see Robin Cook-Deegan and Tom Dedeurwaerdere, *The Science Commons in Life Science Research: Structure, Function and Value of Access to Genetic Diversity*, International Science Journal, Fall 2006, available at <http://www.spatial.maine.edu/icfs/Life%20Sciences%20.pdf>

¹⁵¹ See <<http://www.wikiprofessional.info>>. The project launched in March 2007. Other examples of wiki tools being utilised include protein information from Swiss-Prot and gene descriptions from Gene Ontology. It weaves together existing archives to create what - from the user’s point of view - appears to be a single database. For example, the page on the muscular-dystrophy protein dystrophin contains data from Swiss-Prot, links to disease information from the US National Library of Medicine and links to relevant publications in PubMed Central.

¹⁵² J Reichmann and P Uhler, ‘A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment’ (2003) 66 *Law and Contemporary Problems* <<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp.>>.

2. Key Examples of Data Sharing Frameworks

4.07 The following examples of legal frameworks for data sharing have been selected because they (typically) represent complex, high-budget projects. We have attempted to determine for each example:

- Ownership - who owns rights in relation to the relevant data or dataset;
- Control - any restrictions or conditions applying to the use of the data and any arrangements allowing a person other than the rights-owner to control the data; and
- Access and Use - the right of entry provided users seeking access to and use of the data.

Example 1 - The National Institutes of Health (NIH) Databases

4.08 As the world's largest funder of biomedical research, the National Institutes of Health (NIH) spends around US\$27 billion on internal and external research. It is a leading provider of databases (together with software and bioinformatics tools), as well as one of the largest users of databases, bioinformatics tools and other research tools.¹⁵³

4.09 Since 1996, NIH has required its institutes and funded researchers to share data in several areas, such as DNA sequences, mapping information and crystallographic coordinates.¹⁵⁴ Arrangements for ownership, use and control of two of NIH's major databases, GenBank and dbGaP are discussed below.

4.10 Typically, for NIH-funded community resource genomic projects, NIH requires research data to be made available to the public for all and any uses (whether for research or commercial use) without restriction. A community resource project is defined as one which is:

Specifically devised and implemented to create a set of data, reagents or other material whose primary utility will be as a resource for the broad scientific community.¹⁵⁵

4.11 NIH requires data generated in such projects to be deposited rapidly – often within 24 hours of generation or at latest soon after the data has been validated – in order to create prior art so that the filing of patent applications is precluded. In some cases, the filing of patents by recipients of funding grants is expressly prohibited. Where there is a possibility of “parasitic” patent claims by third parties who can access the research data,¹⁵⁶ NIH has required data users to enter into an online click-wrap agreement (“non-assert clauses”) in which data users agree that they will not do

¹⁵³ Both through its own NIH employees and recipients of NIH grants. See Claire Driscoll, ‘Community Resource Projects & Genomics: How does the NHGRI Ensure Open Access to Databases and Unrestricted Use of Data & Biological Materials?’ (Slide presentation presented at the NAS Toxicogenomics Workshop, 29 June 2006).

¹⁵⁴ NIH policies on data sharing are discussed further in Chapter 8.

¹⁵⁵ Claire Driscoll, ‘NIH data and resource sharing, data release and IP policies for genomics community resource projects’, *Expert Opin. Ther. Patents* (2005) 15(1), p2-3

¹⁵⁶ See the Patents section of this chapter

anything with the data or enforce rights in relation to data which will block future use of the original data by others. Project teams and data generators may be required to agree not to file patents but to instead publish interpretations of data, thereby creating prior art (for example, disease associations).

Example 2 - NIH: GenBank

4.12 GenBank is part of the International Nucleotide Sequence Database Collaboration (INSD), which is comprised of the DNA DataBank of Japan, the European Molecular Biology Laboratory, and GenBank at the National Center for Biotechnology Information (NCBI). These three research organisations exchange data on a daily basis.

4.13 Each GenBank entry includes a concise description of a DNA sequence, the scientific name of the source organism, and a table of features. GenBank and its collaborators receive DNA sequences produced in laboratories throughout the world from more than 100,000 distinct organisms. GenBank contains more than 40 billion DNA sequences and continues to grow at an exponential rate, doubling every 12 – 15 months.¹⁵⁷

4.14 GenBank is built by direct submissions from individual laboratories,¹⁵⁸ as well as from bulk submissions from large-scale DNA sequencing centres. As GenBank is not moderated, it is difficult to perform rigorous analyses of its contents, for example, some viruses are called by different names in different entries.

Ownership - GenBank

4.15 Although the NIH's Data Sharing Principles (see Chapter 8) intend and encourage contributions to NIH databases to be made freely available for use by other researchers, it is apparent from the database website that some contributors to GenBank retain ownership rights in relation to their data.

Control - GenBank

4.16 Control of data held in the GenBank database is exercised by the owners of rights in relation to the data. This is consistent with the statement on the GenBank website that there are no restrictions on the use that a person may make of the data, subject to the ownership rights in relation to that data.

Access and Use - GenBank

4.17 It is not clear if there are legal agreements between INSD and its member organisations, including GenBank, dealing with obligations in relation to the exchange of DNA sequence data and the usage rights that may be granted to users of such data.

¹⁵⁷ See <<http://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/gbrel.html>> at 26 February 2007.

¹⁵⁸ See J Sulston and G Ferry, "The Common Thread: A Story of Science, Politics, Ethics and the Human Genome", 193 (2002)

4.18 It is instructive to examine the NIH Data Sharing Policy (discussed in detail in Chapter 8) for an idea of the access and usage rights applying to GenBank data. In effect, each contributor (such as a laboratory or large-scale centre producing DNA sequence data) would be required to grant to the relevant database organisation (including GenBank) a licence to use its DNA sequence data. Such a licence would:

- be unrestricted: there must be no restrictions at all on the use of the DNA sequence data;
- be irrevocable: the data must be permanently available;
- be no-cost; and
- include the right to sub-licence or redistribute to any other person the right to use the DNA sequence for any purpose.

4.19 Such a licence would be subject only to the obligation to acknowledge the parties who have rights in relation to the DNA sequence data. Additionally, contributors would need to warrant that they have the authority to grant the necessary rights to the database (as set out in the licence) and that the DNA sequence data is accurate and of high quality.

4.20 It is important to note that such a grant of licence need not be in writing between the contributor and the database organisation. It may be implied in certain circumstances.

4.21 However, the GenBank website states:

Databases of molecular data on the NCBI Web site include such examples as nucleotide sequences (GenBank), protein sequences, macromolecular structures, molecular variation, gene expression, and mapping data. They are designed to provide and encourage access within the scientific community to sources of current and comprehensive information. Therefore, NCBI itself places no restrictions on the use or distribution of the data contained therein. However, some submitters of the original data (or the country of origin of such data) may claim patent, copyright, or other IP rights in all or a portion of the data (that has been submitted). NCBI is not in a position to assess the validity of such claims and, therefore, cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in the molecular databases.¹⁵⁹

Example 3 - NIH: Genetic Association Information Network (GAIN)¹⁶⁰

4.22 The Genetic Association Information Network (GAIN) is a community resource project which involves a public-private partnership between the NIH and several private pharmaceutical companies, including Pfizer, Affymetrix and Abbot Laboratories. GAIN aims to understand the complex genetic factors influencing risks of contracting common diseases by conducting a series of whole genome association studies that use samples from patients with these diseases.¹⁶¹

4.23 A feature of GAIN is that it imposes restrictions on publication by external data users who are permitted to access the data. Under GAIN's publication policy,

¹⁵⁹ <<http://www.ncbi.nlm.nih.gov/About/disclaimer.html>> at 23 January 2007.

¹⁶⁰ See Foundation for the NIH, *GAIN Program* <http://www.fnih.org/GAIN/GAIN_home.shtml>

¹⁶¹ See Foundation for the NIH, *GAIN Program Overview* <<http://www.fnih.org/GAIN/Background.shtml#Program>>.

contributing investigators are given a period of nine months within which they have the exclusive right to submit publications based on their data.¹⁶² Approved users may obtain access to the data on signing a restrictive agreement but are not permitted to publish analyses derived from GAIN project datasets during that nine-month period.¹⁶³

4.24 Under the GAIN Data Use Certification policies and procedures for investigators and their institutions,¹⁶⁴ publications are still permissible during this period if:

- publication only uses GAIN data that was accessed through the *public* section of the GAIN Database; or
- the period described on the database is less than nine months; or
- the user obtains the written consent of the “Contributing Study Investigator”.

Ownership - GAIN

4.25 GAIN does not make any definitive statement as to ownership of rights in relation to the GAIN-supported data, but requests that users (in accordance with GAIN’s Intellectual Property Policy¹⁶⁵) treat the data used as “pre-competitive” and not make any intellectual property claims over it, so that certain applications can remain freely available without the need for licensing requirements.¹⁶⁶

4.26 GAIN’s intention is to promote the dissemination of its datasets as widely as possible and GAIN’s goal is to sustain public benefit by not pursuing intellectual property protection that would prevent or block access to, or use of, any element of its data or any conclusions drawn directly from such data.¹⁶⁷

4.27 Users are responsible for complying with all relevant laws relating to the use of the data. If GAIN does not hold rights in relation to the data, this would mean that users would have to obtain direct permission from the owner of those rights (if any) for usage that might amount to an infringement.

Control - GAIN

4.28 Prospective users must agree to retain control over the data and not to distribute data to anyone else other than their research staff provided that the research

¹⁶² See Foundation for the NIH, *Policies and Procedures: GAIN Publication Policy* <<http://www.fnih.org/GAIN/policies.shtml#Publication>>.

¹⁶³ Ibid *GAIN Publication Policy*. This summary of GAIN is drawn from Arti Rai and Rebecca Eisenberg, *Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative*, Duke Law School’s Science, Technology and Innovation Research Paper Series, Research Paper No 11 (2006) <<http://ssrn.com/abstract+941146>>.

¹⁶⁴ See <http://www.fnih.org/GAIN2/Data_Use_Certification.pdf>

¹⁶⁵ See <http://www.fnih.org/GAIN2/ip_policy.shtml>.

¹⁶⁶ Including but not limited to: the use of markers in developing assays and diagnostic tools utilizing a variety of single or multiple technical platforms, the use of combinations of markers in multiplex assays and the use of markers as guides toward identification of new drug targets - <http://www.fnih.org/GAIN2/Data_Use_Certification.pdf>.

¹⁶⁷ See <http://www.fnih.org/GAIN2/ip_policy.shtml>.

staff have themselves agreed to the terms of the Data Use Certification process for data access and use.

Access and Use - GAIN

4.29 Users can only access data or datasets in accordance with any parameters set out for appropriate research use described on the GAIN website and any limitations on use imposed on any of the GAIN project datasets. A researcher who executes a Data Use Certification will have agreed to its terms and the policies of GAIN for the use of GAIN datasets.

4.30 Users are also requested under the Data Use Certification process to be consistent with the NIH Best Practices for the Licensing of Genomic Inventions policy.¹⁶⁸ This policy recommends that:

- unless private sector investment is involved or significant research and development is required, patent protection for research should not be sought;
- non-exclusive licensing of research should be pursued to facilitate the availability of broad enabling technologies and research uses of inventions;
- licensing policies and strategies that maximise access, as well as commercial and research utilisation should be implemented; and
- funding recipients and the intramural technology transfer community should reserve in their licence agreements the right to use the licensed technologies for their own research and educational uses, and should allow other institutions to engage in the same uses.

Example 4 - NIH - National Human Genome Research Institute (NHGRI)

4.31 NHGRI is NIH's lead entity for advancing human health through genetic research.¹⁶⁹ NHGRI seeks to contribute to genome data and knowledge bases by funding and supporting a range of research projects and public-private consortia efforts, including the HapMap Project, the Trans-NIH Mouse Initiative and the International Human Genome Sequencing Consortium (IHGSC), among others.

Ownership – NHGRI

4.32 The NHGRI website does not contain any express statement as to the ownership of rights in relation to data. However, NHGRI's Copyright Policy¹⁷⁰ goes on to differentiate between:

- data prepared by government employees (which is not being subject to copyright restrictions or conditions); and

¹⁶⁸ See <http://www.ott.nih.gov/policy/lic_gen.html>.

¹⁶⁹ See Claire Driscoll, 'NIH data and resource sharing, data release and IP policies for genomics community resource projects' (2005) 15(1) *Expert Opinion on Therapeutic Patents*, 1; See also: <<http://www.genome.gov/>>

¹⁷⁰ See <<http://www.genome.gov/12514471>>.

- data published on NHGRI with the permission of the copyright owner or holder (which is not owned by NHGRI but by the original contributing owner of that information).

4.33 The NHGRI Copyright Policy states that users who access the NHGRI website will encounter non-government information that has been licensed by private individuals and companies and that may be protected by copyright laws. Transmission or reproduction of these items will require the user to obtain the written permission of the copyright owners and in the case of an identifiable personal image, permission will need to be obtained to avoid infringement of privacy and publicity rights.

Control – NHGRI

4.34 NHGRI has been concerned with the restrictive control exercised over data by some commercial entities, particularly through the use of patents. However NHGRI and NIH support the securing of patents, particularly when the patents are associated with potential products. In developing data access policies, NHGRI has been aware of the need to:

Balance the important dual goals of continuing to rapidly place huge amounts of data in the public domain and encouraging restriction-free sharing of genomic tools, whilst also ensuring that more applied inventions, notably those closer to being an actual product, can be patented.¹⁷¹

4.35 For example, the NHGRI Trans-NIH Mouse Initiative (referred to below) addresses the issue of patenting and data submission¹⁷². It states that institutions can choose to patent technology, but that ‘inappropriate’ enforcement could interfere with the distribution of data. NHGRI goes further in stating that a patented resource must still be made reasonably available and accessible to the research community.

Access and Use – NHGRI

4.36 Since its formation in 1990, NHGRI has had a policy of free and open access to genomic data. In 1997, NHGRI developed a Policy for Release and Database Deposition of Sequence Data. This was in response to the Second International Strategy Meeting on Human Genome Sequencing, which affirmed the principle set down in the First International Strategy meeting in 1996 that primary genomic sequence should be rapidly released (the Bermuda Principles).¹⁷³ The policy required NHGRI grantees engaged in large-scale genomic DNA sequencing to automatically release sequence assemblies of 2kb or larger within 24 hours of their generation.¹⁷⁴ In 2001, the policy was updated to extend not only to initial sequence assemblies and early stage data, but also to data generated while producing finished sequences and data that is assembled with and significantly changes sequence assemblies already released. The updated policy provided:

¹⁷¹ Claire Driscoll, ‘NIH data and resource sharing, data release and IP policies for genomics community resource projects’ (2005) 15(1) *Expert Opinion on Therapeutic Patents*, 2.

¹⁷² See <<http://www.nih.gov/science/models/mouse/sharing/5.html>>.

¹⁷³ <<http://www.genome.gov/1000910>> at 28 February 2007.

¹⁷⁴ Ibid.

Sequence trace data, and all ancillary information specified in a standard format provided by the database, should be released weekly into the NCBI Trace Repository.¹⁷⁵

4.37 All submissions to the Trace Repository carry the following notice:

As a public service to the biological research community, these data are being made available by the sequence producers before assembly and before scientific publication. Once deposited, but prior to the publication of the complete sequence of the relevant genome, the data are available to all as follows:

1. The data may be freely downloaded by all users, for use in all types of analyses (with the single exception described in item iv).
2. The data may be repackaged in other databases, provided that appropriate acknowledgement is given.
3. Users are free to use the data for publication in scientific papers analyzing particular genes and regions; the source of the DNA sequence data should be appropriately acknowledged.
4. The producing laboratories intend to publish the sequence of the genome and certain large-scale analyses of the sequence in a timely manner upon the completion of sequence data acquisition. Therefore, the sole exception to the unrestricted use of these unpublished data is that the data may not be used for the initial publication of the complete genome sequence assembly or other large-scale analyses. In this context, "large-scale" refers to regions the size of the whole genome or individual chromosomes and examples of "large-scale analyses" include identification of regions of evolutionary conservation across an entire genome and identification of complete sets of genomic features such as genes, repeat structures, GC content, etc. The producing laboratories will, however, be open to the possibility of collaboration on such assemblies or analyses."
5. Any redistribution of the data should carry this notice.¹⁷⁶

4.38 The NHGRI website states that it is committed to ensuring that the information on its site is accessible to all users,¹⁷⁷ particularly in relation to large-scale DNA sequence data-sets under its Rapid Data Release Policies.¹⁷⁸

4.39 The website contains certain programs which do impose restrictions and conditions on access and use. For example, the NHGRI Medical Sequencing Program (MSP) requires users to submit a formal data access request and to agree to certain certification terms:¹⁷⁹

- to use the MSP dataset solely in accordance with the research project described in the data access request;
- to use the data in accordance with any parameters and limitations described on the MSP web site;
- to ensure that use of the data is consistent with applicable law (i.e. in accordance with any applicable rights of ownership in relation to data);

¹⁷⁵ Ibid.

¹⁷⁶ Ibid

¹⁷⁷ See <<http://www.genome.gov/accessibility.cfm>>.

¹⁷⁸ See <<http://www.genome.gov/10506537>>.

¹⁷⁹ See <<http://www.genome.gov/20019653>>.

- not to distribute any data to persons outside of the research staff set out in the access form and who do not agree to the terms of certification; and
- to follow the NHGRI intellectual property policy, which urges users to avoid making intellectual property claims on the data, to ensure that derived data and conclusions remain freely accessible without licensing and users are encouraged to disseminate results in peer reviewed journals.

4.40 The specifics of the data release and data sharing policies in use for several NHGRI-funded genomics projects are set out below. Where the project databases do not appear to have a data sharing policy in place, the NHGRI policy is presumed to apply.

Example 5 - NHGRI: International Human Genome Sequencing Consortium (IHGSC)

4.41 IHGSC is a community resource project, the focus of which has been the large-scale DNA mapping and sequencing of the human genome. In 2003, IHGSC celebrated the sequencing of the human genome.¹⁸⁰

The international effort to sequence the 3 billion DNA letters in the human genome is considered by many to be one of the most ambitious scientific undertakings of all time, even compared to splitting the atom or going to the moon.¹⁸¹

Ownership and Control – IHGSC

4.42 IHGSC does not appear to have a website separate from NIH. Issues of ownership of rights in relation to data, and control, access and use by the consortium are subject to the NHGRI copyright and data release policies.

Access and Use – IHGSC

4.43 The focus on community benefit means that it is important for any data or information produced by IHGSC to be readily accessible by the community. As a NHGRI-funded project, IHGSC had in place data release and data sharing policies, based on the Bermuda Principles, from an early stage.

4.44 The sequence data generated by the Human Genome Project has been swiftly deposited into public databases and rapidly released without restrictions on its use or redistribution.

Example 6 - NHGRI: Mouse Genome Sequencing Consortium (MGSC) and the Jackson Laboratory Mouse Genome Informatics Database

4.45 MGSC involves collaboration between four major international sequencing centres – the Broad Institute/MIT Center for Genomic Research, the Washington University Genome Sequencing Center, the Wellcome Trust Sanger Institute, and the Baylor College of Medicine Human Genome Sequencing Center - and collaboration between leading mouse researchers known as the Mouse Sequencing Liaison

¹⁸⁰ Ibid.

¹⁸¹ <<http://www.genome.gov/11006929>> at 28 February 2007.

Group.¹⁸² The project aims to produce a “robust physical map” and a physical genome data sequence of the mouse.¹⁸³ The project is support by NHGRI and the Wellcome Trust, and also has ties with the international database Ensembl, which is a joint project between the European Bioinformatics Institute and the Sanger Centre.¹⁸⁴

4.46 Each of these collaborative partners and researchers has an independent website, which may have differing approaches on the issues of ownership, control and access and use.

4.47 For instance, MGSC collaborates with the Jackson Laboratory¹⁸⁵ for the development of a model organism database known as the Mouse Genome Informatics (MGI) database.¹⁸⁶ It provides integrated access to data on the genetics, genomics and biology of the laboratory mouse.¹⁸⁷

4.48 MGI contains information on mouse genetic markers, molecular segments, phenotypes, comparative mapping data, experimental mapping data, and graphical displays for genetic, physical and cytogenetic maps.

4.49 The following copyright notice appears on the MGI web site:¹⁸⁸

WARRANTY DISCLAIMER AND COPYRIGHT NOTICE

THE JACKSON LABORATORY MAKES NO REPRESENTATION ABOUT THE SUITABILITY OR ACCURACY OF THIS SOFTWARE OR DATA FOR ANY PURPOSE, AND MAKES NO WARRANTIES, EITHER EXPRESS OR IMPLIED, INCLUDING MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE OR THAT THE USE OF THIS SOFTWARE OR DATA WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS, OR OTHER RIGHTS. THE SOFTWARE AND DATA ARE PROVIDED "AS IS".

This software and data are provided to enhance knowledge and encourage progress in the scientific community and are to be used only for research and educational purposes. Any reproduction or use for commercial purpose is prohibited without the prior express written permission of the Jackson Laboratory.

Ownership – MGI

4.50 The copyright notice indicates that the Jackson Laboratory asserts copyright in relation to the MGI database. Any use of the data outside of the circumstances allowed in the copyright notice requires permission from the Jackson Laboratory.

¹⁸² See <<http://www.genome.gov/10001859>>.

¹⁸³ <<http://www.genome.gov/10001859>> at 1 March 2007.

¹⁸⁴ Ibid.

¹⁸⁵ A listing of information resources provided by the Jackson Laboratories is available at <http://www.jax.org/resources/supporting_resources.html>.

¹⁸⁶ See http://www.informatics.jax.org/mgihome/other/collab_and_acknow.shtml#collab.

¹⁸⁷ See <<http://www.informatics.jax.org/>> at 1 March 2007.

¹⁸⁸ See <<http://www.informatics.jax.org/mgihome/other/copyright.shtml>> at 22 May 2007.

Control – MGI

4.51 The copyright notice states that the software and data provided by the Jackson Laboratory are only to be used for research and educational purposes. This is reinforced by the statement that the data is provided to enhance knowledge and encourage progress in the scientific community.

4.52 Any use of the data for commercial purposes is prohibited without the prior express written permission of the Jackson Laboratory. This is not an absolute prohibition on commercial use of the data, but it does require a commercial user to obtain written permission from the Jackson Laboratory. The Jackson Laboratory may require commercial users to enter into a written contract, whereby additional conditions are imposed for commercial use of the data. This allows the Jackson Laboratory to control the use of the data on a case by case basis – for example, more stringent conditions may be imposed on purely commercial users than on researchers with more limited commercial possibilities.

4.53 The MGI web site also proscribes how specific database projects and specific data are to be cited.¹⁸⁹

Access and Use – MGI

4.54 Aside from the controls exercised over use of MGI data as described above, there do not appear to be any mechanisms limiting access to data in the MGI database. Users are not required to subscribe to the database or enter into any click-wrap agreement before gaining access to the database – data can be openly searched and accessed without restriction.

Example 7 - NHGRI: Rat Genome Sequencing Consortium (RGSC)

4.55 RGSC is an international research team dedicated to the large-scale DNA mapping and sequencing of the rat genome. The project is led by the Human Genome Sequencing Center at Baylor College of Medicine in Houston, and is primarily funded by NHGRI in conjunction with the National Heart, Lung and Blood Institute (NHLBI). In March 2004, RGSC announced that it had completed a high-quality, draft sequence of the genome of the laboratory rat.¹⁹⁰ RGSC makes data exploring how the rat's genetic blueprint compares to the genetic blueprint of mice and humans available before scientific publication.¹⁹¹

Ownership - RGSC

4.56 Under the website's "Disclaimer" page,¹⁹² it is stated that all software and data found on the website are the copyright of the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC). The copyright statement also provides:

Unless otherwise noted, all images, forms, pages and scripts were created by an employee or student at BCM-HGSC or is in the public domain. If you see an unauthorized copy of your

¹⁸⁹ See <<http://www.informatics.jax.org/mgihome/other/citation/shtml>> at 22 May 2007.

¹⁹⁰ <<http://www.genome.gov/11511308>> at 1 March 2007.

¹⁹¹ Ibid.

¹⁹² See <<http://www.hgsc.bcm.tmc.edu/docs/disclaimer3.html>>.

picture, please contact us immediately and we will remove it. Permission to reproduce anything on these pages is required.¹⁹³

Control – RGSC

4.57 The website differentiates between data which has been scientifically published and unpublished data.

4.58 Under the Disclaimer page, restrictions will apply to unpublished sequences, until such time as they are formally published in scientific literature by the investigators who originally produced the data.

4.59 Users can still access and use unpublished data in specified research projects (with appropriate acknowledgement), but unpublished data cannot be published or distributed without prior approval from the submitting group.

4.60 The Disclaimer pages also states that no data can be redistributed in any form without the prior permission of BCM-HGSC.

Access and Use – RGSC

4.61 Under the “Conditions for Use” page,¹⁹⁴ data that has been made available before scientific publication may be freely downloaded, used in analyses and repacked in databases. Any redistribution should carry notice of the Conditions for Use.

4.62 However, this statement may be at odds with the Disclaimer statement that no redistribution can occur without prior permission.

Example 8 - NHGRI: The SNP Consortium (TSC)

4.63 TSC was formed as a non-profit foundation in 1999, primarily as a private-sector initiative with some NIH involvement. It focused on discovering single point mutations (called single nucleotide polymorphism or SNPs) in the human genome, and had its final data release in September 2004.

Ownership -TSC

4.64 TSC did have a website in operation; however, it appears to be presently off-line. To this extent, it is unclear whether issues of ownership rights in relation to data, control and access and use by this consortium are subject to the NHGRI copyright and data release policies.

¹⁹³ <<http://www.hgsc.bcm.tmc.edu/docs/disclaimer3.html>> at 13 June 2007

¹⁹⁴ See <http://www.hgsc.bcm.tmc.edu/projects/rat/conditions_for_use.html>

Control – TSC

4.65 TSC was designed to prevent competitors from capturing or restraining use of the data by making “the information related to these SNPs available to the public without IP restrictions”.¹⁹⁵ TSC also adopted:

A policy of waiving the right to receive patent protection on the raw SNP data and agreed to publish the mapped SNPs as quickly as was feasible.¹⁹⁶

Access and Use - TSC

4.66 The information disclosed by TSC has been catalogued on HapMap.¹⁹⁷ New SNPs were publicly released quickly through the public NIH Single Nucleotide Polymorphism Database (dbSNP).¹⁹⁸ When SNPs had been densely genotyped, they were released publicly without restrictions.¹⁹⁹

However, the SNP Consortium did not just dump the data. They filed patent applications and then characterized the SNP markers enough so that they could be sure that nobody else could patent them. At that point, they would abandon the patent. It is a very sophisticated intellectual property strategy that in the end was intended to bolster the public domain.²⁰⁰

4.67 By adopting the policy of making data freely available to researchers, it is estimated that TSC considerably reduced the expenditure that would have otherwise been required to complete the project.²⁰¹

Example 9 - NHGRI - Mammalian Gene Collection (MGC)

4.68 MGC has the goal of creating:

A public collection of affordable, sequence-verified full-length complimentary DNAs (cDNAs) for every known mouse and human gene, and a subset of rat genes.²⁰²

4.69 In 2005, the project added cow cDNAs generated by Genome Canada.²⁰³

¹⁹⁵ Claire Driscoll, ‘NIH data and resource sharing, data release and IP policies for genomics community resource projects’ (2005) 15(1) *Expert Opinion on Therapeutic Patents*, 3

¹⁹⁶ Claire Driscoll, ‘NIH data and resource sharing, data release and IP policies for genomics community resource projects’ (2005) 15(1) *Expert Opinion on Therapeutic Patents*, 4.

¹⁹⁷ <<http://snp.cshl.org/datareleasepolicy.html.en>> at 1 March 2007.

¹⁹⁸ Ibid.

¹⁹⁹ Ibid.

²⁰⁰ Robert Cook-Deegan, Duke University, ‘The Urge to Commercialize: Interactions between Public and Private Research and Development’ in National Research Council of the National Academies (2003) Julie M. Esnau and Paul F. Uhler (eds), *The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium*, National Academies Press, Washington USA, p91

²⁰¹ Amounting to a shared project cost of US \$44 million and yielding 1.8 million SNPs, as opposed to an original estimated cost of US \$250 million to identify 150,000 SNPs: see J Reichmann and P Uhler, ‘A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment’ (2003) 66 *Law and Contemporary Problems* 315, 458 <<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

²⁰² Claire Driscoll, ‘NIH data and resource sharing, data release and IP policies for genomics community resource projects’ (2005) 15(1) *Expert Opinion on Therapeutic Patents*, 5.

²⁰³ <<http://mgc.nci.nih.gov/>> at 1 March 2007.

Ownership – MGC

4.70 The MGC website does not make any express statement as to ownership of rights in relation to data.

4.71 However, it does state that the cDNA clones generated by the MGC are available through the Integrated Molecular Analysis of Genomes and their Expression Consortium (I.M.A.G.E.) clone distribution network and are fully accessible to the community.²⁰⁴

4.72 The page links to the Image consortium website which requires users to enter into an “I.M.A.G.E. Consortium Good Faith Agreement” concerning use and distribution of Arrayed cDNA Clones.²⁰⁵

4.73 Under that agreement, paragraph (f) states that ownership of the unarrayed CDNA “libraries” from which clones were arrayed is retained by the “Originators” of those libraries. “Originators” are defined in the first paragraph as the institutions that developed the original libraries from which those clones were derived.

4.74 It seems from this statement that “originating” contributors and researchers to the MGC database will retain ownership rights in relation to the submitted data.

4.75 It should also be noted that the MGC request for funding proposals contains a “Determination of Exceptional Circumstances”.²⁰⁶ This option under the US Bayh-Dole Act allows a non-profit organisation (such as NIH - NGHRI) to retain title in funded inventions, rather than title in such inventions going to the grantee.²⁰⁷

This administrative and legal tool is only used for programmes in which the main goals are to create data and/or resources that are to be made widely available with minimal restrictions on their use.²⁰⁸

4.76 Importantly, this exceptional circumstances determination may be used to prevent restrictive patents being placed over important publicly-funded data. NIH obtained a determination of exceptional circumstances as a condition of MGC government contracts, so that patent rights in inventions from MGC NIH funded projects are retained by NIH.²⁰⁹

²⁰⁴ See <<http://mgc.nci.nih.gov/Info/Summary>>.

²⁰⁵ See <<http://image.llnl.gov/image/html/GFA.shtml>>.

²⁰⁶ See US Patents Code – Disposition of Rights - 35 USC § 202(a) at <http://www.law.cornell.edu/uscode/html/uscode35/usc_sec_35_00000202----000-.html> (29 May 2007) and Claire Driscoll, “NIH data and resource sharing, data release and IP policies for genomics community resource projects”, *Expert Opin. Ther. Patents* (2005) 15(1), p5

²⁰⁷ For more information, see Robert Cook-Deegan, Duke University, ‘The Urge to Commercialize: Interactions between Public and Private Research and Development’ in National Research Council of the National Academies (2003) Julie M. Esnaú and Paul F. Uhler (eds), *The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium*, National Academies Press, Washington USA, p91

²⁰⁸ Ibid

²⁰⁹ Robert Cook-Deegan, Duke University, ‘The Urge to Commercialize: Interactions between Public and Private Research and Development’ in National Research Council of the National Academies (2003) Julie M. Esnaú and Paul F. Uhler (eds), *The Role of Scientific and Technical Data and*

Control – MGC

4.77 Users must agree to the terms of the I.M.A.G.E. Consortium Good Faith Agreement before being granted access to the data. Users can create “derivative products” for commercial or non-commercial use. Depending upon the actual use sought, users may redistribute or may be prohibited from redistributing cDNA clones and data variations.

Access and Use – MGC

4.78 The MGC website states that its clone data is available without restriction to the scientific community through the I.M.A.G.E. consortium distribution network.²¹⁰ MGC makes the cDNAs available at a price.

4.79 The clone data has been described as able to be freely used for research purposes without onerous intellectual property restrictions or reach-through licensing terms.²¹¹ However, the MGC website implies under the page entitled “Where to Buy” that its clone data is available for purchase through the I.M.A.G.E. consortium network to the scientific community, so it appears that a licensing fee is involved.²¹²

4.80 Whilst paragraph (f) of the I.M.A.G.E. agreement states that any patentable inventions first made by any party using the arrayed clones will remain the property of the inventing party, it goes on to state that the agreement does not constitute the Originator’s waiver of any patent rights.

4.81 It seems that any intention to obtain a patent on work derived from the clone data will be subject to existing patent rights held by the owners of rights in relation to the data. It is unclear from the agreement whether the term “any patent rights” would extend beyond an existing right to the mere potential to acquire patent rights in the clone data.

Example 10- NHGRI - Encyclopaedia of DNA Elements (ENCODE)²¹³

4.82 ENCODE was launched by NHGRI in 2003 as a public research consortium to identify and characterise all functional elements in the human genome sequence. Data produced by ENCODE is released to the public via database deposits.

Ownership – ENCODE

4.83 ENCODE has a Project Data Release Policy which discusses data access and intellectual property considerations and issues.²¹⁴ The Project Data Release Policy does not make any express statement as to rights in relation to data.

Information in the Public Domain: Proceedings of a Symposium, National Academies Press, Washington USA, p91

²¹⁰ See <<http://mgc.nci.nih.gov/Info/Buy>>.

²¹¹ See Claire Driscoll, “NIH data and resource sharing, data release and IP policies for genomics community resource projects”, *Expert Opin. Ther. Patents* (2005) 15(1), p5

²¹² See <<http://mgc.nci.nih.gov/Info/Buy>>

²¹³ See <<http://www.genome.gov/page.cfm?pageID=10005107>>

²¹⁴ See <<http://www.genome.gov/12513440>>.

4.84 ENCODE intends to publish all findings of the consortium within six months of the end of the pilot project. In the meantime, any individual research groups in the consortium who contribute data are free to publish the results of their own efforts at any time.

4.85 All users of consortium data should obtain the consent of the “data producers” before using unpublished data in publications. Appendix B to the policy (setting out ENCODE’s concerns regarding the effect of rights and claims on accessibility to data) further states that if required, NHGRI may eliminate the right of contributing parties under grants to retain title in their submitted data.

4.86 These statements imply that at present, copyright and rights in relation to data remain with the contributors and submitters of that data. However, this may be subject to change in the future.

Control – ENCODE

4.87 The Data Release Policy differentiates between two data types:

- Data verification: assessing the data reproducibility of an experiment; and
- Data validation: referring to data confirmation by other, independent methods.²¹⁵

4.88 Users who are outside of the consortium are required to properly cite data sources and to:

Recognise the interests of the data producers to publish reports on the generation and analysis of their data.²¹⁶

4.89 This means that users must refer to the copyright interests of the data producers before publishing such data.

4.90 In addition:

- the data producers themselves are requested not to do anything that would restrict the use of data by others; and
- researchers wishing to join the consortium must abide by the Policy and all users (whether consortium members or not) who incorporate data into a patentable invention are expected to non-exclusively licence the patent so that it does not restrict access of others to ENCODE data.

Access and Use – ENCODE

4.91 The data produced by ENCODE Consortium members are deposited to public databases²¹⁷ and non-sequence based data is made available on public

²¹⁵ For each data type, the Consortium seeks to identify a minimal verification standard necessary for public release of each data type. The Consortium members will also identify additional levels of validation that will be applied in subsequent analyses of the data or with additional experimentation where appropriate.

²¹⁶ Under *ENCODE Publication Policy/Intellectual Property Considerations*
<<http://www.genome.gov/12513440>>

databases²¹⁸ and are available for use to both consortium members and non-members all to use without restriction.²¹⁹

4.92 “Without restriction” has to be read in the context of ownership and the controls described above and NHGRI’s primary interest, being the widespread availability of all information and any inventions generated during the ENCODE project.²²⁰

Example 11 - The National Centre for Biotechnology Information (NCBI) Databases

4.93 The National Centre for Biotechnology Information (NCBI), a division of the National Library of Medicine (NLM)²²¹ at the NIH, was established in 1988 as a national resource for molecular biology information. It is responsible, amongst other things, for the development of:

New information technologies to aid in the understanding of fundamental molecular and genetic processes that control health and disease. More specifically, the NCBI has been charged with creating automated systems for storing and analysing knowledge about molecular biology, biochemistry, and genetics; facilitating the use of such databases and software by the research and medical community; coordinating efforts to gather biotechnology information both nationally and internationally; and performing research into advanced methods of computer-based information processing for analysing the structure and function of biologically important molecules.²²²

4.94 NCBI carries out its responsibilities in a number of ways including:

- collaboration with other NIH institutes, academia, industry and governmental agencies;
- the development, distribution and support of access to a variety of databases and software for the scientific and medical communities; and
- development and promotion of standards for databases, data deposition and exchange, and biological nomenclature.

Example 12 - NCBI: Database of Genotype and Phenotype (dbGaP)

4.95 The Database of Genotype and Phenotype (dbGaP)²²³ is a web-based searchable repository of disease-related data collected over many years, which was launched by NIH in December 2006.²²⁴ It archives and distributes data from genome

²¹⁷ Such as the University of California, Santa Cruz ENCODE Project Genome Bioinformatics Group website: <<http://genome.ucsc.edu/ENCODE/>> at 1 March 2007

²¹⁸ Such as the Gene Expression Omnibus: <<http://www.ncbi.nlm.nih.gov/projects/geo/>> at 1 March 2007; and ArrayExpress: <<http://www.ebi.ac.uk/arrayexpress/>> at 1 March 2007

²¹⁹ <<http://www.genome.gov/page.cfm?pageID=10005107#4>> and

<<http://www.genome.gov/12513440>> at 1 March 2007.

²²⁰ Under *Appendix B: ENCODE Intellectual Property Issues* <<http://www.genome.gov/12513440>>

²²¹ See <<http://www.nlm.nih.gov/>>.

²²² See <<http://www.ncbi.nlm.nih.gov/About/glance/ourmission.html>> at 22 January 2007.

²²³ <http://www.ncbi.nlm.nih.gov/entrez/query/Gap/gap_tmpl/about.html>

²²⁴ See Gene Russo, *NIH offers free access to wealth of disease data*, *Nature*, 21 December 2006

wide association studies to enable links to be made between genes and disease. The launch press release explains:

[Genome wide association] studies explore the association between specific genes (genotype information) and observable traits, such as blood pressure and weight, or the presence or absence of a disease or condition (phenotype information). Connecting phenotype and genotype data provides information about the genes that may be involved in a disease process or condition, which can be critical for better understanding the disease and for developing new diagnostic methods and treatments.²²⁵

4.96 The dbGaP brings together, in a central location, data collected in numerous epidemiological studies. It enables interested researchers to see all study documentation, view summaries of the measured variables and search vast amounts of genetic, phenotypic and study-protocol data simultaneously. It has made the results of some large studies available to interested parties for the first time and greatly increases the scope and efficiency of access to the data. Among the studies included in dbGaP are the 600-subject Age-Related Eye Diseases Study (AREDS) supported by the National Eye Institute,²²⁶ the 2,573-subject Parkinsonism Study conducted by the National Institute of Neurological Disorders and Stroke (NINDS)²²⁷ and the landmark 14,000-subject Framingham Heart Study funded by the National Heart, Lung and Blood Institute,²²⁸ which has followed successive generations since 1948. It is intended that data from the GAIN project will also be included in the dbGaP database.

4.97 In order to protect the privacy of research participants, all datasets in dbGaP have two levels of access:

- public or “open-access”; and
- “controlled-access”.

4.98 Open-access data consists of data that is released publicly, without an approval system for access, although some data may carry restrictions on its use. This open-access data, consisting of studies, documents, variable summaries and association results, is to be made available through dbGaP’s FTP webpage.²²⁹ Controlled-access data relates to specific individuals, such as phenotype trait measurements for individuals or a set of test subject genotypes, and is subject to access restrictions and an approval system. This controlled-access data is only accessible through dbGaP’s “Authorized Access” webpage.²³⁰

²²⁵ See <http://www.nlm.nih.gov/news/press_releases/dbgap_launchPR06.html>

²²⁶ See <<http://www.nei.nih.gov>>. This study was a multi-centre, case-controlled, prospective study of the clinical course of age-related macular degeneration and age-related cataracts.

²²⁷ See <<http://www.ninds.nih.gov>>. This was a case-controlled study that gathered DNA, cell line samples and detailed phenotypic data.

²²⁸ See <<http://www.nhlbi.nih.gov>>

²²⁹ See <<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>> link to

<ftp://ftp.ncbi.nlm.nih.gov/dbgap/> (19 June 2007)

²³⁰ See <http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login> (19 June 2007)

Ownership - dbGaP

4.99 The owner of rights in relation to either open access data or controlled data held by dbGaP appears to be the original creator (or collector) of that data.²³¹ Generally the creator or collector of such data is a research organisation, which may be government or privately funded.

Control - dbGaP

4.100 Control of dbGaP open-access or controlled-access data remains with the owner of rights in relation to the data. The dbGaP database managers may appear to exercise control over the data, but this control is subject to rights granted and restrictions imposed by the rights-owner.

4.101 In relation to controlled-access data, potential users must:

- be associated with a scientific organisation recognised by and registered for an account with the NIH eRA commons system;²³²
- be identified by their research organisation to the NIH eRA as a “Principal Investigator”; and
- lodge a request for controlled-access data, stating the users research objectives.

4.102 The request lodged by the user is given to a representative of the user’s organisation. This representative must certify that the organisation will be bound by the controlled access studies access restrictions. The Principal Investigator also provides certain assurances for its access and use, known as the “Data Use Certification” (DUC – see below).

4.103 The request is then sent to a NIH “Data Access Committee” (DAC) which decides whether or not the user’s research objectives are in accordance with any “Access Restrictions” (see below) and that the Principal Investigator has made the necessary assurances required for the DUC.

4.104 It is unclear whether, in making their decision, the DAC consults with the owners of rights in relation to the controlled-access data or whether the DAC is empowered to make such decisions without having to refer back to the rights owners.

Access and Use - dbGaP

4.105 Ensuring the privacy of personal information in the clinical and biological data contained in dbGaP has been a primary focus of the database managers. Open-access data study pages can be browsed online or downloaded from dbGaP without the need to obtain permission or authorisation and are provided without any information that could identify the subjects. This amounts to a grant by the owners of rights in relation to that open-access study data of an open access licence to any user.

4.106 Some open access study pages used to contain “Access Restrictions” imposed by the different owners of rights in relation to provided controlled-access

²³¹ See <http://www.ncbi.nlm.nih.gov/entrez/query/Gap/gap_tmpl/about.html>

²³² See <<https://commons.era.nih.gov/commons-help/161.htm>> at 19 June 2007.

data. These restrictions were apparently determined by the terms of the original Individual Informed Consent Documents, which gave test subjects elective options as to how their data was to be used. For example, the “Access Restrictions” published with the NINDS Parkinsonism study page stated:

This data will be used only for research purposes. It will not be used to determine the individual identity of any person or their relationship to another person. The investigator will acknowledge the source of the data when publishing results based on this data, and also, acknowledge the NINDS DNA and Cell Line Repository in any publications, including posters, platform presentations, articles, press releases, and manuscripts. The recipient acknowledges that they have complied with will applicable state, local, and federal laws or regulations and institutional policies regarding human subjects and genetics research. Secondary distribution and shared use of this data without registration by secondary parties is prohibited.²³³

4.107 These “Access Restrictions” have now been characterized on this page as “Use Restrictions”.²³⁴ It seems that open access to this data is still “open” in that it is still publicly accessible, but that there are still conditions on its actual use. Using the NINDS example, the Use Restrictions statement amounts to a grant of licence by NINDS to any person to use this data for a purpose associated with research only, subject to the following conditions. The user must:

- not use the data to determine the individual identity of any person or their relationship to another person;
- acknowledge the source of the data when publishing results based on the data;
- acknowledge the NINDS DNA and Cell Line Repository in any publications, including posters, platform presentations, articles, press releases, and manuscripts;
- not distribute or share the data with a secondary party unless the secondary party is registered; and
- comply with applicable state, local, and federal laws or regulations and institutional policies regarding human subjects and genetics research.

4.108 Researchers who wish to obtain access to data classified as controlled-access data may be granted a licence to use the data by the owner of rights in relation to this data. Each rights owner may impose different conditions in relation to such a licence, but generally a potential user must provide the following assurances under the DUC to obtain such a licence:²³⁵

- the data will only be used for research approved by a DAC;
- data confidentiality will be protected and appropriate data security measures must be in place to prevent unauthorised access;

²³³ See <<http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000003>> (24 January 2007)

²³⁴ See <<http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000003>> (25 June 2007)

²³⁵ See <http://www.ncbi.nlm.nih.gov/entrez/query/Gap/gap_tmpl/about.html> at 24 January 2007.

- all applicable laws, local institutional policies, and terms and procedures specific to the study's Data Access Policy will be followed;
- no attempt will be made to identify or contact individual study participants from whom data was obtained;
- data will not be sold or shared with third parties;
- the owner of rights in relation to the data (and the funding organisation) will be acknowledged in publications resulting from the analysis of the data;
- all NIH supported genotype/phenotype data and conclusions derived directly from the data will be licensed to any person for any use at no cost and with no restrictions; and
- an annual research progress report will be submitted to the study's data access committee.

Example 13 - NCBI: Entrez

4.109 Entrez is an online gateway established by the NCBI, which enables cross searches to be carried out for sequences across thirty interlinked databases.²³⁶

4.110 Clicking on the “Disclaimer” tag leads to a web page describing the copyright status of NCBI material.²³⁷ This copyright statement is applicable to use of both the NCBI Entrez site and the NCBI dbGaP site.

Ownership – Entrez

4.111 The web page does not put forward any position on ownership, but goes onto state that in relation to information incorporating non-government material:

All persons reproducing, redistributing or making commercial use of this information are expected to adhere to the terms and conditions asserted by the copyright holder. Transmission or reproduction beyond that of fair use as defined in the copyright laws requires the written permission of the copyright owners.

4.112 It is apparent that Entrez does not own the data and that the contributors of non-government material retain ownership rights in relation to the data.

Control - Entrez

4.113 The web page states that any use of the copyrighted material is subject to the terms and conditions of use established by the journal or publisher.

²³⁶ See <<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>>.

²³⁷ See <<http://www.ncbi.nlm.nih.gov/About/disclaimer.html>>.

Access and Use – Entrez

4.114 The Entrez web page states that government information from the site is in the public domain and any public domain information on the web pages may be freely distributed and copied. Non-government material is subject to the conditions and restrictions imposed by the parties who own rights in the data.

Example 14 - The California Stem Cell Initiative (CIRM)

4.115 The California Stem Cell Research and Cures Initiative, also called Proposition 71, was approved by Californian voters in November 2004.²³⁸ It provided \$3 billion in funding for stem cell research at Californian universities and research institutions.²³⁹ The Initiative established the California Institute of Regenerative Medicine (CIRM), a state agency which allocates the funding grants for stem cell research.²⁴⁰

4.116 CIRM's Intellectual Property Policy for Non-Profit Organizations (IPPNO) endorses the broad principle that all types of intellectual property created with CIRM funding should be made "as freely available as possible in the public domain."²⁴¹ "Intellectual property" is defined in the IPPNO as including, but not limited to:

Data, databases, biomedical materials, patents, scientific articles, research tools and software [protected by] patents, copyrights and trade secret information.²⁴²

4.117 However, while encouraging the broad dissemination of CIRM-funded intellectual property:

CIRM has acknowledged competing interests that might limit such sharing, such as bringing scientific advances to the public through commercialisation and providing a financial benefit to the State of California through revenue sharing. Indeed, the text of Proposition 71, the initiative that created CIRM, explicitly sets forth these conflicting interests.²⁴³

4.118 CIRM requests the final manuscripts of all scientific articles supported in whole or in part by CIRM to be deposited into an approved repository, preferably PubMed Central.²⁴⁴ However, there is no requirement (or request) that data be deposited into a repository to be accessed by the general public. Instead, the IPPNO

²³⁸ See the text of Proposition 71 at <<http://www.cirm.ca.gov/prop71/pdf/prop71.pdf>> at 24 January 2007.

²³⁹ <<http://www.cirm.ca.gov/about/default.asp>> at 24 January 2007.

²⁴⁰ <<http://www.cirm.ca.gov/about/default.asp>> at 24 January 2007.

²⁴¹ *IP Policy for Non-Profit Organizations* (2006) 26, 28

<<http://www.cirm.ca.gov/policies/pdf/IPPNO.pdf>> at 24 January 2007. This policy was approved by the Independent Citizens Oversight Committee (ICOC) (the governing board for the CIRM) on 2 October 2006.

²⁴² See *IP Policy for Non-Profit Organizations*, Approved by the Independent Citizens Oversight Committee (ICOC) (the governing board for the CIRM) Chapter III paragraph J. on 2/10/06 <<http://www.cirm.ca.gov/policies/pdf/IPPNO.pdf>> at 24 January 2007, p26

²⁴³ R Eisenberg and A Rai, "Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California's Stem Cell Initiative" (2006) 21 *Berkley L.J.* 1187 at 1188.

²⁴⁴ See *IP Policy for Non-Profit Organizations* (2006) 24

<<http://www.cirm.ca.gov/policies/pdf/IPPNO.pdf>> at 24 January 2007. For more on PubMed Central, see <<http://www.pubmedcentral.nih.gov/>> at 24 January 2007.

provides that data should be provided to researchers on request and on certain conditions. This is discussed under “Control” and “Access and Use” below.

Ownership - CIRM

4.119 Ownership rights in relation to data in CIRM-funded projects remain with the funded researcher or research organisation.²⁴⁵ However, there is a requirement under the IPPNO that the researcher allows CIRM to:

reproduce, publish or otherwise use the copyrighted material for public benefit so long as such use is not in violation of any copyright held by another party.²⁴⁶

4.120 This licence must be granted for free and the researcher must notify CIRM of patents involving CIRM-funded inventions, but not the development of copyright works under a CIRM grant.

Control - CIRM

4.121 CIRM exercises some limited control over data generated by funded researchers. Although ownership of rights in relation to the data and other research material remains with the researcher, the IPPNO provides that researchers will not be funded unless:

- they grant to CIRM a licence to use the data for public benefit; and
- they grant to persons who request the material a licence to use the material for research purposes in California.

4.122 The control exercised by CIRM over research-generated data is apparent from the outset in its conditions for research funding. Yet aside from the conditions imposed by CIRM in relation to granting the above licences, researchers are not restricted from granting other licences for purposes and in circumstances other than those prescribed in the IPPNO.

Access and Use - CIRM

4.123 CIRM requires funded researchers to grant a licence for use of materials (i.e. data) described in published scientific articles to a person requesting the material for research purposes in California within 60 days of the request.

4.124 The obligation imposed by CIRM on funded researchers is limited in the following ways:

- the requirement only applies to materials “described in a publication.” There is no requirement to share data acquired by a researcher that is not described in any publications written by the researcher; and

²⁴⁵ *IP Policy for Non-Profit Organizations* (2006) 23

<<http://www.cirm.ca.gov/policies/pdf/IPPNPO.pdf>> at 24 January 2007.

²⁴⁶ *IP Policy for Non-Profit Organizations* (2006) 23

<<http://www.cirm.ca.gov/policies/pdf/IPPNPO.pdf>> at 24 January 2007.

- material must only be licensed if the following conditions are satisfied:
 - (a) the material must be used for research purposes in California; and
 - (b) the material cannot be used for purely commercial purposes.

4.125 CIRM allows the researcher to impose costs upon the requestor for provision of the materials, but these costs are limited to the costs of production and distribution of the requested material.

Example 15 - Malaria Genomic Epidemiology Network (MalariaGEN)

4.126 MalariaGEN is an international research consortium whose aim is to use genomic epidemiology to identify molecular mechanisms of protective immunity against malaria.²⁴⁷ The MalariaGEN consortium brings together clinical researchers, epidemiologists, immunologists, genome researchers and statisticians from 20 countries in Africa, Asia, Europe and North America. The project will utilise clinical data from tens of thousands of individuals and generate billions of genetic data points.

4.127 In order to achieve the main goals of MalariaGEN a number of key scientific targets have been identified including:

- building a global epidemiological infrastructure for genomic association mapping;
- collecting DNA and clinical data from individuals with different manifestations of malaria infection;
- characterising genetic variation in malaria-endemic populations;
- identifying genetic variants that protect against severe malaria;
- identifying genetic variants that protect against severe malaria; and
- defining immunological mechanisms in genetic variants that protect against severe malaria.

4.128 Two fundamental principles were proposed to operate in determining policies and principles about data sharing and intellectual property for the MalariaGEN consortium. These are:

- (a) impediments to innovation should be minimised; and
- (b) research outcomes should be made as widely accessible as possible.

Ownership - MalariaGEN

4.129 Without having viewed the exact terms of the MalariaGEN agreements between the consortium, researchers and third parties, it appears that upon submission

²⁴⁷ <<http://www.malariagen.net>>.

of samples or data to MalariaGEN, the contributing researcher retains all ownership rights in relation to the samples or data.

4.130 Under the MalariaGEN draft Data Release Policy,²⁴⁸ ownership rights in relation to the data will remain with the institution or researcher who has generated the data even after it has been entered into the MalariaGEN database.

4.131 As a condition of submission, the contributing researcher grants other researchers in the consortium a revocable, non-exclusive licence to access the samples or data for the sole purpose of undertaking consortium experiments.

Control - MalariaGEN

4.132 As a condition of accessing the samples and data within the MalariaGEN consortium, researchers in the consortium are required to make available to the contributing researcher all primary genotype data as soon as it is generated through the web-based system. The contributing researcher also retains all access rights to their clinical databases and can limit access to shared data within the consortium.

4.133 Under the MalariaGEN draft Data Release Policy, the data can only be used for research purposes.

Access and Use – MalariaGEN

4.134 MalariaGEN are currently in the process of developing a legal framework to clarify some of the data sharing issues and other issues facing the consortium. Issues being considered include:

- *when to release data generated across the whole consortium* – the consortium may decide that there are reasons to delay the general internal release of data. This would allow the researchers who have invested time and effort in collecting samples and clinical data the first opportunity to perform analyses. In the case of the MalariaGEN consortium, a period not exceeding a few months has been suggested as appropriate;
- *the access of data and samples from outside the consortium* – this issue includes an assessment of the ethical implications which need to be made prior to the release of samples and data. For example, the MalariaGEN consortium involves the specific research area of genomic epidemiology. This genetic data may in certain circumstances indirectly identify individuals within a well-defined study population;
- *intellectual property considerations* – for example, the capacity to obtain a patent may be jeopardised by the premature release of information to the public; and

²⁴⁸ See Draft Data Release Policy, p2

<http://www.malariagen.net/tennis/app/malwik_page?frm=malwik_page&hdcon=information&mw_do_main=2&page_mode=1&page_number=506&last_rev_was=83&image_id=&page_to_order=&location_tree_admin_mode=>

- *attribution* – the release of research data and samples could in some circumstances undermine the proper assignment of credit to the major contributors.

4.135 Under the MalariaGEN draft Data Release Policy, any release of data is “subject to appropriate protections”. The draft Data Release Policy does not clarify what this statement means, but it probably extends to protection of ownership rights in relation to data. The policy goes on to state that data that cannot be released (which may refer to third party rights in the data) must not be contributed to the MalariaGEN database.²⁴⁹

4.136 The draft policy proposes that access will be made subject to a “legally-binding” agreement that sets out conditions and restrictions. This may include an obligation that if use of the data leads to development of intellectual property that could support health benefits in the developing world, the owner of intellectual property developed from such data agrees to licence it on a reasonable basis for use in the developing world.²⁵⁰

Example 16 - UniProtKB/Swiss-Prot Protein Knowledgebase

4.137 UniProtKB/Swiss-Prot²⁵¹ is a protein knowledgebase established in 1986 and maintained since 2003 by the UniProt Consortium. The UniProt Consortium is a collaboration between the Swiss Institute of Bioinformatics and the Department of Bioinformatics and Structural Biology of the Geneva University, the European Bioinformatics Institute (EBI) and the Georgetown University Medical Centre's Protein Information Resource.

4.138 UniProtKB/Swiss-Prot, together with UniProtKB/TrEMBL, its computer-annotated supplement, constitutes the UniProt Knowledgebase (UniProtKB), a major project of the UniProt consortium. UniProtKB/Swiss-Prot and UniProtKB/TrEMBL give access to all publicly available protein sequences.

4.139 The UniProt Knowledgebase consists of sequence entries. Sequence entries are composed of different line-types, each with their own format.²⁵²

Ownership - UniProtKB

4.140 The UniProtKB website states that the protein data is sourced from:

scientific publications, that report new sequence data, and/or review articles to periodically update the annotations of families or groups of proteins. We also make use of external experts, who have been recruited to send us their comments and updates concerning specific groups of proteins.²⁵³

²⁴⁹ Malariagen, *Draft Data Release Policy*, 3

<http://www.malariagen.net/tennis/app/malwik_page?frm=malwik_page&hdcon=information&mw_do_main=2&page_mode=1&page_number=506&last_rev_was=83&image_id=&page_to_order=&location_tree_admin_mode=> at 22 May 2007.

²⁵⁰ Ibid.

²⁵¹ See <<http://au.expasy.org/sprot/>> and <<http://www.expasy.uniprot.org/>>.

²⁵² <<http://www.ebi.ac.uk/swissprot/>>.

²⁵³ <http://au.expasy.org/sprot/sprot_details.html> at 23 January 2007.

4.141 The data held within UniProtKB includes protein sequences, current knowledge on each protein, core data (sequence data; bibliographical references and taxonomic data) and further annotation. The database is organised through a web interface that displays the data associated with each protein sequence. It is unclear who owns rights in relation to any of this data. However, examination of a data record within the database reveals the copyright heading:

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>. Distributed under the Creative Commons Attribution-NoDerivs Licence.²⁵⁴

4.142 When clicked, the link within the copyright heading displays the following statement:²⁵⁵

We have chosen to apply the Creative Commons Attribution-NoDerivs Licence to all copyrightable parts of our databases. This means that you are free to copy, distribute, display and make commercial use of these databases, provided you give us credit. However, if you intend to distribute a modified version of one of our databases, you must ask us for permission first.

DISCLAIMER

We provide all data on an 'as-is' basis. We make no warranties regarding the correctness of the data, and disclaim liability for damages resulting from its use. We cannot provide unrestricted permission regarding the use of the data, as some data may be covered by patents or other rights.

4.143 The presence of the disclaimer makes it unclear as to precisely who owns rights in relation to the UniProtKB data, notwithstanding the copyright statement.

4.144 EBI is one of the UniProtKB collaborative partners and provides “Terms of Use” for access to its public databases (including UniProtKB).²⁵⁶ These terms specify that whilst no restrictions are imposed on the use or redistribution of data, the original data may be subject to ownership rights. Users are responsible for ensuring that, in exploiting the data, their acts do not infringe rights of third parties. This indicates that the contributors of data to UniProtKB will retain ownership rights in relation to the data.

Control - UniProtKB

4.145 It is unclear who owns rights in relation to the UniProtKB data, and therefore it is difficult to comment on control issues. The EBI Terms of Use indicate that the users of UniProtKB are directly responsible to any owners of rights in relation to the data for its control and exploitation.

Access and Use - UniProtKB

4.146 The UniProt Consortium states that the public databases maintained by UniProt Consortium members are freely available to any individual and for any purpose.

²⁵⁴ See UniProtKB/Swiss-Prot entry *P66513* <<http://au.expasy.org>> at 23 January 2007.

²⁵⁵ <<http://www.uniprot.org/terms>>

²⁵⁶ See <<http://www.ebi.ac.uk/Information/termsuse.html>>.

4.147 The UniProtKb open access system has been described as operating on an “honour system” on the basis that:

- the user community is so small and so accurately monitored by electronic tracking that non-compliance would risk unacceptable costs in loss of reputation, peer pressure and possible denial of privileges; and
- administrators can rely on default rules derived from the European Communities Database Directive to make every user who fails to comply with its access and use conditions liable for infringement.²⁵⁷

4.148 However, the UniProt Consortium does not grant unrestricted permission regarding the use of data. The UniProt Consortium:

Places no restrictions on the use or redistribution of the data available. However, some of the original data may be subject to patent, copyright or other IP rights claimed by third parties. It is the responsibility of users ... to ensure that exploitation of the data does not infringe the rights of such third parties ...²⁵⁸

Example 17 - BRENDA

4.149 BRENDA is a comprehensive database of molecular and metabolic information on more than 83,000 enzymes from 9,800 organisms.²⁵⁹ BRENDA is maintained and developed at the institute of Biochemistry at the University of Cologne, Germany and is distributed by the corporation, Biobase. The purpose of the database is to provide a systematic collection of functional data for gene products. Genome data can be difficult to acquire, because it is often broadly distributed among journals from different fields and can be subject to experimental conditions.²⁶⁰ BRENDA enables genome data to be more easily collected, analysed and applied in different scientific fields, especially medicine.

4.150 Data contained in the BRENDA database is extracted directly from primary literature by scientists holding a degree in Biology or Chemistry. Each dataset on a classified enzyme is checked manually by one biologist and one chemist.²⁶¹

Ownership - BRENDA

4.151 The copyright notice on the BRENDA website provides that “copyright of the printed version is held by Springer Publishers.”²⁶²

²⁵⁷ J Reichmann and P Uhler, ‘A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment’ (2003) 66 *Law and Contemporary Problems* 315, 323

<<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

²⁵⁸ <<http://www.ebi.ac.uk/Information/termsofuse.html>> at 25 January 2007.

²⁵⁹ <<http://www.biobase-international.com/pages/index.php?id=brendadatabase>> at 24 January 2007.

²⁶⁰ See <<http://www.brenda.uni-koeln.de/information/introduction.php4>> at 24 January 2007.

²⁶¹ See <<http://www.brenda.uni-koeln.de/information/introduction.php4>> at 24 January 2007.

²⁶² <<http://www.brenda.uni-koeln.de/index.php4?page=information/introduction.php4>> at 24 January 2007.

4.152 Springer is the world's second largest specialist publisher in the science, technology and medicine sector and the top specialist information provider in German-speaking countries.²⁶³

4.153 The BRENDA copyright notice does not define what is meant by "printed version". It may mean the printed version of the data itself, or it may mean the printed primary scientific literature containing the data.

Control - BRENDA

4.154 The BRENDA database creators and distributors, including Biobase, exercise control over data in the BRENDA database by imposing restrictions on its use. The primary restriction is the non-commercial use term under "Licence Information", which also prevents the alteration of files obtained from BRENDA by users.²⁶⁴

4.155 Permission must be acquired from the copyright owner (Springer Publishers) for users to further distribute a copy of any BRENDA file. This therefore gives Springer Publishers some control over the data in BRENDA.

4.156 The scientists who have produced the data do not appear to have any control over how it is used once the data is deposited into the database.

Access and Use - BRENDA

4.157 BRENDA is available free of charge to academic, non-profit users. These users are granted a limited licence to access and use the database and the data therein for personal, non-commercial uses. The licence terms are set out in the copyright notice:

Limited licence is granted to individuals accessing BRENDA and its component documents and/or files for the following personal, non-commercial uses:

- Retrieving, Printing, or Electronically Storing a Copy of Any BRENDA Document or FILE
 - Establishing a Link or Links to Any BRENDA Document or File Mounted on this Server
- Individuals accessing this document and its component documents and/or files are NOT GRANTED licence to:
- Alter a Copy of Any Retrieved, Printed, or Stored BRENDA Document or File from this Server
 - Distribute a Copy (Electronic or Otherwise) of Any BRENDA Document or File from this Server without Permission from the copyright owner.
 - Charge for a Copy (Electronic or Otherwise) of Any BRENDA Document or File.²⁶⁵

4.158 Any commercial use of BRENDA, inclusion of its components into other databases or redistribution of BRENDA requires a commercial licence. Commercial licences are obtained from Biobase.²⁶⁶ There are no commercial licences accessible from the Biobase website, so it is unclear whether Biobase provides a standard form

²⁶³ <<http://www.springer-sbm.com/index.php?id=11893&L=0>> at 25 January 2007.

²⁶⁴ See <<http://www.brenda.uni-koeln.de/index.php4?page=information/introduction.php4>>.

²⁶⁵ <<http://www.brenda.uni-koeln.de/index.php4?page=information/introduction.php4>> at 24 January 2007.

²⁶⁶ See <<http://www.biobase-international.com/pages/index.php?id=company>> and <<http://www.biobase-international.com/pages/index.php?id=brendadatabase>> at 24 January 2007.

commercial licence to users or whether the terms of the licence are negotiated on a case-by-case basis with each individual commercial user.

Example 18 - Nanowerk

4.159 Hawaii-based Nanowerk maintains a database devoted to materials useful in nanoscience and nanotechnology research.²⁶⁷ Nanoscience is:

The study of phenomena and manipulation of materials at atomic, molecular and macromolecular scales, where properties differ significantly from those at a larger scale.²⁶⁸

4.160 The Nanomaterial Database maintained by Nanowerk allows users to search for nanoparticle items and then link directly to the item's manufacturer(s) in order to obtain the item for cost. Nanowerk also offers discounts of up to 10% on manufacturer's prices.

Ownership - Nanowerk

4.161 Condition 1 of Nanowerk's Terms of Use provides that "all proprietary content" on the Nanowerk website is "owned or licensed by us".²⁶⁹ It is unclear whether "proprietary content" includes data, although the terms do expressly refer to Nanowerk's property in "product data sheets". If this term does not extend to submitted data and datasets, then ownership rights in relation to data will remain with the data submitters or the website linked manufacturers who have provided data regarding nanoparticle items.

Control - Nanowerk

4.162 The Nanomaterial Database is freely accessible to all web users. However, nanomaterials accessed through the database must usually be acquired from a manufacturer and paid for. Additionally, Nanowerk strictly controls use of website content. Condition 1 in the Terms of Use prohibits users from:

- copying or reproducing;
- framing;
- publishing or republishing;
- modifying;
- downloading or uploading;
- decoding;
- transmitting; or
- distributing website content, whether partially or completely, without the advanced written authorisation of Nanowerk.

Access and Use - Nanowerk

4.163 Nanowerk provides an extremely limited licence under Condition 1 to users to make a single copy of website content that is for:

²⁶⁷ See <http://www.nanowerk.com/phpscripts/n_dbsearch.php> at 25 January 2007.

²⁶⁸ <http://www.nanowerk.com/n_nanomaterials_1.html> at 25 January 2007.

²⁶⁹ <http://www.nanowerk.com/n_terms.html> at 25 January 2007

- personal and private use; and
- non-commercial use.

Example 19 - Mars Climate Database (MCD)

4.164 The Mars Climate Database is a database of statistics describing the climate and environment of the Martian atmosphere.²⁷⁰ It is a collaborative project of European researchers based in the UK, France and Spain:

It is constructed directly on the basis of output from multannual integrations of a Global Climate Model (GCM) developed by Laboratoire de Météorologie Dynamique du CNRS, France in collaboration with the University of Oxford, UK, the Instituto de Astrofísica de Andalucía, Spain, SA, France with support from the European Space Agency (ESA) and Centre National d'Etudes Spatiales (CNES).²⁷¹

4.165 The MCD copyright notice states:²⁷²

COPYRIGHT

The data calculated by the Martian Global Circulation Model and contained in the Mars Climate Database are made available to the public on the condition that we make no representations or warranties regarding the reliability or validity of the data nor the use to which such data should be put, disclaim any and all responsibility for any errors or inaccuracies in the data and bear no responsibility for any use made of this data by any party.

Scientific use of the database is freely allowed provided that the origin of the data is correctly quoted in all publications and that we are kept informed of usage and developments.

This database may not be put to any commercial use without specific authorization.

Ownership - MCD

4.166 The only express reference to copyright ownership appears on the web site under “General description of the database”.²⁷³ This provides:²⁷⁴

270

<http://www-mars.lmd.jussieu.fr/mars/info_web_v4/node1.html> at 25 January 2007.

271

Ibid.

272

<<http://www-mars.lmd.jussieu.fr/mars/access.html>> at 22 May 2007.

273

<<http://www-mars.lmd.jussieu.fr/mars.html>> at 23 May 2007.

274

< http://www-mars.lmd.jussieu.fr/mars/info_web_v4/node6.html> at 23 May 2007.

This document was generated using the 324HLaTeX2HTML translator Version 2002-2-1 (1.70)

Copyright © 1993, 1994, 1995, 1996, 325HNikos Drakos,
Computer Based Learning Unit, University of Leeds.
Copyright © 1997, 1998, 1999, 326HRoss Moore,
Mathematics Department, Macquarie University, Sydney.

4.167 It is not clear whether this notice applies to the entire database or just the written section describing the database. If the latter, then the wording of the copyright notice would indicate that the parties who have generated the database (and possibly the parties who have funded the database) are the copyright owners. These parties are:

- Laboratoire de Météorologie Dynamique du CNRS, France (the primary developer);
- University of Oxford, UK;
- the Instituto de Astrofisica de Andalucia, Spain;
- SA, France;
- the European Space Agency; and
- Centre National d'Etudes Spatiales.

Control - MCD

4.168 Before being able to enter the Mars Climate Database, users are directed to a screen displaying the copyright notice. They are not required to click anything to assent to its terms, but this does indicate an intention on the part of MCD that users read the copyright notice.

4.169 The copyright notice provides that scientific use of the database is allowed, subject to the following requirements:

- the origin of the data must be correctly quoted in all publications; and
- MCD must be kept informed of all usage and developments of the data.

4.170 Additionally, the copyright notice provides that users must not undertake commercial use of the database without specific authorisation.²⁷⁵ This implies that MCD data can only be freely used for non-commercial purposes. It is not clear from whom authorisation must be sought or how authorisation is sought for commercial use of the database.

Access and Use - MCD

4.171 The MCD database has been openly available to the community since 1999.²⁷⁶ The online database is recommended for “moderate use only.”²⁷⁷ A DVD

²⁷⁵ <<http://www-mars.lmd.jussieu.fr/mars/access.html>> at 25 January 2007.

²⁷⁶ Ibid.

ROM version of the database is made freely available for “intensive or precise work.”²⁷⁸

Example 20 - Western Australian Genetic Epidemiology Resource (WAGER)

4.172 The Western Australian Genetic Epidemiology Resource (WAGER) is an initiative of the University of Western Australia.²⁷⁹ It is designed to integrate disease-specific clinical, epidemiological and genetic resources available in Western Australia (WA) with biospecimen banks and with the WA Data Linkage System (WADLS).²⁸⁰

4.173 The WADLS is maintained by the Data Linkage Unit (DLU) - a collaboration between the Information Collection and Management Branch of the WA Department of Health, the Centre for Health Services Research at the University of WA, the Division of Health Sciences at Curtin University of Technology and the Telethon Institute for Child Health Research. The WADLS consists of links within and between the State’s seven core population health datasets which contain data collected over a period of 35 years relating to health events for individuals in WA. Each link in the system is associated with a record in a range of core data sets, including: the Hospital Morbidity Data System, Mortality records, the Mental Health Information System, the Cancer Registry and the Midwives Notification System (collectively, the “Core Data”). Essentially, WAGER provides the infrastructure to enable researchers to use the Core Data and develop Clinical Studies Data, which is then linked to the existing WADLS.²⁸¹

4.174 The resulting composite entity of Core Data and Clinical Studies Data will comprise one of the largest and best-characterised population-based enabling facilities for epidemiological and genetic epidemiological research in the world, and will considerably enhance Australia’s medical research capacity.

Ownership - WAGER

4.175 Data comprising the Core Data is collected by the WA government. The WA government is the owner of rights in relation to the Core Data, subject to any other arrangements relating to the generation of the data. It seems that rights in relation to the derivative Clinical Studies Data are owned by the research organisations that generate them. The WAGER Data Access Policy states:

As WAGER's Metadata databases enable the linking of data between studies and with the core WA Data Linkage System, researchers are asked to lodge a copy of any new data produced on WA research subjects as a result of collaborations enabled by WAGER with WAGER once the study is completed and the data has been checked and cleaned. WAGER considers that the data collected by researchers for a particular study belongs to the researchers. It will hold copies of the data in confidence and will not make a study's data available to other researchers

²⁷⁷ Ibid.

²⁷⁸ Ibid see also <http://www-mars.lmd.jussieu.fr/mars/info_web_v4/node1.html> at 22 May 2007

²⁷⁹ <<http://www.wager.org.au/access.html>>.

²⁸⁰ <<http://www.wager.org.au>>

²⁸¹ See “The Data Linkage System” at

<<http://www.populationhealth.uwa.edu.au/welcome/research/dlu/linkage/system>> and “Uses of Linked Data” at <<http://www.populationhealth.uwa.edu.au/welcome/research/dlu/linkage/uses>> at 27 March 2007.

unless it has permission from the original researchers concerned. If they so wish, the original researchers may place conditions on access to the data by third parties²⁸²

Control - WAGER

4.176 There is no express statement as to what control rights are granted to users or institutional ethics committees or the effect of WA legislation in relation to privacy, freedom of information or other relevant issues. The WAGER Data Access Policy states:

WAGER will facilitate access to individual WA clinical, epidemiological and/or biospecimen resources for the broader Australian research community. *In keeping with existing DLU policy*²⁸³ *access to any of the contributing datasets can not be guaranteed but requires the approval of each of the data custodians involved in the request.* While clear guidelines and protocols for access have been developed, external users will be required to enter into negotiations with individual custodians within WAGER for access that will necessarily involve defining authorship and cost recovery. It is however reasonable to anticipate that the infrastructure established under WAGER will greatly increase the opportunities for Australian researchers to access both large population based genetic epidemiology resources and smaller research datasets held by individual research groups in WA. Such collaborations already occur in an *ad hoc* fashion and the enhancements enabled by WAGER will simply ensure that future collaborations occurred in a systematic and efficient manner and involved the broadest range of potential users within Australia.²⁸⁴

Access and Use - WAGER

4.177 Restrictions apply both in relation to the parties that may obtain access to the Core Data and Clinical Studies Data and the uses that may be made of the data.²⁸⁵ Subject to the Data Access Policy – WA Data Linkage System,²⁸⁶ the owners of rights in relation to the Core Data and Clinical Studies Data grant a licence:

- to external users approved by an institutional ethics committee;
- to use Core Data and Clinical Studies Data only for a purpose approved by both the owners of rights in relation to that data and the institutional ethics committee.

4.178 The grant of a licence to use the Core Data and Clinical Studies Data is subject to the user agreeing to supply to WAGER:

A copy of any report, journal article, conference presentation or other publication that is generated from project data.²⁸⁷

²⁸² See <<http://www.wager.org.au/access.html>>

²⁸³ See <<http://www.populationhealth.uwa.edu.au/welcome/research/dlu/linkage>>

²⁸⁴ See <<http://www.wager.org.au/access.html>> at 27 March 2007.

²⁸⁵ See *Access and Application*

<http://www.populationhealth.uwa.edu.au/welcome/research/dlu/linkage/Access_and_Application> at 27 March 2007.

²⁸⁶ See *Data Access Policy*

<http://www.populationhealth.uwa.edu.au/__data/page/38779/DLU_Access_Policy_Apr_2006.pdf> at 27 March 2007.

²⁸⁷ See <http://www.populationhealth.uwa.edu.au/__data/page/38779/Terms_and_Conditions.pdf> at 24 January 2007.

Example 21 - The RettBase and the InterRett Databases

4.179 The Rett database and the InterRett database are funded by the International Rett Syndrome Association (IRSA), an organisation that is committed to understanding the causes and potential remedies of Rett syndrome (RS), a rare debilitating neurological disorder seen almost exclusively in females.

4.180 The mission of the IRSA is to:

Support and stimulate biomedical research that will determine...treatments and cures for Rett syndrome.²⁸⁸

4.181 To further this mission, the IRSA aims to:

Collect and disseminate accurate and objective information regarding the cause, identification, treatment, prediction, prognosis, analysis, prevention and cure of RS.²⁸⁹

4.182 The InterRett database was developed to enable information about RS to be collected and widely disseminated. It allows clinicians to deposit scientific data about RS into the online database, and also allows families with members suffering from RS to deposit data about their personal experiences of living with RS.

4.183 The Rett database seeks to publish mutations and polymorphisms and invites researchers and diagnostic laboratories to contribute their data to the database. The eventual goal is to discover phenotype/genotype correlations from such submitted data. This database is still in its development stages.

Ownership - IRSA

4.184 It appears that IRSA owns rights in relation to the data in the InterRett database, or it at least holds a licence from depositors to use and publish the gathered data. Depositors are required to complete a consent form when depositing data and answering questionnaires, which provides:

Research data gathered from the results of this study may be presented or published, provided that names are not used.²⁹⁰

4.185 Identifying information is removed from the questionnaires and data before inclusion in the database, to protect the confidentiality of the families concerned.

Control - IRSA

4.186 The InterRett database is available for free from the IRSA website. There do not appear to be any restrictions on who may access the database and what can be done with the data contained in it. In fact, the website boasts data collected from and used by people from all around the world, including the USA, UK, Spain, Argentina, China, Turkey, Belgium, Norway, India and Mexico.²⁹¹ While it is clear that the

²⁸⁸ <<http://www.rettssyndrome.org/content.asp?pl=433&contentid=459>> at 25 January 2007.

²⁸⁹ <<http://www.rettssyndrome.org/content.asp?pl=433&contentid=459>> at 25 January 2007.

²⁹⁰ <http://www.ichr.uwa.edu.au/rett/irsa/pages/consent/Info_ConsInterRett.pdf> at 29 January 2007.

²⁹¹ <<http://www.rettssyndrome.org/content.asp?pl=450&sl=482&contentid=536>> at 25 January 2007.

IRSA exercises control over the InterRett database itself, little control is asserted over the data contained in the database.

Access and Use - IRSA

4.187 IRSA and its affiliated organisation, Telethon Institute for Child Health Research, use the information collected to:

Carry out research projects looking for causes or ways to prevent particular childhood conditions...[and]...perform medical research and statistical analyses into the general health of populations to inform health providers and assist in government policy planning.²⁹²

4.188 For other users, InterRett is available online and provides access to data collected from families of children with Rett syndrome. The database is also interactive, allowing users to create graphs comparing and analysing data collected from different sources, for example a graph could be created showing the relationship between history of seizures and each of the common mutations associated with RS.

4.189 Researchers (with appropriate ethical approvals) will have access to de-identified InterRett data.²⁹³

Example 22 - Earth System Grid (ESG)

4.190 The ESG is one of several virtual collaborative environments that link distributed centres, users, models and data throughout the United States. Data for the project, which is expected to revolutionise understanding of climate change, is being collected from a wide range of sources, including ground and satellite-based sensors, computer-generated simulations and thousands of independent scientists.²⁹⁴

Ownership – ESG

4.191 The ESG is funded by the US Department of Energy. It is a collaborative project of:

- Argonne National Laboratory;
- Lawrence Berkeley National Laboratory;
- Lawrence Livermore National Laboratory;
- Los Alamos National Laboratory;
- National Center for Atmospheric Research;
- Oak Ridge National Laboratory; and
- University of Southern California/Information Sciences Institute.²⁹⁵

²⁹² <<http://www.ichr.uwa.edu.au/rett/irsa/pages/login.lasso>> at 29 January 2007.

²⁹³ See <http://www2.ichr.uwa.edu.au/rett/irsa/pages/data_access.html>.

²⁹⁴ See <<http://www.earthsystemgrid.org/home/home.htm>> at 23 May 2007.

²⁹⁵ Ibid.

4.192 However, copyright is expressed to be owned by the University Corporation for Atmospheric Research (UCAR):²⁹⁶

Portal Software version 4.4 © UCAR, all rights reserved.

Control – ESG

4.193 There is no statement on the ESG web site about how data that is accessed may be used. However, much of the data requires registration with ESG before it can be accessed. Registration requires the submission of a simple “request form” setting out the user’s contact details and the data which the user is interested in downloading.²⁹⁷ It does not contain terms of use. It is likely that if registration is accepted, conditions will be imposed on the user about how they may use the data provided. The position for data that does not require registration is unclear. It is possible that this data is made openly available for reuse, but it is equally possible that the copyright owner will need to be contacted for permission to reproduce the data.

Access and Use - ESG

4.194 A search bar allows dataset metadata to be searched on the ESG database.²⁹⁸ The web site also states that some portal functions require registration.²⁹⁹ A login page is provided where users must enter a username and password to access these portal functions.³⁰⁰ In reality, much of the data available on the ESG web site requires registration for access. These include:

- IPCC Working Group 1 Data;³⁰¹
- Parallel Ocean Program Data;³⁰²
- Community Climate System Model Data;³⁰³ and
- Community Land Model Data.³⁰⁴

Example 23 - Australian Social Science Data Archive (ASSDA)

4.195 The Australian Social Science Data Archive (ASSDA)³⁰⁵ was established in 1981 as a data archive for social, political and economic affairs and to make such data available for further analysis.³⁰⁶ ASSDA is located in the Research School of Social

²⁹⁶ Ibid.

²⁹⁷ <<http://www.earthsystemgrid.org/security/accountRequestPage.do>> (28 May 2007).

²⁹⁸ <<http://www.earthsystemgrid.org/data/searchPage.do?>> at 23 May 2007.

²⁹⁹ <<https://www.earthsystemgrid.org/security/loginout.htm?>>, see also

<<http://www.earthsystemgrid.org/index.jsp?>> and

<<http://www.earthsystemgrid.org/security/accountRequestPage.do>> at 23 May 2007.

³⁰⁰ <<https://www.earthsystemgrid.org/security/loginout.htm?>> at 23 May 2007.

³⁰¹ <<https://esg.llnl.gov:8443/index.jsp>> at 23 May 2007

³⁰²

<<http://www.earthsystemgrid.org/browse/browse.htm?uri=http://datagrid.ucar.edu/metadata/lanl/pop/thredds/lanl.pop.thredds>> at 23 May 2007.

³⁰³

<<http://www.earthsystemgrid.org/browse/browse.htm?uri=http://datagrid.ucar.edu/metadata/cgd/ccsm/thredds/ccsm.thredds>> at 23 May 2007.

³⁰⁴

<<http://www.earthsystemgrid.org/browse/browse.htm?uri=http://datagrid.ucar.edu/metadata/cgd/clm/thredds/clm.thredds>> at 23 May 2007.

³⁰⁵ See <<http://assda.anu.edu.au>>.

³⁰⁶ See <<http://assda.anu.edu.au/about.html>>.

Sciences at the Australian National University (ANU). Collected data includes data from participating institutions including universities, market research companies, and government organisations.

Ownership – ASSDA

4.196 The ASSDA website does not make any express statement as to ownership of rights in relation to the data.

4.197 However, the website contains a “Licence Form” for depositors of data which refers to depositors as “the owner of copyright in the material” who grants a non-exclusive licence to the ANU to publish the data.³⁰⁷ It also states that:

It is essential for users to check what access conditions were set by the owners of the data when they deposited their data with ASSDA.³⁰⁸

4.198 It follows that copyright and the ownership of rights in relation to the data remains with the data contributors.

Control – ASSDA

4.199 All users must sign an “Undertakings Form” and provide it to ASSDA before being allowed to access and use data.

4.200 Different undertakings apply depending on whether the user is a member of the Australian Consortium for Social and Political Research Incorporated and whether the data has been deposited under “Restricted Data” conditions.³⁰⁹ The terms in these forms are designed to be changed or removed, depending on any conditions or restrictions that may be placed by a depositor on their data.

4.201 For instance, there is an undertaking to use the material for statistical purposes only. Another undertaking provides that material will not to be used for non-statistical or commercial or financial gain without the express written permission of the Data Archive Manager.³¹⁰

Access and Use – ASSDA

4.202 The website states that copies of all datasets acquired through ASSDA's own acquisition activities are available to members of the International Federation of Data Organisations and other interested users.³¹¹

4.203 ASSDA goes onto designate three levels of access:

- Unrestricted Access: this data is designated by the depositor to be provided to any person who provides the correct undertakings;

³⁰⁷ See <http://assda.anu.edu.au/forms/ASSDA_Deposit_Licence.pdf>.

³⁰⁸ See <<http://assda.anu.edu.au/data.html>>.

³⁰⁹ See <<http://assda.anu.edu.au/data.html>>.

³¹⁰ See <<http://assda.anu.edu.au/forms/GeneralUndertakingACSPRI.pdf>>.

³¹¹ See <<http://assda.anu.edu.au/about.html>>.

- Notification of Access: this data is designated by the depositor to be provided to any person who provides the correct undertakings, subject to the Data Archive Manager notifying the depositor regarding its supply; and
- Depositors Permission: this data can only be accessed if the depositor provides written permission to ASSDA to allow access. If a depositor fails to respond to an ASSDA notice requesting such access within 30 days, consent to access is deemed to have been given.³¹²

4.204 Data will be made available, provided that the undertakings and applicable access conditions and restrictions pertaining to data have been adhered to and processed.

Example 24 - Marine Themes (MT)

4.205 Marine Themes is an Australian-hosted database holding the world's largest internet based stock image library dealing with all aspects of marine wildlife.³¹³

Ownership – MT

4.206 The MT website contains several statements as to copyright and ownership:

- the database “is intended for commercial image buyers only. This site contains no royalty free images. All images are copyright”;³¹⁴
- “All text on the Knowledge Archives pages are copyright protected: © 2006 marinethemes.com/Kelvin Aitken. All images in the Knowledge Archives are copyright protected by the photographer stated on each page. All rights reserved”;³¹⁵
- the “Frequently Asked Questions” page for photographers states that:
 - a contract should be entered into between the photographer and Marine Themes as the photographer’s “agent” and “legal representative” and that the photographer has the option to give Marine Themes either exclusive or non-exclusive worldwide rights to the photographer’s images.³¹⁶ (There is no example of such a contract on the website); and
 - upon termination of a contract the photographer “will always own [their] original photographs as well as the copyright to those photographs and any scans supplied by [the photographer, but] the scans made by Marine Themes will remain the property of Marine Themes and unless other arrangements are made, will be destroyed on termination of [the photographer’s] contract [with Marine Themes].”³¹⁷

³¹² See < http://assda.anu.edu.au/forms/ASSDA_Deposit_Licence.pdf.

³¹³ <<http://www.marinethemes.com/>>.

³¹⁴ See <<http://www.marinethemes.com/indexFrame.html>>

³¹⁵ See < <http://www.marinethemes.com/indexFrame.html>>

³¹⁶ See < <http://www.marinethemes.com/indexFrame.html>>

³¹⁷ Ibid.

4.207 The site also contains two links to its “Copyright Notice” which states:³¹⁸

Copyright Notice

All images and text on this web site are protected by international copyright law. Breach of copyright is a criminal offence, not to mention dishonest and unfair to our photographers who risk life and income to capture the images presented in this site. You may use the photographs in the Marine Themes Stock Library database and website for rough layouts, comps, storyboards or in-house submissions or presentations only as long as the Marine Themes copyright watermark is not removed.

Any other use, including but not limited to inclusion in websites, any form of copying or publication or use as derivative works, is prohibited without prior written consent and payment of the stipulated licensing fee. This means that use of any image where payment is not made means that you do not have copyright release and you are in breach of copyright.

Breach of copyright will be vigorously pursued and prosecuted to the fullest extent of the law. Payment of a fee to use an image allows you to use that image only as specifically stated on the agreement and invoice and does not acquire for you any ownership, title or rights to the image or its copyright.

4.208 Reading all of these statements in context:

- photographers submit photographs to MT for approval and then enter into either an exclusive or non-exclusive licence with MT, allowing MT to make image scans from the photographs which are incorporated into the database;
- ownership and copyright of the original photographic material submitted to MT remains the property of the contributing photographer; and
- MT and/or the MT database manager seek to assert a property right in the physical scans made from the submitted photographic material, but not copyright in the scanned images. MT’s property in the physical scans is limited because MT either destroys the scans or transfers them to the photographer upon termination of the licence.³¹⁹

Control – MT

4.209 Users who wish to access the image database state their agreement with the copyright notice by clicking on “I Agree” button displayed under the copyright notice.³²⁰ In return for payment of the “stipulated licensing fee”, they are granted the right to use the image specified in the agreement and invoice but do “not acquire any

³¹⁸ See <<http://www.marinethemes.com/indexFrame.html>> and <<http://www.marinethemes.com/cumulus/catalogue.html>>

³¹⁹ See <<http://www.marinethemes.com/indexFrame.html>>

³²⁰ Ibid

ownership, title or rights to the image or its copyright”. A licensing fee is always payable by users at a “on a fair and reasonable rate” with the possibility of a reduced fee for non-profit organisations.³²¹ The copyright notice also grants a general, non-contractual permission to use certain photographs in the MT Stock Library database and website for “personal use for display in private homes only”³²² and for “rough layouts, comps, storyboards or in-house submissions or presentations” provided the MT “copyright watermark is not removed”. Any other use, including inclusion of the photographs in “websites, any form of copying or publication or use as derivative works” is expressly prohibited “without prior written consent and payment of the stipulated licensing fee”.

Access and Use – MT

4.210 It seems that MT sub-licenses the scanned images for access and use to third party users. The images can then be used in a variety of ways, subject to non-removal of the MT image watermark. The website does not expressly state whether such a sub-licence will be exclusive or non-exclusive. This may depend on the nature of each transaction and the non-exclusive or exclusive terms of the licence between the contributing photographer and MT.

Example 25 - Australian Ocean Data Centre Joint Facility (AODCJF) and the Australian Ocean Data Centre (AODC)

4.211 The Australian Ocean Data Centre Joint Facility (AODCJF) manages the Australian Ocean Data Centre (AODC) database and seeks to create a national data management system in relation to scientific ocean data through a network of partner agencies.³²³ Data is submitted by the venture partners whom include the Australian Institute of Marine Science, the Australian Antarctic Division, the Bureau of Meteorology, CSIRO Marine and Atmospheric Research, Geoscience Australia and the Department of Defence.³²⁴

4.212 AODCJF has adopted the National Service Improvement Framework (NSIF) structure as its management model³²⁵ - a tiered system of template collaboration agreements, principles and explanatory statements that describe how documents and tools for collaborative working arrangements between government agencies must be written and enforced.³²⁶

Ownership –AODC

4.213 The AODC’s website copyright notice under the “Privacy & Disclaimer” tag states that:³²⁷

³²¹ See <<http://www.marinethemes.com/indexFrame.html>> and

<<http://www.marinethemes.com/cumulus/catalogue.html>> (28 May 2007)

³²² See <<http://www.marinethemes.com/indexFrame.html>> (25 June 2007)

³²³ <<http://www.aodc.gov.au>>

³²⁴ See <<http://www.aodc.gov.au/index.php?id=19>>

³²⁵ <<http://www.nsif.gov.au>>

³²⁶ See <http://www.nsif.gov.au/index.php?node=207_206>

³²⁷ See <<http://www.aodc.gov.au/index.php?id=23>>

COPYRIGHT NOTICE

© Commonwealth of Australia 2006

This work is copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organization. Apart from any use as permitted under the *Copyright Act 1968*, all other rights are reserved.

4.214 This notice is in accordance with the Commonwealth government's standard copyright notice template for "Internet Sites".³²⁸ The question is whether the term "this material" in the AODC copyright notice extends to the actual data on the database. If it can be assumed that it does, then the next question is who owns copyright in relation to the submitted data. The website does not expressly state that the Commonwealth acquires copyright, but the Commonwealth Copyright Administration states that the Commonwealth claims copyright in "all material written by its employees".³²⁹ Since Commonwealth government agencies and departments are Commonwealth employees, data submitted to AODC which is capable of copyright protection will be the property of the Commonwealth.

Control – AODC

4.215 For use and exchange of data between the AODC partners, the "Tier 2 Statements of Intent" document under the NISF management system applies.³³⁰ Term 17 of these statements requires that:

- intellectual property must be addressed in inter-agency collaboration agreements;
- any pre-existing intellectual property will be identified and 'respected' under such agreements; and
- pre-existing intellectual property can only be used as authorised by the owner of the intellectual property or as permitted under law'.³³¹

4.216 The term "respected" is not defined but probably means compliance with the copyright of other parties, at the very least. This statement also means that the agency partners will have to obtain written authorisations from any owners of rights in relation to data, before they submit such data to AODC or before exchanging it between each other.

³²⁸ See the Commonwealth Copyright Administration
<http://www.ag.gov.au/www/agd/agd.nsf/Page/Copyright_CommonwealthCopyrightAdministration_CommonwealthCopyright> at 28 May 2007.

³²⁹ Ibid.

³³⁰ Tier 2 Statement of Intent '17. Intellectual Property'

<http://www.nisf.gov.au/index.php?node=255_221_209_206> at 18 May 2007.

³³¹ Ibid.

Access and Use - AODC

4.217 The website states an intention to deliver interoperability to the marine data community and to undertake projects to enable universities, marine science researchers and portal users and contributors to access and contribute to the AODC virtual national ocean data centre.³³²

4.218 The website does not expressly state whether such access will be for no cost or for an agreed fee. However, the copyright notice implies that use of the data in an unaltered form will be without charge, provided that it is used for non-commercial purposes.

Example 26 - The Australian Digital Thesis Program

4.219 The Australian Digital Thesis (ADT) Program is a collaborative project for a distributed database of digital versions of theses produced by postgraduate students.³³³ ADT digitises student theses as part of the deposit process and also digitises a number of commonly requested theses.³³⁴

4.220 The project members include universities from Australia and New Zealand and the collection of data is managed by the Council of Australian University Librarians (CAUL). All CAUL members are also members of the ADT.³³⁵

Ownership – ADT

4.221 The “Facts and Questions” page³³⁶ states that:

- the author of the thesis retains copyright, in most cases; and
- the author may also choose to place restrictions on access to their thesis.

4.222 An ADT 2005 Report states that the author’s permission for the republication of their thesis on ADT will still have to be obtained.³³⁷

4.223 This means that students who write theses will retain copyright in them, even after they are published on the ADT website.

4.224 Authors will need to be made aware of the implications of submission of their thesis with any commercial partners to the thesis and the author’s obligation to

³³² See <<http://www.aodc.gov.au/index.php?id=34>>

³³³ <<http://adt.caul.edu.au/>>

³³⁴ See B Fitzgerald et al, *Open Access to Knowledge Oak Law Project No.1: Creating a Legal Framework for Copyright Management of Open Access within the Australian Academic and Research Sector* (2006) Ch 6, [6.27]

³³⁵ See *Australian Digital Theses Program – Expansion and Redevelopment*, 7 August 2003 <<http://adt.caul.edu.au/downloads/arricProposal2003.pdf>>

³³⁶ See <<http://adt.caul.edu.au/about/faq/>>

³³⁷ See *Australian Digital Thesis program Expansion and Redevelopment Project Partnership Activities with ProQuest Information and Learning Trial of the Digital Commons Deposit and Repository Software Report*, 23 November 2005 at <<http://adt.caul.edu.au/downloads/ADTPQILTrialReport.pdf>>

obtain permission (before submission to ADT) to use any third party copyright that may be in the thesis.

4.225 In many member universities, it is a requirement of students completing a PhD that their thesis is made available through ADT. Each of the member universities is required to provide the theses created by their students through a server located within their institution.³³⁸

4.226 This means that a contributing student submits their thesis for ADT incorporation through their own university.³³⁹ Each of university member servers uses an identical database configuration, standards and metadata,³⁴⁰ which is then accessed by CAUL under a non-exclusive royalty free licence to access the theses metadata on the servers and then to harvest and copy such metadata into the ADT database repository.³⁴¹

4.227 ADT does not provide a standard submission form for use by its university members in accepting theses submissions from students. It was considered that as each university server site has different copyright, deposit and open access policies and regulations, a generic submission form would not be feasible.³⁴²

4.228 The process is dealt with by each member university directly with its students and each university site will have to be examined to determine each member's submission terms. The question remains open as to whether individual members seek a licence or an assignment of copyright from the submitting student and will have to be gauged against ADT's statements on copyright remaining with the author.

4.229 The "Home" page also contains two notices,³⁴³ the first states that :

Information on this website is copyright. Apart from fair dealing as permitted under the copyright law of your country, no material may be copied or further disseminated without the written permission of the legal holder of that copyright, 322HCouncil of Australian University Librarians.

4.230 This notice seems to assert copyright ownership in database information both by author students in their theses (the "legal holder of that copyright") and by CAUL.

³³⁸ See <<http://adt.caul.edu.au/about/aimsoverview/>>

³³⁹ See B Fitzgerald et al, *Open Access to Knowledge Oak Law Project No.1: Creating a Legal Framework for Copyright Management of Open Access within the Australian Academic and Research Sector* (2006) Ch 6, [6.28]

³⁴⁰ Ibid.

³⁴¹ See *ADT Program Copyright and Metadata Policies* – draft September 2002, at <http://www.caul.edu.au/adt/ADTmetadata_v2_2002.doc>

³⁴² See *Australian Digital Thesis Program Expansion and Redevelopment Project Partnership Activities with ProQuest Information and Learning Trial of the Digital Commons Deposit and Repository Software Report*, 23 November 2005, 4

³⁴³ See <<http://adt.caul.edu.au/>>

4.231 This is clarified by the second smaller notice, which states that copyright is held by CAUL.³⁴⁴ This link only leads to the CAUL homepage and not to any further statement on ownership.

4.232 In this context, it seems CAUL is only asserting copyright in the web text and pages of the ADT database and not in the actual theses data.

Control – ADT

4.233 Access restrictions on theses can be imposed:

- either by the student authors given their copyright in their work; or
- by project members.³⁴⁵

4.234 However, this right has to be weighed against ADT's intention to make theses available on a worldwide basis.³⁴⁶

Access and Use – ADT

4.235 ADT states that it seeks to improve access to and enhance the transfer of research data contained in theses by providing full text theses that are accessible worldwide.³⁴⁷

4.236 This is reinforced under the ADT Program draft Copyright and Metadata Policies, which state that members must allow_free access to the theses metadata by any users for personal, educational and research use and for international discovery purposes.³⁴⁸

Example 27 - Capture Wales Digital Storytelling Database - BBC.Co.UK³⁴⁹

4.237 Hosted by BBC Wales, the Capture Wales project is an online database showcasing digital stories made in Wales:

BBC Wales is working with digital storyteller Daniel Meadows from the Centre for Journalism Studies at Cardiff University. We run monthly workshops around Wales, working with members of the public to help them create their own digital stories.³⁵⁰

4.238 The digital stories are placed online, with some also shown on television. The digital stories may consist of photographs, other pictures and visual media, text, music and sound.

³⁴⁴ See <<http://adt.caul.edu.au/about/faq/>>

³⁴⁵ Ibid.

³⁴⁶ Ibid.

³⁴⁷ See <<http://adt.caul.edu.au/about/aimsoverview/>>

³⁴⁸ See <http://www.caul.edu.au/adt/ADTmetadata_v2_2002.doc>

³⁴⁹ See <<http://www.bbc.co.uk/wales/capturewales/>>

³⁵⁰ <<http://www.bbc.co.uk/wales/capturewales/about/>> at 24 May 2007

Ownership – Capture Wales

4.239 A link on the Capture Wales website takes users to the “Terms of Use”. These are the BBC Terms of Use that apply generally to the central BBC website (bbc.co.uk) and other BBC hosted sites.

4.240 Term 9 provides for contributions made to bbc.co.uk:³⁵¹

Where you are invited to submit any contribution to bbc.co.uk (including any text, photographs, graphics, video or audio) you agree, by submitting your contribution, to grant the BBC a perpetual, royalty-free, non-exclusive, sub-licenseable right and license to use, reproduce, modify, adapt, publish, translate, create derivative works from, distribute, perform, play, make available to the public, and exercise all copyright and publicity rights with respect to your contribution worldwide and/or to incorporate your contribution in other works in any media now known or later developed for the full term of any rights that may exist in your contribution, and in accordance with privacy restrictions set out in the BBC's 323HPrivacy Policy. If you do not want to grant to the BBC the rights set out above, please do not submit your contribution to bbc.co.uk.

4.241 This term provides for a licence to BBC and not an assignment, suggesting that the makers of the digital stories displayed on Capture Wales retain copyright ownership in their work. A distinction will need to be drawn between the actual digital stories, which are owned by the maker, and other content on the BBC website (such as graphics), which will be owned by BBC.

Control – Capture Wales

4.242 Term 3 of the BBC Terms of Use regulates use of bbc.co.uk:³⁵²

You may not copy, reproduce, republish, download, post, broadcast, transmit, make available to the public, or otherwise use bbc.co.uk content in any way except for your own personal, non-commercial use. You also agree not to adapt, alter or create a derivative work from any bbc.co.uk content except for your own personal, non-commercial use. Any other use of bbc.co.uk content requires the prior written permission of the BBC.

4.243 Term 3 is a restriction on the kind of uses that can be put to material on the BBC website. However, it also operates as a permission to use material for personal, non-commercial use only. Any use other than personal, non-commercial use requires prior written permission from BBC.

Access and Use – Capture Wales

4.244 There do not appear to be any limitations on who may access material on the Capture Wales database. Digital stories are grouped together and searchable according to broad themes. From there, it is simply a matter of clicking the story title or accompanying blurb to access and view the digital story. Viewing requires a Real Player plugin, which can be downloaded for free.

³⁵¹ See <<http://www.bbc.co.uk/terms/>> at 24 May 2007

³⁵² Ibid

Example 28 - Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)

4.245 PARADISEC provides a digital conservation facility for endangered materials from the Pacific region, that is, Oceania and East and Southeast Asia.³⁵³ PARADISEC is also a national repository for recorded material relating to indigenous cultures of regions in and around Australia.³⁵⁴ As well as preserving materials that may otherwise be lost, PARADISEC provides access to these materials to researchers and other authorised users.³⁵⁵ In preserving and providing access to materials in the repository, PARADISEC takes responsibility for safeguarding the interests and sensitivities of the indigenous people to whom the material relates.³⁵⁶ In 2006, the PARADISEC collection contained almost 4,000 items. An associated project, EthnoER, focuses on supporting secure and distributed collaborative research based on digital media and data repositories.³⁵⁷

4.246 PARADISEC is a consortium of four universities:

- University of Sydney;
- University of Melbourne;
- University of New England; and
- ANU.³⁵⁸

Ownership – PARADISEC

4.247 Contributors deposit material – typically as old as 50 years - with PARADISEC for conservation. Non-digital data (eg tapes and transcripts) are digitised. The “Deposit of Material” form, which must be completed by persons depositing material into the repository, states that only the “owner of material” or the “owner’s delegate” can enter into the deposit agreement and deposit material.³⁵⁹ However, “ownership” is defined in terms of rights in the tangible article only and expressly excludes intangible property and associated rights, whether in the form of intellectual property or traditional cultural knowledge:

The term ownership refers to ownership of the physical copy of the material being lodged with PARADISEC. It is not a wider claim to the intellectual property or ownership of any traditional knowledge, mythology, personal information or expression which relates to or derives from other objects, individuals or communities contained in the material lodged. If the material was written, photographed, drawn, recorded or filmed by you, then you are the creator and owner of the physical copy of the material, or if you have collected, found, or inherited the material you are the owner of the physical copy of the material and therefore you or your delegate are in a legal position to enter into this agreement.³⁶⁰

4.248 By restricting the meaning of “ownership” to rights in the tangible article, the deposit form clearly attempts to achieve a cultural balance between the indigenous

³⁵³ Ibid

³⁵⁴ See <<http://www.paradisec.org.au/PdsCdeposit.rtf>> at 28 May 2007

³⁵⁵ <<http://www.paradisec.org.au/about.html>> at 28 May 2007

³⁵⁶ See <<http://www.paradisec.org.au/PdsCdeposit.rtf>> at 28 May 2007

³⁵⁷ See <<http://ethnoer.unimelb.edu.au>> at 15 June 2007.

³⁵⁸ <<http://www.paradisec.org.au/about.html>> at 28 May 2007

³⁵⁹ <<http://www.paradisec.org.au/PdsCdeposit.rtf>> at 28 May 2007

³⁶⁰ Ibid

people who may claim rights in traditional knowledge and mythology and other persons who may record or capture the traditional knowledge and mythology on a tangible medium such as film, photographs or drawings. While legal possession of a tangible copy of the material is a pre-requisite for deposit of the material into the PARADISEC repository, deposit does not in itself correlate with, or amount to a representation about, ownership of intangible rights in that material. As Shadbolt et al explain in the APSR report, *Sustainable paths for Data-intensive Research Communities at the University of Melbourne* (2006):

For the PARADISEC collection there are moral and ethical issues associated with the deposit and ownership of materials. The overall mission is to ensure preservation of the item by the legal custodian who may not necessarily be the ‘owner’ of the IP or cultural IP of the item as such³⁶¹.

4.249 The deposit agreement acknowledges the existence of intellectual property rights in material by advising depositors to negotiate with any other person or body who has rights in the material to allow deposit of the material in PARADESIC on suitable terms.³⁶² The PARADISEC information leaflet (available on the website) also states that “normal copyright restrictions” apply to each item in the PARADISEC collection “as specified by the depositor and the performer”.³⁶³ This indicates that the copyright owner retains copyright related rights in materials deposited into the repository.

Control – PARADISEC

4.250 PARADESIC gives the depositor of material a certain amount of control over the material included in the database. Upon deposit, the depositor completes a “Deposit of Material” form where they indicate what restrictions they want imposed upon access to their material. The depositor selects:

- free access as specified in the PARADISEC “Conditions of Access” form;
- access by permission of depositor only; or
- access on other terms as specified by the depositor.³⁶⁴

4.251 PARADESIC will not hold material on permanent closed access.³⁶⁵

4.252 The depositor is also given the option of preventing catalogue information (metadata) about their collection being publicly listed, although PARADESIC discourages this practice.³⁶⁶

³⁶¹ A Shadbolt, D van der Kniff, E Young and L Winton, “Sustainable paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories”, APSR, August 2006 at p 27.

³⁶² <<http://www.paradisec.org.au/PdsCdeposit.rtf>> at 28 May 2007

³⁶³ <http://www.paradisec.org.au/PARADISEC_PR04.pdf> at 28 May 2007

³⁶⁴ <<http://www.paradisec.org.au/PdsCdeposit.rtf>> at 28 May 2007

³⁶⁵ Ibid

³⁶⁶ Ibid

4.253 The PARADISEC “Conditions of Access” form imposes the following general conditions on persons accessing the data collections:

- the depositor’s permission for copying must be obtained if the depositor requires this in the “Deposit of Material” form;
- data may be used for private scholarly research or educational purposes only;
- data may not be used for profit (although publication in a scholarly context of analyses or interpretations derived from the data is permitted);
- any work that is based whole or in part on analyses of the data must contain an acknowledgement of both the original depositors of the data and PARADISEC;
- data must not be copied in whole or in part except for security purposes or personal use;
- data must not be distributed to third parties or published or reproduced in any way;
- access to the data may only be given to persons directly associated with the user or working under the user’s control, and these persons must provide signed undertakings not to use the data except in connection with the user’s academic or educational purposes and not to give others access to the data;
- undertakings must be made available to PARADISEC on request; and
- where substantial revisions or reformatting of the data is carried out as part of research, the user must inform PARADISEC and the original depositor (if this is a condition of deposit) of the nature of the revisions and must deposit the revised version in the PARADISEC database.³⁶⁷

Access and Use - PARADISEC

4.254 Access to the PARADISEC database is only provided to users who have completed the “Conditions of Access” form. This form must be printed out, signed and returned by mail to the PARADISEC Project Manager.³⁶⁸ When the signed form is received by PARADISEC, the user is provided with a username and password login to access the database.³⁶⁹

4.255 Access to content metadata from the PARADISEC catalogue is unrestricted and can be freely searched by both registered and unregistered users,³⁷⁰ subject to:

- any conditions nominated by the depositor in the “Deposit of Material” form; and
- PARADISEC being given authority by depositors in the “Deposit Form” to decide questions of access and use in order to “safeguard the interests and sensibilities” of indigenous cultures.³⁷¹

Example 29 - Sloan Digital Sky Survey (SDSS)

4.256 SDSS is a project that aims to conduct wide-angle surveys of the sky in order to advance science in the area of extra-galactic astronomy.³⁷² When the survey is

³⁶⁷ <<http://www.paradisec.org.au/PDSCaccess.rtf>> at 28 May 2007

³⁶⁸ <<http://www.paradisec.org.au/PDSCaccess.rtf>> at 28 May 2007

³⁶⁹ See, for example, <<http://azoulay.arts.usyd.edu.au/paradisec/login.php>> at 28 May 2007

³⁷⁰ <<http://www.paradisec.org.au/database.html>> at 28 May 2007

³⁷¹ ³⁷¹ <<http://www.paradisec.org.au/PdsCdeposit.rtf>> at 28 May 2007

complete, it will provide detailed optional images of the sky and a 3-dimensional map of close to a million galaxies.³⁷³ The project website describes SDSS as “the most ambitious astronomical survey ever undertaken.”³⁷⁴

4.257 SDSS is managed by the Astrophysical Research Consortium (ASTRORC) for the Participating Institutions.³⁷⁵ There are currently 25 Participating Institutions, including the American Museum of Natural History, Cambridge University, the Chinese Academy of Sciences, Princeton University and the United States Naval Observatory.³⁷⁶

4.258 The main product of the SDSS is the Science Archive. This is a database of astrophysical data sets, data products and software interfaces.³⁷⁷ SDSS releases data to the scientific community and the general public in yearly increments.³⁷⁸

Ownership – SDSS

4.259 IP appears to be owned by the “SDSS Collaboration”, which comprises of all Participating Institutions and Participants as listed and defined in the SDSS Principles of Operation.³⁷⁹

4.260 The SDSS Principles of Operation provide:

Initiatives to use the intellectual property of the SDSS Collaboration will be covered by the Intellectual Property policy of ARC, Inc., a copy of which is available from the ARC Secretary.³⁸⁰

4.261 The ASTRORC Intellectual Property Policy generally applies to IP: “invented, discovered, created or authored by faculty members or other professional employees of ARC member institutions.”³⁸¹ It defines Intellectual Property as:

All technology such as patentable or unpatentable inventions, discoveries, computer software, trade secrets, copyright and any directly related technical or scientific know-how.³⁸²

4.262 The ASTRORC Intellectual Property Policy provides:

³⁷² See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, p3

³⁷³ <<http://www.sdss.org/>> at 29 May 2007

³⁷⁴ Ibid

³⁷⁵ <<http://www.sdss.org/>> at 29 May 2007

³⁷⁶ Ibid

³⁷⁷ See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, p3

³⁷⁸ <<http://www.sdss.org/>> at 29 May 2007

³⁷⁹ See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, see also <<http://www.sdss.org/collaboration/>> at 4 June 2007

³⁸⁰ See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, 13-14

³⁸¹ Astrophysical Research Consortium Intellectual Property Policy, adopted by ARC Board effective 6/10/94, available on request from the ARC Secretary, see the ARC website at <<http://www.astro.washington.edu/arc/>> at 5 June 2007

³⁸² Ibid.

In general, it is the policy of [the ASTRORC member institutions] that their faculty and other professional employees agree to assign title to Intellectual Property they develop to the institution...[ASTRORC] requires that Intellectual Property resulting from [ASTRORC] activities disclosed to [ASTRORC] member institutions be reported, within three (3) months of disclosure, to the [ASTRORC] Treasurer or other corporate officer responsible for administering [ASTRORC's] intellectual property policy.³⁸³

4.263 Where intellectual property is created by persons other than faculty members or professional employees of ASTRORC member institutions, intellectual property ownership and rights are negotiated with ASTRORC on a case-by-case basis.³⁸⁴

In each such case, [ASTRORC] will assert a position of ownership and control of the Intellectual Property unless persuaded by the pertinent circumstances that a lesser position for [ASTRORC] is appropriate and justified. At the least, [ASTRORC] will expect to receive an appropriate share of any licensing fees or royalties resulting from the Intellectual Property developed in [ASTRORC] activities.³⁸⁵

Control – SDSS

4.264 SDSS Participating Institutions aim to make data public within 18 months of data collection, with this reducing to 12 months as the survey progresses.³⁸⁶ However, prior to public release, data is tightly controlled by the SDSS Collaboration.

4.265 Under the SDSS Principles of Operation:

Participants have unlimited access to the Science Archive and may use this access to pursue any scientific project, subject to the policies and procedures specified [in the Principles of Operation]. Participants must read and abide by the [Principles of Operation], and are responsible for protecting the scientific integrity of the SDSS and the data rights of other Participants.³⁸⁷

4.266 “Participant” is defined as:

- at ASTRORC Participating Institutions, full-time PhD-level scientific staff that receive at least 50% of their annual salary from that institution and who have Principal Investigator status at that institution or who have performed the equivalent of two years service to SDSS; and
- at non-ASTRORC Participating Institutions, Participants as defined in the individual Memorandums of Understanding with those institutions.³⁸⁸

4.267 The Spokesperson for the SDSS is required to maintain a projects list, a publications list and a thesis list. Any Participant using or sponsoring the use of SDSS data must inform the Spokesperson of:

³⁸³ Ibid

³⁸⁴ Astrophysical Research Consortium Intellectual Property Policy, adopted by ARC Board effective 6/10/94, available on request from the ARC Secretary, see the ARC website at <<http://www.astro.washington.edu/arc/>> at 5 June 2007

³⁸⁵ Ibid

³⁸⁶ See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, 4

³⁸⁷ Ibid, 9

³⁸⁸ See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, 9

- the initiation of any scientific project using SDSS data that is expected to ultimately result in publication;
- any manuscript using SDSS data that is being published (and must submit to the Spokesperson a copy of the manuscript at least three weeks prior to submission for publication); and
- any information needed to create and maintain an accurate and up-to-date posting for use of the data by PhD students.³⁸⁹

4.268 Anyone who submits for publication results based on SDSS data before the data is publicly released, without notifying the SDSS Collaboration, may be barred from further access to SDSS data.³⁹⁰

4.269 Once data is publicly released, the conditions imposed upon use of the data are:

- the data may not be used for commercial publication or any other commercial purpose except with the explicit approval of the ASTRORC; and
- data used in non-commercial scientific and technical publications should be properly cited and the publication should contain a proper acknowledgement of SDSS funding sources and Participating Institutions.³⁹¹

Access and Use - SDSS

4.270 Publicly released data is accessible by anyone via the SDSS website.

4.271 Prior to public release, Participants are able to access SDSS data for use in scientific projects.

4.272 Non-Participants may only be granted limited access to SDSS data prior to public release if:

- they are conducting a specific research project in collaboration with Participants;
- they are a post-doctoral fellow at a Participating Institution where more than 50% of their salary is paid by that institution and a Participant agrees to sponsor their access;
- they are a full-time student at a Participating Institution and a Participant at that institution agrees to sponsor their access; or
- they are an external collaborator working on a specific project and a Participant will agree to sponsor their access.³⁹²

Example 30 – The University of Melbourne – Experimental Particle Physics (EPP) – Collaboration with the ATLAS Experiment

4.273 The Experimental Particle Physics (EPP) research group in the University Of Melbourne School Of Physics is a participating member in the ATLAS Collaboration. (ATLAS). ATLAS is an international group of research bodies collaborating in the

³⁸⁹ Ibid, 10-11

³⁹⁰ Ibid, 11

³⁹¹ See <<http://www.sdss.org/dr1/coverage/credits.html>> at 29 May 2007

³⁹² See Astrophysical Research Consortium Principles for Operation for the Sloan Digital Sky Survey <http://www.sdss.org/policies/sdss_poo.html> at 28 May 2007, 9-10

development of a large hadron collider accelerator machine at CERN, the European Laboratory for Particle Physics in Geneva. The collider accelerator is expected to generate discoveries in the field of physics from testing head-on collisions of protons of high energy. The resultant data from such testing will be distributed to and analysed by collaboration members.

Ownership – EPP

4.274 In the report *Sustainable paths for Data-intensive Research Communities at the University of Melbourne* (2006), APSR examined the EPP-ATLAS project and observed that:

[p]rime responsibility for the data remains with the global collaboration. Post processing, including analysis of the datawhere Melbourne physicists will be doing their data analysis as individuals, remains with the individual and the local team. This is essentially where researchers try and compete with the rest of the collaboration to be the first to make these discoveries that come about. Where individual researchers generate output from shared data it is considered individual IP, but publications coming from this data must include the collaboration on the author list. So the researcher can never claim this as solely his/her own work.³⁹³

4.275 The copyright notice on the EPP homepage states:

Copyright © 2000-2007 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.³⁹⁴

4.276 The ATLAS project itself has two websites: the ATLAS Public website containing general information for the public at large and the ATLAS Collaboration website containing research data submitted by collaboration members and which is only accessible by collaboration members. There is no direct statement as to ownership of rights in relation to data on the public website, although it does provide a fact sheet,³⁹⁵ which states that after a research group has produced a draft of its research findings:

[b]oth the analysis and its description in the draft will then be subject to comments by ATLAS collaborators, to careful review, and eventually to discussion within a plenary meeting...a consensus, and the agreed-upon paper can be submitted for publication.

Internal publications within the Collaboration, usually with one or a few authors, will document the individual contributions. These can be made known to the scientific community. Also, leading contributions are often recognized by asking the person in question to present results at the conferences. Often major results are obtained in a collective way, because people need to share the tasks.³⁹⁶

4.277 These statements suggest that individual EPP researchers retain rights in intellectual property or in relation to data which they develop and generate from their research for the ATLAS project. In practice, the process of sharing, commenting

³⁹³ Anna Shadbolt, Dirk van der Kniff, Eve Young and Lyle Winton, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories*, August 21 2006, <http://www.apsr.edu.au/aeres/sustainable_paths.pdf> at 18 June 2007, 9

³⁹⁴ See <<http://epp.ph.unimelb.edu.au/EPP>> at 18 June 2007

³⁹⁵ See <http://atlasexperiment.org/pdf/atlas_factsheet_8.pdf> at 19 June 2007

³⁹⁶ See <http://atlasexperiment.org/pdf/atlas_factsheet_8.pdf> at 18 June 2007

upon and publicizing this data may increase the numbers of contributing researchers who may also obtain rights in relation to such data.

Control – EPP

4.278 EPP-ATLAS data is subject to security restrictions (for example user passwords and authentication through a Grid computing network). Members are required to pay an annual contribution towards the maintenance and operation of equipment and infrastructure for the continued progress of the ATLAS project.

Access and Use – EPP

4.279 Membership of the ATLAS Collaboration entitles all members to access the research data submitted to ATLAS by other members.³⁹⁷ This research data is not accessible by non-members,

Example 31 – Household Income and Labour Dynamics in Australia Survey (HILDA)

4.280 HILDA is a database containing data on economic and subjective well-being, labour market dynamics and family dynamics³⁹⁸ in Australia collected through a nationwide survey of individuals and households. HILDA was instigated and is funded by the Australian Federal Government Department of Families, Community Services and Indigenous Affairs.

Ownership – HILDA

4.281 The HILDA “Deed of Licence for General Release Dataset R5 for Australian Academic and Government Researchers” for users of HILDA data states that the Department owns “all Intellectual Property rights in the Dataset”.³⁹⁹ “Dataset” is defined in the Deed as Releases 1 to 5 of HILDA’s general release household and person level unit record dataset from the HILDA survey, including any variations. Users own intellectual property in any material which they create from the Dataset, except for the Dataset itself. Users are required to acknowledge the use of the HILDA Dataset in any publication.⁴⁰⁰

4.282 APSR’s report, *Sustainable paths for Data-intensive Research Communities at the University of Melbourne* (2006), observed that:

³⁹⁷ Anna Shadbolt, Dirk van der Kniff, Eve Young and Lyle Winton, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories*, August 21 2006,

<http://www.apsr.edu.au/aeres/sustainable_paths.pdf> at 18 June 2007, 10

³⁹⁸ See <<http://www.melbourneinstitute.com/hilda/>> at 19 June 2007

³⁹⁹ Clause 12.1. See

<<http://www.melbourneinstitute.com/hilda/data/Deed%20of%20Licence%20for%20General%20Release%205%20Aust%20govtacademicNEW.pdf>> at 19 June 2007

⁴⁰⁰ Clause 12.2 of the

All IP and copyright of the data belongs to the Australian Government. Scholarly output from the data is owned by the researcher.⁴⁰¹

4.283 It is clear that the Department owns rights in relation to the HILDA survey data.

Control – HILDA

4.284 The HILDA data files are divided into two categories:

- “General Release”: files from which all information that could identify survey subjects has been removed; and
- “In-Confidence”: files from which subject names and addresses have been removed but which still contain information such as postcodes, income levels and geographic information.⁴⁰² Users of this data must have a site assessment carried out to meet the Department’s data security requirements.

4.285 HILDA has separate Deeds of Licence depending upon whether the user is:

- an academic, researcher, student or a Commonwealth or State government employee;
- a private research organization;
- an overseas academic, student or researcher; or
- a user of the In-Confidence data files.

4.286 The only difference between the Deeds of Licence for the various categories of user is that for private research organizations, overseas users and users of In-Confidence data files, the maximum term for which access to the Dataset is granted is three years. After that time, these users are obliged to delete and destroy all copies of the Dataset in their possession.⁴⁰³

4.287 The HILDA Deed of Licence requires users to comply with the following conditions:

- confidentiality and non-disclosure obligations;
- restrictions limiting use to research topics approved by the Department;
- notified security obligations; and
- privacy obligations under the *Privacy Act 1988* (Cth).

⁴⁰¹ Anna Shadbolt, Dirk van der Kniff, Eve Young and Lyle Winton, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories*, August 21 2006,

<http://www.apsr.edu.au/aeres/sustainable_paths.pdf> at 18 June 2007, 29

⁴⁰² See the HILDA User Manual – Release 5

<<http://www.melbourneinstitute.com/hilda/doc/HILDA%20User%20Manual%20Release%205.0.pdf>> at 19 June 2007

⁴⁰³ See clauses 2.1 and 8.3 of the Deed of Licence for General Release Dataset R5 for Australian (other than Academic and Government Researchers) and Overseas Researchers

<<http://www.melbourneinstitute.com/hilda/data/Deed%20of%20Licence%20for%20General%20Release%205%20Overseas%20and%20Aust%20non-govtacademicNEW.pdf>> at 19 June 2007

4.288 Users must notify HILDA of the reason for access and how user's research will be consistent with the Department's objectives. The Department then considers whether or not the proposed access and use is consistent with the purpose for which the data was collected by HILDA and whether the user is capable of meeting prescribed data security requirements. Applications for access by private research organizations are considered on a case-by-case basis.

4.289 Users are not permitted to share data with other parties unless the parties with whom the data is shared are also authorized users. Each user in a research group must apply to use the data.⁴⁰⁴

Access and Use – HILDA

4.290 HILDA data is only accessible by authorized users who are granted a non-exclusive, non-transferable licence under the Deed of Licence by the Department⁴⁰⁵ to use, copy, adapt and modify the Dataset. Users must also complete an order and payment form and pay the prescribed access fee.

⁴⁰⁴ See <<http://www.melbourneinstitute.com/hilda/data/datafaq.htm>> at 19 June 2007

⁴⁰⁵ See <<http://www.melbourneinstitute.com/hilda/data/datafaq.htm>> at 19 June 2007

3. Observations on Data Sharing Practices

4.291 The databases examined in this preliminary survey dealt with rights in relation to data and controls on access to and reuse of data. From the practices observed in these databases, it is possible to make some observations on current data management practices and data sharing policies.

Clarity and Understanding

4.292 For many of the databases examined, ascertaining the access and reuse rights and restrictions applying to the data was time-consuming and cumbersome. To understand the data deposit process and the conditions applying to data access and sharing, it was necessary to search through multiple web pages, policies, statements, copyright notices and links to other web pages.

4.293 Consequently, it is recommended that each database should have a single comprehensive web page which definitively:

- states whether rights exist in relation to the data contained in the database;
- identifies who owns or retains rights in relation to data, whether this be the database manager or otherwise;
- differentiates between data that is subject to copyright and/or other legal restrictions and data that is not subject to such restrictions;
- identifies the different levels of access and use rights and restrictions attaching to the data;
- addresses the rights and obligations of users in relation to sharing of the accessed data with third parties;
- highlights where copyright attaches to data, states whether there is a general copyright licence applying to the data and sets out the terms of such licence; and
- distinguishes data ownership from data control and explains the control exercised by the relevant parties.

4.294 This single web page could provide hyperlinks to the relevant policies, statements and notices and should be easily accessible from an immediately-identifiable link on the database's homepage.⁴⁰⁶ This web page could also provide

⁴⁰⁶ For example, see the Nature Precedings website at <<http://precedings.nature.com/>> at 26 July 2007. Nature Precedings is a database where researchers in biology, medicine (except clinical trials), chemistry and the earth sciences can share pre-publication research, unpublished manuscripts, presentations, posters, white papers, technical papers, supplementary findings, and other scientific documents. The home page provides a link to the "Terms and Conditions" page, which sets out Nature Precedings entire policy and conditions relating to contributions to the database, use of the database, intellectual property rights and data protection and privacy. See

links to examples of licences and/or agreements used by the database for the submission and use of data, which would provide immediate guidance on issues of data rights ownership, control, access and use.

4.295 Certain terms used to describe the database's access and sharing practices should be clearly and precisely defined. For example:

- terms such as “open”, “open access”, “free”, “public”, “research” and “commercial” - which do not have a single, commonly-understood meaning - should be defined so that the sense in which they are used in the database is clear. This would ensure that their intended meaning in that context will not be misunderstood by data submitters or database users;
- “data” should be defined to clarify what is included in this term and whether there are different kinds of data or data with different levels of access and use restrictions;
- “owner” and “ownership” should be defined to clarify exactly who the owner is and how the ownership rights arise.

4.296 Copyright notices could be improved by identifying the exact material to which the copyright notice applies. This would clarify whether the database manager claims copyright only in the database webpages, the data record, the text and pictures on the website and the underlying software, or the actual data that has been submitted to the database, or all or a combination of these.

4.297 If the matters described above were expressly addressed on database websites, users would be provided with information about their rights of access and use and any restrictions applying to the data as soon as they enter the database website.

Consistency of Language

4.298 Databases could ensure that language relating to issues of ownership, control, restrictions, access and use is used in a consistent manner, by cross-checking defined terms across the database web pages, policies and statements. Consistent usage of terms avoids any ambiguity, imprecision and contradiction which is otherwise likely to occur and avoids confusion about and difficulties in enforcing rights in relation to data. Greater consistency in data licensing, accessibility and use can also be achieved by adopting existing information exchange frameworks (such as the Creative Commons and Science Commons frameworks).

4.299 There is scope for these practices to be further examined and analysed, and for practical solutions to be developed. This is discussed further in Chapter 9 and Chapter 10.

4. Patents and Open Licensing

Inventing human genes is impossible. So every discovery relating to genes – their sequence, functions and everything else – should be placed in the pre-competitive arena. After all, one goal of the patent process is to stimulate competition. The most valuable gene-related applications are often far removed from the first easy steps. So this is a matter of science, not just a matter of principle.⁴⁰⁷

4.300 Data or information can be practically applied in such a way that it forms part of or gives rise to an invention capable of being patented. Patents protect products and processes that are novel, useful and involve an inventive or innovative step,⁴⁰⁸ by conferring on a patentee the exclusive right to exploit the patented product or process for a period of time (usually 20 years from the time the application is filed). During this time, only the patent owner or licensees may make, use or sell the patented invention. Unlike copyright, which arises automatically upon the creation of an original work, to obtain a patent it is necessary to file an application with the Patent Office and comply with formal requirements.

4.301 Discussion surrounding patents and data has most often centred on the patenting of genomic data. In the 1990s, several groups of scientists embarked on the Human Genome Project, with the objective of sequencing the human genome. In 1996, the various groups of scientists involved in the Human Genome Project came together with the intention of sharing their research data in order to accelerate their common goal.⁴⁰⁹ Yet despite this early spirit of sharing, some firms started looking at the possibility of commercialising the results through patenting and charging licensing fees for access to private databases.⁴¹⁰ Researchers and commentators became concerned that successful patentees would unreasonably restrict access to and use of the genome data, stifling research and innovation. The response of participants in the Human Genome Project, as well as later models that built upon the approach adopted in the Human Genome Project, are discussed below.

What Are Patents?

4.302 In Australia, patents are granted under the *Patents Act 1990*. Section 13 of the *Patents Act 1990* provides:

13 Exclusive rights given by patent

(1) Subject to this Act, a patent gives the patentee the exclusive rights, during the terms of the patent, to exploit the invention and to authorise another person to exploit the invention.

(2) The exclusive rights are personal property and are capable of devolution by law.

4.303 “Exploit” is defined in the Dictionary in Schedule 1 to the *Patents Act 1990* as follows:

⁴⁰⁷ John Sulston, ‘Heritage of Humanity’ (2002) *Le Monde diplomatique* <<http://modediplo.com/2002/12/15/genome>> at 21 May 2007.

⁴⁰⁸ See Section 18 *Patents Act 1990*

⁴⁰⁹ Andres Guadamuz, ‘Open Science: Open Source Licences in Scientific Research’ (2006) 7(2) *North Carolina Journal of Law & Technology*, 333.

⁴¹⁰ *Ibid*, 334.

(a) where the invention is a product--make, hire, sell or otherwise dispose of the product, offer to make, sell, hire or otherwise dispose of it, use or import it, or keep it for the purpose of doing any of those things; or

(b) where the invention is a method or process--use the method or process or do any act mentioned in paragraph (a) in respect of a product resulting from such use.

4.304 The term of a standard patent is 20 years, which typically runs from the date on which the complete patent application is filed: s 67 *Patents Act 1990*.

4.305 The grant of exclusive rights to exploit the patented invention does not mean that patent laws operate to restrict the communication of information or knowledge about the patent. In fact, they have exactly the opposite effect. The “price” a patentee pays for the patent monopoly is the disclosure into the public domain of all information necessary to understand and perform the patented invention.

What Are The Concerns Relating To Patents?

4.306 Concerns relating to patents depend upon whether a researcher intends to obtain a patent for an invention based on their research data. For researchers intending to seek patent protection for inventions derived from their research, a primary concern is whether they will be able to obtain a patent and whether disclosure of their data to other researchers could prevent them from obtaining a patent. For researchers who do not intend to patent, a concern is whether another person could secure a patent over an invention that encompasses the researcher’s data.

Researchers Intending To Patent

4.307 To obtain a patent in Australia, what is described in the patent must be novel and inventive and must not have been secretly used at the time the patent application is filed.⁴¹¹ Disclosing data (for example, to other researchers or to a funding body) could ruin a researcher’s chances of obtaining a patent, because if the information or invention is in the public domain, then it is not “novel” or “inventive” as required under patent law.

4.308 Section 9 of the *Patents Act 1990* sets down a number of acts that are not taken to be secret use of an invention by a potential patentee and that do not therefore prevent a researcher from obtaining a patent. These include use for the purpose of reasonable trial and experiment only and use occurring solely in the course of a confidential disclosure of the invention.

4.309 Section 24(1)(a) *Patents Act 1990* and Regulation 2.2 *Patents Regulations 1991* provide for circumstances that are to be disregarded in deciding whether an invention is novel or involves an inventive step. These circumstances include:

- publication or use of the invention within 12 months before the filing date of a complete patent application;
- the showing or use of the invention at a recognised exhibition;

⁴¹¹ See *Patents Act 1990* (Cth) s 18.

- publication of the invention in a paper written by the inventor and read before a learned society or published by or on behalf of a learned society; and
- the working in public of the invention within the period of 12 months before the priority date of a claim for the invention for the purposes of reasonable trial and because of the nature of the invention it is reasonably necessary for the working to be in public.

4.310 Despite these exceptions in the *Patents Act 1990*, a researcher who intends to patent their data should only disclose their data under confidentiality agreements. This would ensure that the data is kept secret and out of the public domain so that a patent can still be obtained.

4.311 *The Patents Act 1990* makes some provision for disclosures made in breach of a confidentiality agreement. Section 24(1)(b) provides that in deciding whether an invention is novel or involves an inventive step, any information made publicly available without the consent of the patentee through any publication or use of the invention by another person who derived the information from the patentee is to be disregarded.

Researchers Not Intending To Patent

4.312 Some researchers will be more interested in making their data openly available to advance research than in commercialising patented products or processes derived from their research. These researchers will not be concerned that public disclosure of their research data could constitute prior art which would prevent them obtaining a patent because the invention is no longer novel or is obvious.

4.313 A range of approaches have been developed by researchers and research institutions committed to ensuring that their data remains openly accessible. These approaches are described below.

The Bermuda Principles Approach – Releasing Data to Pre-empt Patents

4.314 The Bermuda Principles, developed in 1996 by a consortium of researchers involved in the Human Genome Project, require gene sequence assemblies larger than 1 kb to be released automatically, preferably within 24 hours. Through the rapid public disclosure of genomic data, the Bermuda Principles sought to create prior art which would defeat potential patents filed by high-profile private sector competitors.⁴¹²

4.315 However, disclosure of data, in itself, will not always be enough to prevent patenting, as was demonstrated in the race to sequence the human genome. In 1998, the private sector firm Celera Genomics began competing with the publicly funded

⁴¹² See G Parchomovsky, “Publish or Perish”, (2000) 98 Mich. L. Rev. 926; D Lichtman, K Kraus and S Baker, “Strategic Disclosure in the Patent System”, (2000) Vand. L. Rev. 2175; R Eisenberg, “The Promise and Perils of Strategic Publication to Create Prior Art: A Response to Professor Parchomovsky”, (2000) 98 Mich. L. Rev. 2358; and R Eisenberg and A Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 Berkeley L.J. 1187 at 1195 – 1996.

Human Genome Project to be the first to sequence the human genome.⁴¹³ Celera Genomics intended to commercially exploit its research findings, by selling licences to access their data using discriminate pricing – higher prices for pharmaceutical companies and lower prices for academic institutions and researchers and for the Howard Hughes Medical Institute.⁴¹⁴

Celera initially announced that it would seek patent protection on “only 200-300” genes, but ...eventually filed preliminary (“place-holder”) patent applications on 6,500 whole or partial genes. Celera also promised to publish their findings in accordance with the terms of the 1996 “Bermuda Statement,” by releasing new data quarterly (the [Human Genome Project] released its new data daily), although, unlike the publicly funded project, they would not permit free redistribution or commercial use of the data.⁴¹⁵

4.316 Ultimately, the sequenced human genome was made openly available for researchers to use. Nevertheless, the sequencing race demonstrated the risks associated with the approach taken in the Bermuda Principles. The problem arising from the public release of data is that it leaves the way open for another party to make improvements to the disclosed data and then make these improvements proprietary.⁴¹⁶ As Eisenberg and Rai explain, in the context of the Human Genome Project:

Some public sector scientists ... did not view creating prior art as the best weapon for defeating proprietary claims. Because the data were freely available, those who accessed the data could blend it with their own privately-held information and make the combination proprietary. These scientists suspected that Craig Venter, the major private sector challenger to the [Human Genome Project], had adopted this approach.⁴¹⁷

4.317 Claire Driscoll of the NIH describes the dilemma as follows:

It would be theoretically possible for an unscrupulous company or entity to add on a trivial amount of information to the published...data and then attempt to secure ‘parasitic’ patent claims such that all others would be prohibited from using the original public data.⁴¹⁸

4.318 Where information or data is used to develop a patentable invention, the subsequent patent rights may be broad enough to cover use of the actual data forming part of the invention. As Eisenberg and Rai explain:

Although raw genomic data would not undermine claims to specific genes of identified function, annotated data might do so. A major goal of annotation is to identify coding regions

⁴¹³ See <http://en.wikipedia.org/wiki/Human_Genome_Project> at 22 May 2007; Robert Cook-Deegan, ‘The Urge to Commercialize: Interactions Between Public and Private Research Development’, Steering Committee on the Role of Scientific and Technical Data and Information in the Public Domain, Office of International Scientific and Technical Information Programs, National Research Council, *The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium* (2003) The National Academies Press, Chapter 11, available at <http://books.nap.edu/openbook.php?record_id=10785&page=87> at 22 May 2007

⁴¹⁴ Ibid, see especially Robert Cook-Deegan at p90

⁴¹⁵ <http://en.wikipedia.org/wiki/Human_Genome_Project> at 22 May 2007

⁴¹⁶ See, for example, Donna M Gitter (2007) ‘Resolving the Open Source Paradox in Biotechnology: A Proposal for a Revised Open Source Policy for Publicly Funded Genomic Databases’ 43(4) *Houston Law Review* available at <<http://ssrn.com/abstract=901994>> at 11 May 2007, 4

⁴¹⁷ R Eisenberg and A Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 *Berkley L.J.* 1187 at 1207.

⁴¹⁸ Claire T Driscoll, ‘NIH data and resource sharing, data release and intellectual property policies for genomics community resource projects’ *Expert Opin. Ther. Patents* (2005) 15(1), 4

in the genome and add information about the function of the protein for which the region codes.⁴¹⁹

4.319 Consequently, subsequent research projects relied on licensing methods in an attempt to restrict the development of intellectual property in downstream discoveries based on the disclosed data, rather than simply releasing the data into the public domain.

The HapMap Approach – Conditional Access Licensing

4.320 The International Haplotype Project (HapMap) was a three-year project which ran from 2002 to 2005. It involved international collaboration among a consortium of research organisations from Japan, the UK, Canada, China, Nigeria, and the United States. The aim of the project was to develop a haplotype map of the human genome, mapping common patterns of DNA sequence variation by determining the genotypes of one million or more sequence variants, their frequencies and the degree of association between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The HapMap will allow the discovery of sequence variants that affect common disease, will facilitate development of diagnostic tools and will enhance our ability to choose targets for therapeutic intervention.

4.321 The HapMap project was established on the basis that all the data should be freely accessible to the public and that restrictive patents should not be filed on inventions based on submitted data. The project website states that:

The Project will release all the data it produces into the public domain, so that any researcher can use the information.⁴²⁰

4.322 The raw data on single base DNA variations – also known as single nucleotide polymorphisms (SNPs) – and allele frequencies were made available at dbSNP, a publicly accessible database. However, in order to guard against the possibility of applications being filed for parasitic patents, the HapMap Steering Committee developed an innovative defensive strategy. According to Eisenberg and Rai, the approach toward data sharing in the HapMap project appears to have been influenced by the frustrations of public sector scientists with the public domain approach that had been adopted in the Human Genome Project.⁴²¹

4.323 Individual genotypes were made available on the HapMap website, but anyone seeking to use the research data was first required to register via the website and enter into a click-wrap licence for the use of the data. The licence entered into, the International HapMap Project Public Access Licence, was explicitly modeled on the General Public Licence (GPL) used by open source software developers.⁴²² A

⁴¹⁹ R Eisenberg and A Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 Berkley L.J. 1187, footnote 55 at 1202.

⁴²⁰ <<http://www.hapmap.org/abouthapmap.html>> at 22 May 2007.

⁴²¹ R Eisenberg and A Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 Berkley L.J. 1187 at 1207.

⁴²² The HapMap Licence included an acknowledgement of the GNU General Public Licence of the Free Software Foundation.

central term of the licence related to patents. It allowed users of the HapMap data to file patent applications on associations they uncovered between particular SNP data and disease or disease susceptibility, but the patent had to allow further use of the HapMap data. The licence specifically prohibited licensees from combining the HapMap data with their own in order to seek product patents on combinations of SNPs known as haplotypes (although it may have allowed claims for certain uses of SNPs and haplotypes).⁴²³ Paragraph 2(b) of the licence stated:

You shall not file any patent applications that contain claims to particular uses of any SNP, genotype or haplotype data obtained from the Genotype Database or any SNP, haplotype or haplotype block based on data obtained from the Genotype Database, unless such claims do not restrict, or are licensed on such terms that they do not restrict, the ability of others to use at no cost the Genotype Database or the data that it contains for other purpose.⁴²⁴

4.324 The requirement to enter into an agreement before being given access to the data was discontinued in December 2004 in order to allow the HapMap data to be integrated into other open access databases.

4.325 More recent community resource projects have adopted a less aggressive approach towards the use of data in future patents than that adopted in the HapMap project.⁴²⁵ Eisenberg and Rai point to the approach taken in the Genetic Association Information Network (GAIN) project:

Like the HapMap licence, the GAIN Data Use Certification requires those who access the data to refrain from disclosing the data to anyone who is not bound by the same agreement. It also urges registrants not to rely on GAIN-supported data to seek patents on markers that might be useful in diagnosis or identification of drug targets. However, the language is entirely hortatory, calling upon approved users to “acknowledge the intent” of the GAIN IP policy, reminding them that “in this spirit, it is expected” that data and conclusions will remain freely available, and stating that GAIN “encourages” compliance with various NIH policies that favour sharing. Further, the document explicitly recognizes the importance of the later development of IP on downstream discoveries, especially in therapeutics.⁴²⁶

The CAMBIA Approach – Open Licensing

4.326 CAMBIA is an international, independent non-profit research institute led by world-renowned scientist, Richard Jefferson. CAMBIA was designed to “foster innovation and a spirit of collaboration in the life sciences.”⁴²⁷ This goal is achieved through four interconnected work products:

⁴²³ R Eisenberg and A Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 Berkley L.J. 1187 at 1208 – 1209.

⁴²⁴ HapMap, *Project Public Access Licence*, previously at: <<http://www.hapmap.org/cgi-perl/registration>>; See also Andres Guadamuz, ‘Open Science: Open Source Licences in Scientific Research’ (2006) 7(2) *North Carolina Journal of Law & Technology*, 349-50.

⁴²⁵ For criticisms of the HapMap Project Model see Donna M Gitter (2007) ‘Resolving the Open Source Paradox in Biotechnology: A Proposal for a Revised Open Source Policy for Publicly Funded Genomic Databases’ 43(4) *Houston Law Review* available at <<http://ssrn.com/abstract=901994>> at 11 May 2007.

⁴²⁶ R Eisenberg and A Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 Berkley L.J. 1187 at 1209 – 1210.

⁴²⁷ <<http://www.cambia.org/daisy/cambia/home.html>> at 27/02/2007.

- *Patent Lens*, which provides tools to make patents and patent landscapes more transparent;
- *Biological Open Source Initiative* (BiOS), which advocates for the sharing of life sciences technology and data through a series of licences;
- *BioForge*, a research portal (or repository) that makes data and technologies openly available for others to use in new innovations, whether for research, commercial use, or humanitarian use; and
- *CAMBIA's Materials*, new technologies developed by CAMBIA, particularly in the field of genetics, which CAMBIA makes openly available under a BiOS licence.⁴²⁸

4.327 CAMBIA has also applied for and been granted twelve patents for biological inventions in different patent offices around the world.⁴²⁹ CAMBIA's approach involves obtaining patents over products or processes based on or encompassing data, but then licensing the use of the data under open terms. A primary object of this is to ensure that the biological material is not patented by others under restrictive terms which do not allow for open access and use by others. Another object is to encourage innovation. CAMBIA:

Strives to create new norms and practices for dynamically designing and creating the tools of biological innovation, with binding covenants to protect and preserve their usefulness, while allowing diverse business models for wealth creation, using these tools.⁴³⁰

4.328 CAMBIA has developed two open source licences relevant to data – the BiOS Plant Enabling Technology Licence and the BiOS Genetic Resource Technology Licence. Paragraph 2.1 of each licence grants to licensees:

[a] worldwide, non-exclusive, royalty-free right and licence to make and use the IP & Technology for the purpose of developing, making, using, and commercializing BiOS Licensed Products without obligation to CAMBIA, including a sub-licence...⁴³¹

4.329 Licensees are given the right to sub-licence the material, as long as it is sub-licensed under the same terms as contained in the original licence agreement. In this way, the BiOS licences are similar to Creative Commons Share-Alike licences.

4.330 CAMBIA's model allows and even encourages researchers to obtain patents over inventions that build upon CAMBIA's research data. However, instead of using patent licences to "extract a financial return from a user of a technology," CAMBIA advocates using a patent licence to "impose a covenant of behaviour."⁴³²

⁴²⁸ Ibid.

⁴²⁹ CAMBIA, *IP* <http://www.cambia.org/daisy/cambia/intellectual_property.html>

⁴³⁰ R Jefferson, 'Science as Social Enterprise: The CAMBIA BiOS Initiative' (2006) *Innovations* 13 at 22 available at

<http://www.bios.net/daisy/bios/3067/version/default/part/AttachmentData/data/INNOV0104_pp13-44_innovations-in-practice_jefferson.pdf> at 9 May 2007, 16

⁴³¹ CAMBIA, *Biological Innovation for Open Science, About BiOS (Biological Open Source) Licences* <<http://www.bios.net/daisy/bios/398>> at 28 February 2007.

⁴³² R Jefferson, 'Science as Social Enterprise: The CAMBIA BiOS Initiative' (2006) *Innovations* 13, 27

4.331 According to CAMBIA, the purpose of the BIOS licences is that:

[i]nstead of royalties, BIOS licensees must agree to legally binding conditions in order to obtain a licence and access to the protected commons. These conditions are that improvements are shared, and that licensees cannot appropriate the fundamental “kernel” of the technology and improvements exclusively for themselves. Licensees obtain access to improvements and other information, such as regulatory and biosafety data, shared by other licensees. To maintain legal access to the technology, licensees must agree not to prevent other licensees from using the technology in the development of different products.⁴³³

4.332 By making the licence cost-free, CAMBIA hopes to encourage what founder Richard Jefferson terms:

[t]he most valuable contribution to the license community: “freedom to innovate”.⁴³⁴

4.333 CAMBIA is working on developing a range of open source licences that can be applied in different fields (beyond biosciences) and that “allow institutions to preserve their own cultures and priorities.”⁴³⁵

<http://www.bios.net/daisy/bios/3067/version/default/part/AttachmentData/data/INNOV0104_pp13-44_innovations-in-practice_jefferson.pdf> at 9 May 2007.

⁴³³ Ibid.

⁴³⁴ Ibid 29

⁴³⁵ Ibid 30.

5. Researchers' Data Sharing Attitudes and Practices

4.334 While the benefits of data sharing have been widely acknowledged by scientific organisations, there has been some reluctance among researchers to embrace data sharing practices. Further, there appears to be a low level of awareness among researchers of the opportunities for employing an open access approach to enable licensing of data to permit re-use.⁴³⁶

4.335 Surveys of researchers' attitudes to data sharing indicate that their reluctance to share their research data is based on a range of concerns including:

- lack of acknowledgement;
- use of data for unintended purposes;
- misinterpretation of data;
- publication and use of the data without the knowledge of owners of rights in relation to data; and
- publication of the data in such a way as to prevent subsequent publication by the person who own rights in relation to the data.

4.336 Reichman and Uhler suggest that the pressure on universities to commercialise publicly funded research causes researchers to refrain from fully divesting their data and to treat data collections as private goods under restrictive conditions.⁴³⁷ Researchers in "small science" environments retain considerable discretion in determining the amount and conditions of disclosure, subject to issues of peer pressure, self interest and informal data sharing networks:

By withholding data, [researchers] temporarily promote their individual competitive advantages. At the same time, they may lose opportunities to improve their positions by gaining mediated access to other's data.⁴³⁸

4.337 Brandt-Rauf states that access to data amongst researchers has been restricted because of a desire for ownership of the data and for credit in relation to its development.⁴³⁹ Resistance to data sharing can:

⁴³⁶ Alan Tonge and Peter Morgan, *Project SPECTRa: Submission, preservation and exposure of Chemistry Teaching and Research Data*, JISC Final Report (2007) <http://www.lib.cam.ac.uk/spectra/documents/SPECTRa_Final_Report_v10.doc>.

⁴³⁷ J Reichmann and P Uhler, 'A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment' (2003) 66 *Law and Contemporary Problems* 315, 342 <<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

⁴³⁸ J Reichmann and P Uhler, 'A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment' (2003) 66 *Law and Contemporary Problems* 315, 346 <<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

⁴³⁹ Sherry Brandt -Rauf, 'The Role, Value, and Limits of S&T Data and Information in the Public Domain for Biomedical Research', in National Research Council of the National Academies (2003)

- waste resources by leading to duplication of efforts;
- slow the progress of science because scientists cannot weasily build on the efforts of others or discover errors in completed work, and
- lead to a generalized level of mistrust and hostility among scientists in place of what should be a community of scientists.⁴⁴⁰

4.338 Eisenberg and Rai observe that there is emerging evidence to suggest that some research communities in the life sciences are reluctant to share data even after publication. In support of this observation they refer to a survey conducted by Eric Campbell and his colleagues which found that 47% of academic geneticists who had made a request to another academic had been denied access to data or materials associated with a published article at least once in the preceding three years.⁴⁴¹ Eisenberg and Rai considered various reasons why scientists may be reluctant to share data:⁴⁴²

- (i) the perception that sharing data, even after an initial publication, is providing advantages to competitors in the race to generate further publications;
- (ii) involvement in commercial activities; and
- (iii) sharing may imperil patent applications or destroy confidentiality.

4.339 Scientific competition and commercial involvement were both found to be important reasons for refusal to share data. Concerns over sharing of data have been discussed in the literature in the context of specific disciplines, for example neuroscience.⁴⁴³

4.340 The need to address researchers' concerns about whether their contributions in creating data will be recognised if they participate in data sharing arrangements has been acknowledged by the ICSU. It has recommended that scientists should be recognised and given credit for the scientific contribution of the datasets they produce as well as for the analysis of the data.⁴⁴⁴

4.341 Recent surveys of the Australian research community provide indications about current attitudes and practices in relation to data ownership and sharing. Consultations conducted by NCRIS on Platforms for Collaboration indicate that while data volumes are growing rapidly, much research data is held in isolated forms which are not readily accessible and there are different "missions" with respect to data. A more extensive survey is currently being undertaken by the School of Law at QUT, to

Julie M. Esnau and Paul F. Uhler (eds), *The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium*, National Academies Press, Washington USA, p.66.

⁴⁴⁰ Ibid.

⁴⁴¹ Eric Campbell et al, 'Data Withholding in Academic Genetics: Evidence from a National Survey' (2002), 287 *JAMA* 473, 477.

⁴⁴² Rebecca Eisenberg and Arti Rai, 'Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California's Stem Cell Initiative' (2006) 21 *Berkley Law Journal* 1187.

⁴⁴³ S H Koslow, 'Sharing primary data: a threat or asset to discovery?' (2002) 3 *Nature Reviews Neuroscience* 311.

⁴⁴⁴ ICSU, *Scientific Data and Information: A report of the CSPR Assessment Panel* (2004) 21.

gather more detailed and specific information on these issues and related topics regarding collaboration and e-Research.

Australian eResearch Sustainability Survey

4.342 This APSR project (funded by DEST) brings together a number of Australian universities to focus on “the management of scholarly assets in digital form”.⁴⁴⁵ In October 2006 APSR published the results of its Australian eResearch Sustainability Survey (“AERES”).

This national project surveyed data-intensive communities with a view to establishing current capabilities for the storage, access, and long term management of research data. Awareness of sustainability issues and how these are being addressed were also canvassed. The project examined how available resources may be applied to sustainability issues for these communities.⁴⁴⁶

4.343 The report, *Sustainability Issues for Australian Research Data: the report of the Australian e-Research Sustainability Survey Project*,⁴⁴⁷ is based on feedback from 80 research or support staff, across several research organisations and research service providers throughout Australia. The study found that:

The surveyed researchers had a strong awareness of the e-Research agenda and issues. This awareness extends to the specific issue of the enduring value of research data. All groups were tracking NCRIS activities and, where applicable, NCRIS timelines in their respective disciplines.

There is a general belief that publicly funded data should be made available to all, within appropriate access and security guidelines. This belief is more strongly held by researchers with management responsibilities than at the practitioner level.⁴⁴⁸

4.344 The AERES study found a distinct lack of formal policies for management of data. Rather, it found that there were informal practices which were adhered to – to a greater or lesser degree – depending on the perceived value of the data to the original researchers.⁴⁴⁹ Data management practices were not usually considered as part of the project plan for research projects, as the research data was viewed as an input to the research process rather than an outcome of intrinsic value separate from the report or article generated from it.

4.345 The report concluded that current data practices generally see data:

- managed sufficiently for research needs, but not professionally;
- discoverable through scholarly publication, but not otherwise;
- having a value placed on it for present needs, but not for the future;
- lost through commission and omission; and

⁴⁴⁵ See <<http://www.apsr.edu.au/aeres/>> page 2.

⁴⁴⁶ <<http://www.apsr.edu.au/currentprojects/currentprojects.htm>>.

⁴⁴⁷ M Buchhorn and P McNamara, *Sustainability Issues for Australian Research Data: The Report of the Australian e-Research Sustainability Survey Project*, Australian Partnership for Sustainable Repositories (2006) <<http://www.apsr.edu.au/aeres/>>.

⁴⁴⁸ Page 4 <<http://www.apsr.edu.au/aeres/>>.

⁴⁴⁹ Page 5 <<http://www.apsr.edu.au/aeres/>>.

- accessible only through approach to the author of the related publication.⁴⁵⁰

4.346 The AERES study found that researchers believed that data belonged to them specifically. The report surmised that this belief could be explained by the lack of formal examination of the value of data by researchers until now. Researchers have traditionally maintained their own data for their entire career.

They ask the question, ‘Who else could this data belong to?’

4.347 Consequently, researchers were reluctant to share their data, and believed that they had an exclusive right to exploit the data before it became available to others. Some of the surveyed domains have established policies which permit time limited data exploitation by the original researchers.⁴⁵¹

4.348 When researchers were made aware of the issues around data custodianship and questioned as to when data should be transferred to a repository, the study found that:

[t]he majority of groups consulted saw the data as belonging with the group for the life of the project. A minority thought that data lodgement for management and preservation needed to take place during the life of a project not at the end of the project. This view was more prevalent for groups in organisations with data centres. The reasoning was that lodging data at the end of the project reduced the chances of the data being lodged at all because the researchers would be focussed on the next project.⁴⁵²

4.349 The AERES report also highlights the need for a repository to develop trust relationships with the serviced academics, as:

[t]he level of knowledge about central services is directly related to whether researchers trusted them. If researchers did not know of any services that would meet their data management needs they would then be reluctant to use them until a level of knowledge and trust was established. Where services did exist, the fact that researchers did not know about them was considered to be an indication that the services were unreliable.⁴⁵³

NCRIS Platforms for Collaboration Data Management Survey

4.350 Another survey undertaken by Information Services at the University of Melbourne, *NCRIS Platforms for Collaboration Data Management Survey* (NCRIS survey), was directed towards questioning the key stakeholders in the management of research data throughout Australia.

4.351 Some of the views on data ownership reported by the NCRIS survey included:

- “the owners of the data remain the local researchers who collected and generated the data.... data ownership and access does not change over the data lifecycle”;
- “it seems reasonable to assume that ultimately data is owned by the funding body, government or otherwise”;

⁴⁵⁰ Page 44 <<http://www.apsr.edu.au/aeres/>>.

⁴⁵¹ Page 7 <<http://www.apsr.edu.au/aeres/>>.

⁴⁵² Page 7 <<http://www.apsr.edu.au/aeres/>>.

⁴⁵³ Page 9 <<http://www.apsr.edu.au/aeres/>>.

- “the owner is the researcher, the research group, or the principal investigator of the research group”;
- “the owners of the data are the original data collectors”;
- “this is a complex question as there are a variety of rights issues associated with data records as they move as they move from the environment of creation and initial use into a longer term historic framework”; and
- “one of the key issues for resolution around data ownership is that many researchers will not know who owns the data, and who has rights, but does not own it”.

4.352 From these comments, it is apparent that while some researchers are aware of the complexity of the issues involved in data ownership, most have only a rudimentary understanding. It is clear that there is a need for researchers to be provided with guidelines and a data management toolkit to assist in developing a better understanding of data ownership and management.

4.353 The NCRIS survey included the following comments on data sharing:

- “data is shared where there is a need to do so. Large and commonly used data sets such as astronomical, ocean and climate data, or genetic information are shared as a matter of community interest”;
- “most researchers reported that they did not use data from other researchers and that others did not use theirs”;
- “some researchers reported that carrying out carrying out quality control on their own data had made them aware of the possibilities for error in data collections”;
- “all researchers who worked with data collected from survey communities expressed strong concern about the importance of maintaining a trust relationship with that community”;
- “researchers are often reluctant to share their data. Data is viewed as a personal good, created by researchers and to be exploited by them”;
- “while many researchers feel data collections should be available to the research community, there is a very strong and unanimous view that researchers should be able to exploit “their” data for a period of time before it becomes available to others”;
- “there was concern that shared data may allow others to gain from the work of a group without recognition of that group”; and
- “there was a concern that analysis of data by others may diminish the impact of the research already reported”.

4.354 These opinions suggest that while there is an awareness of the potential benefits of data sharing within the Australian research community, there are also concerns about the exploitation of the data by others, particularly if this diminishes the credit attributed to the researchers who generated the data in the first place. The survey indicates that understanding and acceptance of the benefits derived from data sharing is not yet well established in Australia.

4.355 In its March 2007 report, *Public Support for Science and Innovation*, the Australian Government Productivity Commission noted that DEST had made it clear that the *NCRIS Platforms for Collaboration* was intended to play:

[a] central role in promoting access to and utilisation of research data, facilitating information flows throughout Australia's science and innovation system.⁴⁵⁴

QUT e-Research survey

4.356 In 2007 QUT's School of Law conducted a survey of researchers, focusing on legal issues arising in collaborative research projects and e-Research. The objective of the QUT survey was to obtain views from the research community across the academic, government and private sectors, including views on data and databases. The results are still being processed, but it is anticipated that a sufficiently large and statistically significant sample of responses will be available for analysis.

4.357 Researchers were asked to answer questions relating to data including:

- Do you presently take steps to ensure research data is available in a form which can be readily stored and accessed?;
- Does your organisation currently have defined mechanisms to assist you in storing and accessing your data in the long term?;
- How important are online tools, databases and electronic resources to your collaborative products?;
- Do you see exchange of data or information as informal or formal collaborations or agreements?;
- Is data ownership and access an issue that causes problems in negotiating formal agreements?;
- How important do you feel that the issue of data ownership or access will become with the increase of e-Research?;
- What types of data does your research routinely use/generate?;
- Do you have a clear understanding of who owns the data you use in your research?;
- Do you have a clear understanding of who owns the data you generate in your research?;
- Do you have any reservations about people outside your project or organisation having access to data created as a result of your research project and why?; and
- If you have objections to people outside your project or organisation having access to data created as a result of your research project, would your concerns be reduced by having a legally binding agreement that clearly defined ownership and limited your liability for the recipient's use of the data?

4.358 In relation to databases, researchers were asked a series of questions including:

- How often do you access external databases in conjunction with your research activities?;

⁴⁵⁴ Productivity Commission, *Public Support for Science and Innovation*, Research Report (2007) 212 <<http://www.pc.gov.au/study/science/finalreport/index.html>> at 3 April 2007.

- Are you required to register in order to access the databases?;
- In which geographical areas are the databases located?;
- Are you aware of legal restrictions associated with copying, extraction or reuse of information from the databases you access?;
- Do you always comply with restrictive notices presented on databases (particularly copyright restrictions)?;
- Would it facilitate your research to have clearer definitions on what can be legally copied, extracted or reused from particular databases?;
- Are you actively involved in creating a database as part of your research?;
- Does your research usually result in the production of data or datasets that are dumped into a database?;
- If your data or datasets are deposited into a database, was this database created by you or your research organisation, or some other body or institution?;
- Is this database located outside Australia?;
- Is your data available for access and use by other researchers?;
- If your data or datasets are deposited into a database available for access and use by other researchers, on what basis is it made available to them? (Open access or restricted access?);
- Does your organisation have a policy setting out the basis on which research data should be deposited into databases for access by other researchers?;
- If your organisation has a policy setting out the basis on which data should be deposited into databases for access by other researchers, can you provide any details of the policy?;
- If your organisation has a policy setting out the basis on which data should be deposited into databases for access by others, are researchers provided with guidelines on how the policy is to be put into practice?;
- Do you (or your organisation) prepare plans for the management and/or sharing of your research data?;
- At what stage of a research project are your plans for the management and/or sharing of your research data provided?; and
- Would it assist you to have access to a plain English “how to guide” explaining the legal restrictions associated with databases?

4.359 From a preliminary analysis of the survey results (for those respondents who answered the questions on data and databases) the following observations can be made:

- almost 90% of the respondents to the survey were in favour of databases having a plain English “how to” guide that explains legal requirements and restrictions relating to the database;
- almost three-quarters of respondents expressed the view that clearer explanations of what can be legally copied, extracted or reused from a database would help facilitate research;

- overwhelmingly, respondents supported the inclusion on databases of clearer definitions of legal ownership, liability, control, access and use of research data; and
- more than half of the respondents identified data ownership or access as very important issues in future contracting arrangements with other organisations or researchers.⁴⁵⁵

4.360 The observations of researchers surveyed are consistent with the observations made in relation to the analysis of existing database practices set out above. The researchers also identify current and future issues relating to the access and use of databases. Potential management solutions to these issues are suggested in Chapter 9 and Chapter 10 of this Report.

⁴⁵⁵ The survey results are forthcoming. See the Legal Framework for e-Research website at <<http://www.e-research.law.qut.edu.au/>>

KEY POINTS

In developing policies and systems for access to and sharing of research data, it is important to understand the legal frameworks and infrastructure for research data that already exist in the Australian and international research environment.

Many databases have developed protocols addressing the ownership of rights in relation to data, control of data, and access to and use of data. Some funding bodies and other organisations, such as the National Institutes of Health (NIH) in the United States, have established research databases across a wide range of research disciplines. Some of these databases give effect to data access policies favouring broad availability of research data for access and reuse, whereas others provide access to data on a more restricted basis. Where access and use are restricted, contractual licences may operate to regulate access to the database and the use of the data deposited in it. Comprehensive data access policies and web pages that define and clarify legal rights and obligations relevant to ownership, control, access and use of data will assist data submitters and users in understanding and following the database's practices and procedures for dealing with data.

It is also important to be aware of data sharing practices between researchers/academics and third parties. These practices will be influenced by the researchers' attitudes to data sharing practices. Surveys indicate that researchers are currently reluctant to share data, for a variety of reasons including concerns that data will be misinterpreted or used for unintended purposes. Increased advocacy and well-drafted data access policies will help to overcome this reluctance.

CHAPTER 5 - COPYRIGHT

“Legal ownership and control over the products of innovative and creative effort has typically been recognised as one or more forms of IP rights. Several of these different forms of IP have become directly relevant to the practice of e-science, which may generate valuable ideas, methodologies, data, databases, texts, software, and even tangible research tools.”⁴⁵⁶

⁴⁵⁶ Dan Burk, *Intellectual Property in the Context of E-Science*, Minnesota Legal Studies Research Paper No 06-47 (2006) <<http://ssrn.com/abstract=929479>> at 18 August 2006.

COPYRIGHT

Aims:

1. Provide an overview of the principles of copyright law relating to data, information, datasets and databases;
2. Explain how copyright can be protected by the law and through the use of technological mechanisms such as digital rights management (DRM) and electronic rights management information (ERMI);
3. Consider how permission can be given for the use of copyright material, including bare permissions, contractual licences, statutory licences and open content licences such as Creative Commons licences; and
4. Consider the use of copyright and other laws to protect databases, comparing the legal position in Australia, the United States, the United Kingdom and Europe.

1. Introduction

5.01 Among the various forms of intellectual property relevant to e-Research, copyright is of primary importance. Copyright is a proprietary right, meaning that like conventional forms of property it can be owned, sold (assigned) and licensed.

5.02 This chapter examines copyright as it relates to data, datasets and databases. It looks at the different types of material that can be deposited in a database, including raw data, written documents such as tables, graphs, articles and reports, and visual material such as diagrams and charts. Additionally, this chapter considers how copyright applies to subsidiary material such as metadata and the software underlying a database. As the majority of these materials will fall within the “literary works” category, and most of the rest will fall into the “artistic works” category, the focus of this chapter will be the law under Part III of the *Copyright Act 1968*.

2. Materials Protected by Copyright

5.03 Copyright applies to a range of materials, including written text, visual images, music and sound recordings, performances and computer software. Under the *Copyright Act 1968* these materials are divided into two categories – “works” under Part III and “subject matter other than works” under Part IV. “Works” include literary, dramatic, musical and artistic works. Computer programs are protected under the *Copyright Act 1968* as literary works. “Subject matter other than works” includes cinematograph films (which encompasses feature films, videotapes, commercials, television programs and video clips), sound recordings and sound and television broadcasts.

3. Requirement of Material Form

5.04 Copyright comes into existence automatically, without the need for registration, as soon as a work which meets certain qualifying criteria is expressed in material form. It is a fundamental principle of copyright law that copyright does not protect ideas, information or facts but instead protects the form in which those ideas, information or facts are expressed in a material form. “Material form” is met when the work is written down, or entered into a computer or stored in some other machine-readable form.

4. Copyright in Data and Databases

5.05 The requirement of material form means that raw data, basic facts or items of information will not, in themselves, attract copyright protection. However, where data, information and facts have been compiled to create a new work, it may be protected by copyright as a compilation if it meets the originality threshold. Compilations are protected by copyright in the literary works category, where the *Copyright Act* defines “literary work” as including:

a table, or a compilation, expressed in words, figures or symbols.⁴⁵⁷

Data may be compiled into datasets, which may then be compiled into a larger database. Both the datasets and the database may be protected by copyright.⁴⁵⁸

5.06 It is not only the database itself that may be protected by copyright. The underlying software will also attract copyright protection. Software is protected by copyright in the literary works category. The definition of “literary work” has been extended to include:

a computer program or a compilation of computer programs.⁴⁵⁹

“Computer program” is defined as:

a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.⁴⁶⁰

5.07 Again, it is not the information or ideas behind the software that is protected, but the material form of the software itself.

With respect to computer programs, copyright protects the expression of the methods or systems used in the program in the form of object code or source code but does not protect the functionality of the program or the ideas and concepts underlying it.⁴⁶¹

5.08 Sometimes, more than simply raw data will be deposited in a database. The database may also contain textual material (such as journal articles, research reports

⁴⁵⁷ *Copyright Act 1968* (Cth) s 10(1).

⁴⁵⁸ For an overview of the application of copyright law in relation to databases of gene sequences and related information, see Australian Law Reform Commission, *Genes and Ingenuity: Gene Patenting and Human Health*, Report No 99 (2004) Ch 28 Copyright and Databases <<http://www.austlii.edu.au/au/other/alrc/publications/reports/99/>>.

⁴⁵⁹ *Copyright Act 1968* (Cth) s 10(1).

⁴⁶⁰ *Copyright Act 1968* (Cth) s 10(1).

⁴⁶¹ *Autodesk v Dyason* (No 2) (1993) 176 CLR 300.

or manuscripts) and visual material (such as diagrams, maps, charts or plans). These materials will be copyright protected in their own right as literary works and artistic works respectively.

5.09 Since the 1950s, questions have arisen as to whether copyright could subsist in biological or gene sequences consisting of alphanumeric characters.⁴⁶² The issue has never been definitively resolved. Academic commentary and official statements by the US Copyright Office favour the view that biological sequences are not suitable for copyright protection.⁴⁶³ In contrast, the English High Court judge Sir Hugh Laddie has expressed support for the idea of allowing copyright to protect sequences,⁴⁶⁴ although his view is not universally accepted.⁴⁶⁵ In Australia, it has been suggested that copyright may subsist in the written representation of biological or gene sequences as a literary work.⁴⁶⁶

5.10 It is important to distinguish between the copyright existing, on the one hand, in the individual components deposited in a database (that is, discrete data items and datasets) and, on the other hand, in the database as a whole (that is, in the arrangement of the collected components). Copyright interests may co-exist independently in components contained within the database and in the database itself, and may be owned by different entities. The fact that a data item is included in a database does not mean that any pre-existing copyright in that item is lost. Nor does a data item which is not protected by copyright (for example, because it consists of a non-original set of facts) attract copyright simply by virtue of its inclusion in a copyright-protected database.

5.11 In order to make a database functional and searchable, metadata will usually be included with individual data items to provide information about the data item, the copyright owner or custodian of the data item and any relevant licensing terms applying to the use of the data. Metadata, on its own, is unlikely to attract copyright. However, metadata or a compilation of numerous metadata records may be protected by copyright if it is arranged or organised in such a way as to satisfy the originality threshold.

⁴⁶² In *DNA: The Secret of Life*, James Watson relates how the question was raised during a seminar at Cold Spring Harbour in 1953, shortly after the double helix structure was published. Academic debate in the 1980s saw a number of proponents for the idea of copyright protection for biological sequences in the United States. Some have argued against copyright protection on the grounds that legislation does not permit copyright protection for a 'discovery': Scott McBride, 'Bioinformatics and IP Protection' (2002) 17 *Berkley Technology Law Journal* 1331. Others have used the doctrine of 'merger', which applies where there is only one way to express an idea or fact, to argue against copyright protection in the U.S.

⁴⁶³ Ibid, see also G Karnell, 'Protection of Results of Genetic Research by Copyright or Design Rights?' (1995) 17 *European IP Review* 355, 357 and L Eckstrom et al, *Eckstrom's Licensing in Foreign and Domestic Operations* (1999) vol 2, [13.10].

⁴⁶⁴ Hugh Laddie, Peter Prescott and Mary Vitoria, *The Modern Law of Copyright and Designs* (2nd ed, 1995) 857. See also A Speck, 'Genetic Copyright' (1995) 4 *European IP Right Review* 171.

⁴⁶⁵ Gunnar Karnell, 'Protection of Results of Genetic Research by Copyright or Design Rights' [1995] 8 *European IP Review* 355.

⁴⁶⁶ See especially, Australian Law Reform Commission, *Genes and Ingenuity: Gene Patenting and Human Health*, Report No 99 (2004) Ch 28

<<http://www.austlii.edu.au/au/other/alrc/publications/reports/99>>; see also Sue Coke, 'Copyright and Gene Technology' (2002) 10 *Journal of Law and Medicine* 97 and Natalie Derzko, 'Protecting Genetic Sequences under the Canadian Copyright Act' (1993) 8 *IP Journal* 31.

5. Duration of Copyright

5.12 As a general rule, copyright lasts for the duration of the life of the person who has created the material, plus a further 70 years.⁴⁶⁷ Questions arise as to how the duration of copyright is to be calculated for electronic databases that are constantly being updated.⁴⁶⁸ The Copyright Law Reform Committee (CLRC) considered this issue in its *Report on Computer Software Protection* (1995). The CLRC noted that for any published database there will be a time when the first edition of the database is made. The CLRC concluded that it is this date from which the copyright term begins. Any addition of information to the database will not give rise to a new edition, unless the database is changed so substantially that it may be regarded as a new work. To determine whether a new work has been created, regard must be had to the quality and quantity of the material added to the database.

6. Originality

5.13 In order to attract copyright protection, a work must be original. Originality in the copyright context does not mean that the work must be inventive or novel. Rather, it means that the work must originate from the author and must not be copied from another person's work. Issues arise in the case of data compilations because data is generally extracted from existing materials which do not originate with the "author" of the database (who will usually be the person who has compiled the data). Questions have therefore arisen about whether data compilations have the requisite originality for copyright protection.

5.14 This issue was addressed in *Desktop Marketing Systems Pty Ltd v Telstra Corporation Ltd*,⁴⁶⁹ which considered the question of whether Telstra held copyright in their Yellow Pages and White Pages directories. The directories contain the names, addresses and phone numbers of certain people and businesses in a given region, listed in alphabetical order. In a landmark judgment, the Full Court of the Federal Court held that Telstra did own copyright in its White and Yellow Pages directories,⁴⁷⁰ thereby confirming that the innovation threshold of originality required to sustain copyright in a compilation is quite low.

5.15 The court in *Desktop Marketing System Pty Ltd v Telstra Corporation Limited*⁴⁷¹ held that copyright can be claimed in a compilation produced as a result of either the exercise of skill, judgment or knowledge in the selection, presentation or arrangement of the materials or for a compilation which has required the investment

⁴⁶⁷ This duration term applies to published literary, dramatic, artistic (apart from photographs) and musical works where the author is living and identified. Anonymous works or works published after the death of the author (except for computer programs) have a copyright term of 70 years from publication. See *Copyright Act 1968* (Cth) s 33.

⁴⁶⁸ Copyright Law Reform Committee, *Report on Computer Software Protection* (1995) Ch 14 <http://www.ag.gov.au/agd/WWW/clrHome.nsf/Page/Overview_Reports_Computer_Software_Protection> at 19 February 2007.

⁴⁶⁹ [2002] FCAFC 112. See further Brian Fitzgerald and Cheranne Bartlett, 'Database Protection under Australian Copyright Law: *Desktop Marketing Systems Pty Ltd v Telstra Corporation* [2002] FCAFC 112' (2003) 7 *Southern Cross University Law Review* 308

⁴⁷⁰ The directories contained the names, addresses and telephone numbers of telephone service subscribers.

⁴⁷¹ [2002] FCAFC 112.

of substantial labour or expense in collecting the material included in the compilation (the so-called “sweat of the brow” approach).⁴⁷² Where copyright subsists in a factual compilation on the basis of the labour or expense required to collect the information, the compiler must show that the labour or expense exceeds a minimum threshold, but this is a question of fact and degree.⁴⁷³ The court in *Desktop Marketing System Pty Ltd v Telstra Corporation Limited* held that since Telstra had undertaken substantial labour and incurred substantial expense in compiling and listing data in the White and Yellow Pages directories - “industrious collection” – it had satisfied the originality requirement, notwithstanding that there may have been minimal intellectual input or creativity in the process of selecting and arranging the material.

5.16 The approach to be taken by the court was explained by Lindgren J as follows:

Comparatively, little work went into the alphabetisation of the entries (as opposed to the considerable labour involved in gathering and checking the data). As will appear below, in my opinion, at least in the case of a factual compilation intended to be a work of utility, infringement must be tested by reference to the interest which copyright is intended to protect in the particular case. In the present case, that interest was the labour and expense of gathering together in the one place the details of all the members of a given universe - all the telephone subscribers in a region.

5.17 Thus, a data compilation, usually in the form of a database, will attract copyright protection provided that a sufficient degree of labour or expense has gone into collecting the data included in the compilation or a sufficient degree of skill, judgment or knowledge has gone into selecting the data for the compilation.

7. Ownership

5.18 The basic principle of copyright ownership is that the author or creator of the material is the first owner of copyright. In the case of databases and datasets, the author will be the person who has compiled the information. For a computer program underlying a database, the author will be the person who has written the code which, in machine readable form, enables the computer to perform a task. Where the information deposited into the database is in the form of an article or other such work, the copyright owner of that individual work will be the person who has written it.

5.19 It will often be the case that data is compiled not only by one person acting alone, but by multiple different people working in a team. If each team member has contributed a separate and distinct part of the overall work, then each individual will hold the copyright in their distinct part only and no single person will hold the copyright in the overall work as a whole. Yet if it is impossible to distinguish between the work of the different authors (i.e. the team was working together so that all output is truly a collaborative effort), then those authors will hold the work as joint owners. This means that the team members all own the copyright equally, and copyright cannot be assigned or licensed without the consent of all owners. At law, joint owners take copyright as “tenants-in-common.” The significance of this is that where copyright is infringed, one of the joint owners can bring an infringement action

⁴⁷² *Desktop Marketing Systems Pty Ltd v Telstra Corporation* [2002] FCAFC 112, [409].

⁴⁷³ *Ibid.*

without having to include the others. Furthermore, it is possible for one joint owner to sue another joint owner for infringement.

5.20 There are, however, exceptions to the general rule that the author of a work is the owner of copyright. Data may be compiled for a particular institution by employees of that institution. If the work has been performed in the ordinary course of an employee's duties, then generally the employing institution will own copyright in the work instead of the employee. If the employing institution is the Crown, that is, a government body, copyright will be owned by the Crown.

8. Assignment

5.21 Even though an author will own copyright in their work in the first instance, they may wish to assign their rights to someone else in return for payment or for the provision of a service. An author may assign copyright in full or only partially. A full assignment involves transferring copyright completely and absolutely to another, leaving the original owner with no residual copyright in the work. A partial assignment limits the assignment to a certain right, place or time. For example, the owner of copyright in a computer program may assign the right of reproduction to another to be exercised in Australia only or for a period of five years only. To be effective at law, assignments must be in writing and must be signed by or on behalf of the assignor.

9. Ownership in Computer-Generated Works

5.22 There has been some discussion as to whether, and how, copyright applies to computer-generated works, such as where a database is created or updated by a computer program. The CLRC considered this issue in their *Report on Computer Software Protection* (1995). The CLRC drew a distinction between materials created with the assistance of computer programs and materials created by computer programs. For the former, they determined that the computer is a mere tool (albeit a sophisticated one) and that copyright does attach to the material as a work. The author, or copyright owner, of the work is:

the person or persons who devised the instructions and originated the data used to control and condition the computer to produce the particular result.⁴⁷⁴

5.23 For the latter, however, the CLRC did not consider that computer-generated materials (i.e. materials created purely by a computer with no human input, such as satellite images picked up by sensors) should be protected as copyright "works", because they lack a human author. The CLRC was of the opinion that if computer-generated materials were to be protected, they should fall into the category of "subject-matter other than works" and attract the neighbouring rights akin to those extended to performers. The author of computer-generated material, if one had to be found, would probably be either the programmer or owner of the copyright in the programs that assisted in creation of the material, or the provider of the data. To date,

⁴⁷⁴ Copyright Law Reform Committee, *Report on Computer Software Protection*, (1995) ch 13 <http://www.ag.gov.au/agd/WWW/clrHome.nsf/Page/Overview_Reports_Computer_Software_Protection> at 19 February 2007.

the suggestions of the CLRC have not been adopted, and the *Copyright Act* makes no specific reference to computer-generated materials.

10. Rights

5.24 Copyright confers on the copyright owner the exclusive proprietary right to do a range of acts in relation to the protected work. The copyright owner also has the exclusive right to authorise the doing of any of those acts by someone else, whether by sale (assignment) or by licence.

5.25 The economic rights of the owner of copyright in a literary, dramatic or musical work are the rights to:

- reproduce the work in a material form;
- publish the work;
- perform the work;
- communicate the work to the public;
- make an adaptation of the work (for example, a translation); and
- control rental of the work, where the work is a computer program or is reproduced in a sound recording.⁴⁷⁵

5.26 For artistic works, the exclusive rights are the rights to:

- reproduce the work in a material form;
- publish the work; and
- communicate the work to the public.⁴⁷⁶

5.27 The primary right of a copyright owner is the right to reproduce. Reproduction includes converting a work into or from a digital or other machine-readable form. A work will be reproduced if it is copied in its entirety or if a substantial part of the work is reproduced. A computer program is taken to have been reproduced if an object code version of the program is derived from the program in source code (for example, by compilation) or if a source code version of the program is derived from the program by object code (for example, by decompilation).

5.28 The right to communicate to the public is an important right in the online environment. “Communicate” is defined to mean:

make available online or electronically transmit (whether over a path, or a combination of paths, provided by a material substance or otherwise) a work or other subject-matter.⁴⁷⁷

5.29 Making a database available online or through a computer network will be an exercise of the copyright owner’s right to communicate the database to the public. A communication may be to the public even though it is made to a relatively small number of individuals, provided the recipients are part of the copyright owner’s public in the sense that he or she could expect to be remunerated for permitting the work to be communicated to that audience.

⁴⁷⁵ *Copyright Act 1968* (Cth) s 31(1).

⁴⁷⁶ *Copyright Act 1968* (Cth) s 31(1).

⁴⁷⁷ *Copyright Act 1968* (Cth) s 10(1).

11. Infringement

5.30 In the absence of a defence or limitation provided by the *Copyright Act 1968*, it is an infringement to do an act within the scope of the exclusive rights of the copyright owner without the owner's prior authorisation. An infringing act may relate to the whole or a substantial part of the copyright material, but will not extend to an "insubstantial" part of a copyright work or other subject matter. It is also an infringement to authorise someone else to do an act within the copyright owner's exclusive rights without the owner's permission. A copyright owner can bring legal action against an infringer and recover damages to compensate them for any loss suffered as a result of the infringement.

5.31 There are defences or exceptions to infringement that allow a person to exercise a copyright owner's exclusive right in certain circumstances or for a particular purpose. The exceptions relevant to data are as follows.

5.32 **Fair dealing for research or study:** It is permissible to copy a "reasonable portion" of a literary, dramatic, musical or artistic work for the purpose of private research or study.⁴⁷⁸ A "reasonable portion" is defined in s 40(5) of the *Copyright Act*, but notably will not apply to a computer program or an electronic compilation such as a database. This fair dealing exception may be relevant to reports and articles contained in a database, but not to the database itself. Additionally, the court has held that:

Multiple copying by an education institution that would otherwise fall under a compulsory licence cannot be excused as fair dealing.⁴⁷⁹

5.33 **Temporary reproduction:** Particularly relevant to users accessing online databases is the exception in s43A *Copyright Act 1968*, which provides that copyright in a work will not be infringed where the work is temporarily reproduced as part of the technical process of making or receiving a communication. This exception is designed to apply to the temporary reproduction of material by a user's computer, which occurs when a user accesses material from a website and views it on their computer screen, even if the material is not downloaded. Similarly, s 22(6A) states that a person is not responsible for determining the content of a communication merely because they take steps in order to gain access to what has been made available online by someone else or to receive the communication.⁴⁸⁰ This provision is intended to make it clear that a person who merely accesses or browses material online is not making a communication to the public and that engaging in the technical processes necessary to obtain access to or receive content that has been communicated by another party is not an act of communication.

5.34 **Interoperable program:** Section 47D *Copyright Act 1968* permits the reproduction or adaptation of a computer program for the purpose of creating an

⁴⁷⁸ *Copyright Act 1968* (Cth) s 40(5).

⁴⁷⁹ *Copyright Licence Ltd v University of Auckland* (2002) 53 IPR 618.

⁴⁸⁰ *Copyright Act 1968* (Cth) s 22(6A) was inserted by the *Copyright Amendment Act 2006*. It provides 'To avoid doubt, for the purposes of [s 22(6)], a person is not responsible for determining the content of the communication merely because the person takes one or more steps for the purpose of:

- (a) gaining access to what is made available online by someone else in the communication; or
- (b) receiving the electronic transmission of which the communication consists.'

interoperable program or article. Additionally, s47B(1) provides that a reproduction of a computer program will not be an infringement where it is:

Incidentally and automatically made as part of the technical process of running a copy of the program for the purposes for which the program was designed.

5.35 Libraries: Under s49(5A) *Copyright Act 1968*, libraries may provide users with access to material acquired in electronic form, provided that access occurs on the library premises and that the user cannot use library equipment to copy or communicate the material.

12. Digital Rights Management (DRM) and Electronic Rights Management Information (ERMI)

5.36 A technological solution that has been developed to protect the interests of copyright owners in the digital environment is digital rights management (DRM). DRM employs technology in the distribution, communication or usage of digital content to express rights in digital materials and regulate the exercise of those rights.⁴⁸¹ Various technologies are available to protect copyright materials from unauthorised access and use, the most common being encryption,⁴⁸² digital watermarking⁴⁸³ and embedding rights management information. Legal recognition of the use of DRM to protect copyright materials was introduced into Australian law by the *Copyright Amendment (Digital Agenda) Act 2000*. The Act contains anti-circumvention provisions which make it unlawful to deal in devices or services designed to circumvent technological protection measures. The Act also contains prohibitions on the removal or alteration of electronic rights management information (ERMI).

5.37 ERMI is a technological mechanism that is increasingly being used by copyright owners to protect their copyright material. ERMI is defined in the *Copyright Act 1968*⁴⁸⁴ as electronic information (including numbers or codes representing such information) which is either attached to or embodied in the copyright material, or appears in connection with a communication or the making available of the copyright material. It typically includes information about the copyright material, the copyright owner's details and the terms and conditions on which the material may be used. ERMI allows digital copyright material to be described, identified, monitored and tracked, enabling a copyright owner to potentially monitor every instance of access to and use of their copyright material.

5.38 It is an infringement of the copyright owner's rights to remove or alter ERMI relating to a copyright work or other subject matter without the permission of the

⁴⁸¹ For an overview of many technological and legal issues relating to digital rights management, see Reihaneh Safavi-Naini and Moti Yung (eds), *Digital Rights Management: Technologies, Issues, Challenges and Systems* (2006). Note, in particular, the chapter by Yee Fen Lim, *Digital Rights Management: Merging Contract, Copyright and Criminal Law*, 66–74.

⁴⁸² Encryption involves the scrambling of the information embedded within a digital object so that it cannot be used without a password.

⁴⁸³ Digital watermarks (which can be visible or invisible) embed information (eg about the author, publisher, terms and conditions of use) into the data and removing them causes the quality of the data to be severely degraded.

⁴⁸⁴ *Copyright Act 1968* (Cth) s 10(1).

copyright owner or exclusive licensee, if the person doing the act knows or ought reasonably to have known that the removal or alteration would induce, enable, facilitate or conceal an infringement of copyright.⁴⁸⁵ Unless the defendant proves otherwise, it is presumed that they knew or ought reasonably to have known that the removal or alteration of ERMI would induce, enable, facilitate or conceal an infringement of copyright.⁴⁸⁶

5.39 Where ERMI relating to a copyright work or other subject matter has been removed or altered, it is an infringement to:

- distribute a copy of that work or other subject matter to the public;
- import into Australia a copy of that work or other subject matter for distribution to the public; or
- communicate a copy of that work or other subject matter to the public.⁴⁸⁷

5.40 To be liable, the person dealing with the work or other subject matter must know that the ERMI was removed or altered without the permission of the copyright owner or exclusive licensee and know or ought reasonably to have known that the distribution, importation or communication of the material would induce, enable, facilitate or conceal an infringement of copyright.

5.41 These actions can also constitute criminal offences under sections 132AQ to 132AS of the *Copyright Act*. Where a person is charged with an offence⁴⁸⁸ (whether or not the person is convicted of the offence), the court may make an order requiring the defendant to destroy, deliver up to the copyright owner or otherwise deal with any article in the defendant's possession that appears to the court to be:

- a circumvention device used or intended to be used in conduct that amounts to an offence;
- an infringing copy; or
- a device or equipment used or intended to be used for making infringing copies.⁴⁸⁹

13. Licences

5.42 A copyright owner can choose to allow a person to do an act that would otherwise infringe the copyright owner's exclusive rights. They can do this by issuing a licence. A licence differs from an assignment in that where a licence is granted, the licensor is only *permitting* a licensee to do certain acts but is not transferring the copyright. The copyright owner retains all of their exclusive rights, unless the licence expressly provides that they cannot exercise those rights in some capacity.

⁴⁸⁵ *Copyright Act 1968* (Cth) s 116B(1).

⁴⁸⁶ *Copyright Act 1968* (Cth) s 116B(3).

⁴⁸⁷ *Copyright Act 1968* (Cth) s 116C(1).

⁴⁸⁸ Except an offence under *Copyright Act 1968* (Cth) s 132AM.

⁴⁸⁹ *Copyright Act 1968* (Cth) s 133.

5.43 A licence can be exclusive or non-exclusive. An exclusive licence grants the relevant right to the licensee only, and the licensor cannot then grant the same right to anyone else. A non-exclusive licence, however, reserves for the licensor the option of granting the same right in different licences to other interested persons. A licence need not be in writing unless it is an exclusive licence.

5.44 A licence can be contractual or non-contractual. A contractual licence, as the name implies, operates like a standard contract. This allows the copyright owner to not only give permission for use of the material, but to also impose restrictions or additional conditions on its use. For example, a copyright owner may licence the use of a dataset compilation to another person on the condition that if the person uses the dataset commercially, they must account back to the copyright owner for some of the profits. A non-contractual licence is essentially a bare permission to use the material for certain purposes. A copyright owner can still limit rights issued under a non-contractual licence – for example, the licence may be for the right of reproduction only, or may only extend to use within Australia – but the other person is not contracting to fulfill any additional obligations. An increasingly common form of non-contractual licensing is the open content licence.

5.45 Whether or not a licence can be revoked will depend on the nature of the licence. A contractual licence should provide for this in its terms, and a non-contractual licence may also state whether the licence can be revoked. In the absence of any express statement, it seems that an exclusive licence will usually be irrevocable, a bare permission may be revocable at will, and any licence falling between these two extremes may only be revoked if there cannot be read into the licence an implied promise not to revoke.⁴⁹⁰

Licences and Databases

5.46 The various types of data included in datasets and databases raise different legal issues which will impact on the licensing practices adopted. Some types of content will simply not be susceptible to licensing re-use at all because of the sensitivity of the data (for example, due to privacy or confidentiality considerations) or because the database manager does not own or has not obtained licences to exercise the intellectual property rights in the data. The issue of third party rights included in datasets must be addressed operationally before the database manager licences other parties to obtain access to or re-use the data. This will involve a consideration of any third party rights limiting or prohibiting certain uses of data inputs as well as a consideration of statutory or policy constraints.

5.47 Licences are particularly relevant to databases because a user accessing material in a database will need to know what rights they have to download and use the material (if it is indeed copyright protected and not simply raw data, which does not attract copyright). It is important for the database manager that the contributor of the data can warrant that they have the necessary rights to deposit the data and to make the data available for access by others. Both the grant of rights to the database and the grant of rights to end-users can be determined and governed by licences. At

⁴⁹⁰ Jill McKeough, Andrew Stewart and Philip Griffith, *Intellectual Property in Australia* (3rd ed, 2004) 202-203.

the time of contribution, a researcher should be required to enter a Deposit Agreement that sets out these rights and responsibilities. The agreement may take the form of a click-wrap online document so that the researcher has ready access to the agreement at the time of deposit. Any rights granted to end-users can be determined either by the Deposit Agreement or by a separate licence the researcher grants directly to end-users, usually through an open content licensing model.

Open Content Licensing

5.48 The development of open content licensing models has made it easier for copyright owners to licence their material to a wider range of people, particularly over the Internet. Open content licensing involves making copyright material available on liberal terms, to ensure that it is readily accessible and available for re-use. A central feature of open content licensing is that while copyright is asserted in the material, the copyright owner exercises their rights to ensure ready accessibility and to permit re-use while still reserving some rights for their sole benefit (for example, to be attributed as the creator of the material).

5.49 Importantly, open content licences grant rights to users to do acts that fall within the scope of the copyright owner's exclusive rights and do not impose further (i.e. non-copyright related) obligations on the users of the copyright material. Open content licences differ from many traditional information licences which seek to impose contractual obligations or constraints on users (for example, restrictions on further dissemination of information or confidentiality obligations). The open content licensing model is of particular relevance to systems designed to facilitate access to and re-use of public sector materials because it acknowledges government ownership of copyright and sets conditions by which public sector information may be accessed and re-used, particularly in the digital environment.

Creative Commons

5.50 The leading model of open content licensing is that developed by the Creative Commons project⁴⁹¹ which was established by Professor Lawrence Lessig (Stanford University) and others in 2001. Through the iCommons project, the Creative Commons licence is now established in about 40 countries worldwide, including Australia. Its aim is to increase the amount of raw source material available online and to make access to such material easier and cheaper.

5.51 The Creative Commons project has developed a set of copyright licences which make copyright works freely available for use, on certain conditions. The Creative Commons project:

Facilitates public access to copyrighted literary and artistic works by devising a set of standard-from contractual templates any author can digitally adopt⁴⁹².

⁴⁹¹ For more information see <<http://www.creativecommons.org>>.

⁴⁹² J Reichmann and P Uhler, 'A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment' (2003) 66 *Law and Contemporary Problems* 315, 431
<<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

5.52 In addition to certain “baseline” rights and restrictions which are included in all Creative Commons licences, the copyright owner can choose from among certain optional features, which can be used alone or in combination.

Creative Commons - Baseline Features:

5.53 The following features are common to all Creative Commons licences:

- licensees are granted the right to copy, distribute, display, digitally perform and make verbatim copies of the work into another format;
- the licences have worldwide application that lasts for the entire duration of copyright and are irrevocable;
- licensees cannot use technological protection measures to restrict access to the work;
- copyright notices should not be removed from any copies of the work; and
- every copy of the work should maintain a link to the licence.

5.54 Additionally, all Creative Commons licences contain the condition:

- Attribution: Others are permitted to copy, distribute, display and perform the copyright work - and any derivative works based upon it - but only if they give credit to the creator of the copyright work.

Creative Commons - Optional Features:

5.55 Copyright owners can choose from among the following optional licence conditions:

- Non-commercial: Others are permitted to copy, distribute, display and perform the copyright work - and any derivative works based upon it – but for non-commercial purposes only;
- No derivative works: Others are permitted to copy, distribute, display and perform only exact copies of the work but cannot make derivative works based upon the original work;⁴⁹³ and
- Share Alike: Others may distribute derivative works only under a licence identical to that in the original work.⁴⁹⁴

5.56 Each Creative Commons licence is expressed in three ways:

- (1) the Commons Deed: a simple, plain-English summary of the licence, together

⁴⁹³ Note that the ‘No derivative works’ option is incompatible with the ‘Share alike’ option.

⁴⁹⁴ Note that the ‘Share alike’ option only applies to derivative works and is in-compatible with the ‘No derivative works’ option.

with the relevant icon/s that indicates the scope of permitted use;

(2) the Legal Code: the “fine print” licence document; and

(3) the Digital Code: the machine-readable translation of the licence that helps search engines and other applications identify the copyright work by its terms of use.

5.57 When a copyright owner chooses to use a Creative Commons licence, they also obtain metadata (licensing information) which is encoded in RDF (Resource Description Framework). Creative Commons metadata can be embedded in the copyright work in a variety of formats.⁴⁹⁵

5.58 The last few years have seen an increasing appreciation of open content licences to grant access to copyright-protected compilations of data in open collaborative research projects. The use of Creative Commons licences to facilitate access to copyright-protected research data was highlighted in an Editorial in *Nature* in December 2005:

Scientists may be justified in retaining privileged access to data that they have invested heavily in collecting, pending publication, but there are also huge amounts of data that do not need to be kept behind walls. And few organizations seem to be aware that by making their data available under a Creative Commons licence (see <http://creativecommons.org/licence>), they can stipulate both rights and credits for the reuse of data, while allowing its uninterrupted access by machines.⁴⁹⁶

The same point has been made more recently by Tapscott and Williams:

Teams of scientists that invest heavily in collecting data, and understandably feel justified in retaining privileged access to it, could apply Creative Commons licences that stipulate rights and credits for the reuse of data, while allowing uninterrupted access by networked computers.⁴⁹⁷

5.59 In the report produced for the JISC-funded GRADE project, *Designing a Licensing Strategy for Sharing and Re-Use of Geospatial Data in the Academic Sector* (2007), Waelde and McGinley commented that where copyright material (such as visual images or text) is deposited into databases by researchers:

consideration should be given as to whether the work might be licensed under Creative Commons or a similar licence.⁴⁹⁸

5.60 As explained on the Science Commons website, Creative Commons licences can be used in relation to databases. An FAQ prepared by Mia Garlick, former

⁴⁹⁵ For further information, see Creative Commons, *Using Creative Commons Metadata* (2006) <<http://creativecommons.org/technology/usingmarkup>> at 30 January 2007.

⁴⁹⁶ Editorial, ‘Let data speak to data’ (2005) 438 *Nature* 531 <<http://www.nature.com/nature/journal/v438/n7068/full/438531a.html>> at 24 April 2007.

⁴⁹⁷ Don Tapscott and Anthony Williams, ‘The New Science of Sharing’, *BusinessWeek.com* 2 March 2007, <http://www.businessweek.com/innovate/content/mar2007/id20070302_219704.htm?chan=technology_technology+index+page_more+of+today's+top+stories> at 12 March 2007.

⁴⁹⁸ Charlotte Waelde and Mags McGinley, *Designing a licensing strategy for sharing and re-use of geospatial data in the academic sector*, GRADE (2007) <<http://edina.ac.uk/projects/grade>>.

General Counsel for Creative Commons, explains that databases are usually comprised of four elements:⁴⁹⁹

- a set of field names identifying the data;
- a structure, which includes the organisation of fields and relations among them;
- data entry sheets; and
- data.

5.61 Where any one or more of these elements attracts copyright, a Creative Commons licence can be used to license that copyright.⁵⁰⁰ It is important to remember that the threshold for copyright protection is higher in the United States than in Australia, with the result that, in the United States, a compilation of data will only be protected by copyright if a sufficient degree of creativity was involved in producing it.⁵⁰¹

5.62 The Science Commons FAQ advises database providers who are applying a Creative Commons licence to their database to make clear:⁵⁰²

- which elements of the database to which the Creative Commons licence attaches, based on the existence of copyright in those elements;
- where applicable, that the licence only applies to the database elements and not the underlying software; and
- which parts of the database, being raw data or information, are not subject to copyright and therefore free to be used and reused, independently of the Creative Commons licence.

5.63 The Creative Commons licence has also been the inspiration for other similar open content licences. For example, BBC developed the Creative Archive Licence to make content from its archive available through an open and transparent process.⁵⁰³ The Creative Archive Licence is essentially the same as the Creative Commons attribution, non-commercial, share-alike licence. However, there are two additional restrictions that are not found in Creative Commons licences – the Creative Archive Licence is restricted to UK use only, and the licensed content cannot be used to promote political, charitable or other campaigning purposes.

⁴⁹⁹ See <<http://sciencecommons.org/resources/faq/databases>> at 18 June 2007, Licensed under a Creative Commons Attribution 3.0 License

⁵⁰⁰ Ibid

⁵⁰¹ Ibid; For information on the operation of Creative Commons licences in relation to the European Database right, see also <<http://sciencecommons.org/resources/faq/databases>>

⁵⁰² Ibid

⁵⁰³ See Creative Archive Licence Group, *The Creative Archive Licence* (2006) <<http://creativearchive.bbc.co.uk/index.html>> at 30 January 2007.

Science Commons

5.64 Science Commons is a project related to Creative Commons that extends the Creative Commons open content principles to scientific data and publications.⁵⁰⁴ The aim of Science Commons is to:

achiev[e] for the world of science and data, what Creative Commons had begun to achieve for the world of culture, art and education material: to ease unnecessary legal and technical barriers to sharing, to promote innovation, to provide easy, high quality tools that let individuals and organizations specify the terms under which they wished to share their material.⁵⁰⁵

5.65 This is a difficult and complicated task. As John Wilbanks, Executive Director of Science Commons, and Professor James Boyle state:

Creating an open regime of sharing and reuse in the sciences is a complicated proposition. Though copyrights guard the final published documents in peer reviewed journals, patents protect inventions (some more unique than others) and a web of handshakes and contracts guard the tools, materials, datasets, databases and informal knowledge transfer of day-to-day science. What works for a biologist will likely fail for a physicist, neither of whose solutions will perfectly solve the legal problems of the anthropologist.⁵⁰⁶

5.66 Science Commons seeks to promote open access to scientific information through the following range of mechanisms:

- encouraging open access publishing and self-archiving in open databases;
- developing author addenda for publishing agreements;
- licensing mechanisms, to apply Creative Commons-type licences to scientific data and publications; and
- streamlining Material Transfer Agreements (commonly used by scientific researchers exchanging information and data) to make data sharing easier.⁵⁰⁷

5.67 The directors of Science Commons have discussed with pharmaceutical companies the possibility of creating a “tox commons” in which researchers could pool toxicity data from failed commercial drug attempts:

The idea is simple. While a successful drug application results in open data...every failed drug results in secrets and obscurity. So a tempting target, tried and again and again, can mean repetition of failure.⁵⁰⁸

⁵⁰⁴ See <<http://sciencecommons.org/>> at 14 June 2007

⁵⁰⁵ John Wilbanks and James Boyle, “Introduction to Science Commons”, August 2006, available at <http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf> at 14 June 2007, p5

⁵⁰⁶ Ibid, p4

⁵⁰⁷ See John Wilbanks and James Boyle, “Introduction to Science Commons”, August 2006, available at <http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf> at 14 June 2007; and <<http://sciencecommons.org/projects/>> at 14 June 2007

5.68 Science Commons is a supporter of the concept of the “semantic web”. The semantic web allows the searching and interchange of data by using advanced automated technologies that allow searching by functions or meaning, not just searching by words.⁵⁰⁹

At its most ambitious, it would allow seamless integration between scholarly articles, the data those articles refer to, and to cross references with other articles dealing with similar processes in different areas of science. But the process of mining, linking, tagging and cross-referencing that the semantic web requires faces extraordinary difficulties...Some of the difficulties involve the coordination of standards and formats for metadata, something that Creative Commons has considerable experience in.⁵¹⁰

Statutory Licences

5.69 The *Copyright Act 1968* establishes several statutory licences which allow use of copyright material without the authorisation of the copyright owner, provided that the copyright owner is compensated for the use of the material. These operate automatically (provided the requisite fee is paid) and are distinct from voluntary licences granted by copyright owners on their own terms.

5.70 The statutory licences most applicable to data are:

- copying by educational institutions and institutions assisting persons with an intellectual or print disability; and
- use of copyright material by Commonwealth and State governments.

5.71 Part VB *Copyright Act 1968* sets out two statutory licences for equitable remuneration for education institutions and institutions assisting persons with disabilities:

- the “hardcopy licence” for reproduction of works or parts of works when the original is in hardcopy; and
- the “electronic reproduction and communication licence” for reproduction and electronic communication of works or parts of works that are in electronic form.

5.72 The Copyright Agency Limited (CAL) is the collecting society for the owners of copyright in works for the purposes of Part VB.

⁵⁰⁸ John Wilbanks and James Boyle, “Introduction to Science Commons”, August 2006, available at <http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf> at 14 June 2007, p8

⁵⁰⁹ See John Wilbanks and James Boyle, “Introduction to Science Commons”, August 2006, available at <http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf> at 14 June 2007, p11; and <<http://sciencecommons.org/projects/>> at 14 June 2007

⁵¹⁰ John Wilbanks and James Boyle, “Introduction to Science Commons”, August 2006, available at <http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf> at 14 June 2007, p11

5.73 CAL is also the collecting agency for instances where the State and Commonwealth governments reproduce copyrighted work for the services of the Commonwealth or State. This arrangement does not apply to the exercise of any of the copyright owner's rights other than the right to reproduce in material form.

14. Other Jurisdictions

5.74 Under Australian copyright law, many factual databases will be regarded as sufficiently original to attract copyright protection. In other countries, the legal principles relating to compilations of data have developed differently. Databases situated in these jurisdictions will often not receive the same protection under copyright law as in Australia. Many of the databases accessed by Australian researchers online are situated overseas, usually in the United States and Europe, and will be subject to the laws of those jurisdictions.

United States

5.75 Copyright protection for US databases was considered in *Feist Publications v Rural Telephone Service*⁵¹¹ ("Feist"). In that case, the Supreme Court of the United States rejected the "sweat of the brow" theory and held that Rural did not have copyright in its white pages telephone directory containing subscriber information arranged alphabetically.⁵¹² Unlike the position in Australia, some degree of creativity must be exercised in the compilation of the database for copyright to exist.

5.76 The Supreme Court found that a compilation, even one composed purely of non-copyrightable elements (such as mere factual information), is protected under copyright law so long as the author of the compilation exercised creativity in the selection, coordination, or arrangement of those elements.⁵¹³ Where the author of the compilation adds no written expression but lets the facts speak for themselves, the expressive element is more elusive. The only conceivable expression is the manner in which the compiler has selected and arranged the facts. If the selection and arrangement are original, these elements of the work are eligible for copyright protection. No matter how original the format, however, the facts themselves do not become original through association.

5.77 This inevitably means that the copyright in a factual compilation is thin. Notwithstanding a valid copyright, a subsequent compiler remains free to use the facts contained in another's publication to aid in preparing a competing work, so long as the competing work does not feature the same selection and arrangement.

5.78 It is important to emphasise that copyright in the database is limited to the particular selection or arrangement of the components. In *Feist* the Supreme Court stated:

⁵¹¹ *Feist Publications Inc v Rural Telephone Service Co Inc*, 499 US 340 (1991).

⁵¹² Jane Ginsburg, 'No Sweat? Copyright and Other Protection of Works of Information After *Feist v Rural Telephone*' (1992) 92 *Columbia Law Review* 338.

⁵¹³ *Feist Publications Inc v Rural Telephone Service Co Inc*, 499 US 340, 349 (1991).

A factual compilation is eligible for copyright if it features an original selection or arrangement of facts, but the copyright is limited to the particular selection or arrangement.⁵¹⁴

5.79 As in Australia, in the US there is no specific legislative protection for databases such as that introduced in Europe under the Database Directive. While there have been attempts in the US to introduce statutory protection for databases, they have not met with success.⁵¹⁵ Burk explains:

Such statutes have been repeatedly introduced into the US Congress, but never enacted due to opposition from a variety of constituencies, most prominently the scientific research community. Scientists in the US have vigorously resisted the introduction of a form of IP that they fear will restrict access to data, and increase the cost of obtaining commercially available data.⁵¹⁶

United Kingdom

5.80 The UK legal position on intellectual property protection for databases is more complex than in Australia or the US. The complexity arises because as well as the availability of copyright protection for databases which meet the originality threshold, there are specific “database rights” which were introduced into UK legislation as required by the European Union Database Directive (1996).⁵¹⁷

Copyright in Databases – United Kingdom

5.81 In the UK, the copyright legislation has been amended to include a “database” in the list of items to be considered as literary works.⁵¹⁸ “Database” is defined as:

A collection of independent works, data or other materials which: (a) are arranged in a systematic or methodical way; and (b) are individually accessible by electronic or other means.⁵¹⁹

5.82 Databases are excluded from the compilations category and there are differences in how databases and compilations are treated in copyright law. Compilations may acquire copyright protection through the exercise of the author’s skill, judgment or labour, but the position in relation to databases has been changed by the European Union Database Directive.

5.83 Copyright protection in databases in the UK extends only to the structure of the database, not its contents.⁵²⁰ Copyright protection will only attach to a database where the selection and arrangement of contents in the database can be regarded as

⁵¹⁴ *Feist Publications Inc v Rural Telephone Service Co Inc*, 499 US 340, 350-351 (1991).

⁵¹⁵ See Stephen Maurer, ‘Coping with change: IP Rights, New Legislation, and the Human Mutation Database Initiative’ (2000) 15 *Human Mutation* 22; J Reichman and P Uhlir, ‘Database protection at the crossroads’ (1999) 14 *Berkeley Technology Law Journal* 793; S Maurer and S Scotchmer, ‘Database protection: is it broken and should we fix it?’ (1999) 284 *Science* 1129.

⁵¹⁶ Dan Burk, *Intellectual Property in the Context of E-Science*, Minnesota Legal Studies Research Paper No 06-47 (2006) 5 <<http://ssrn.com/abstract=929479>>.

⁵¹⁷ For an overview of the UK law in relation to databases, see Charles Oppenheim, *The legal and regulatory environment for electronic information* (4th ed, 2001) 43 – 53.

⁵¹⁸ *Copyright, Design and Patents Act 1988* (UK) s 3(1).

⁵¹⁹ *Copyright, Design and Patents Act 1988* (UK) s3A.

⁵²⁰ European Union Database Directive Article 3(2).

the author's "intellectual creation."⁵²¹ An author of a database is the person or persons who created it.⁵²² Under this framework, a comprehensive database may be excluded from copyright protection but a more restrictive database or dataset would be protected.⁵²³ Where copyright protection for the database does arise, it will last for the life of the author of the database, plus 70 years.

5.84 The exclusive rights of a copyright owner of a database are:

- to reproduce the database, either temporarily or permanently;
- to translate, adapt, arrange or alter the database;
- to distribute to the public;
- to communicate, display or perform to the public; and
- to reproduce, distribute, communicate, display or perform a translation, adaptation, arrangement or alteration of the database.⁵²⁴

5.85 A copyright owner also has the right to authorise someone else to exercise any of these rights.

5.86 An exception to the rights of the copyright owner applies to give lawful users of a database the right to access or use the database or do anything necessary to access or use the database.⁵²⁵ This exception cannot be excluded or restricted by contract.⁵²⁶

5.87 Under the European Union Database Directive, exceptions to the rights of the copyright owner are limited to:

- reproduction of a non-electronic database for private purposes;
- use for the sole purpose of illustration for teaching or scientific research;
- use for the purpose of public security or for the purposes of administrative or judicial procedure; and
- other traditional exceptions to copyright under the national law of the member state.⁵²⁷

⁵²¹ European Union Database Directive Article 3(1).

⁵²² European Union Database Directive Article 4(1).

⁵²³ Professor Cornish has described this distinction as "silly". See William Cornish, *IP: Patents, Copyright, Trade Marks and Allied Rights* (4th ed, 1999) 523.

⁵²⁴ European Union Database Directive; *Copyright, Designs and Patents Act 1988* (UK) s 21.

⁵²⁵ *Copyright, Designs and Patents Act 1988* (UK) s50D; see also European Union Databases Directive Article 6(1).

⁵²⁶ *Copyright, Designs and Patents Act 1988* (UK) s 296B; European Union Databases Directive Article 6(1).

⁵²⁷ European Union Database Directive Article 6.

5.88 Any exceptions must not unreasonably prejudice the copyright owner's legitimate interest in the database and must not conflict with the copyright owner's normal exploitation of the database.⁵²⁸

The European Database Right

5.89 In addition to copyright protection available for databases, a new *sui generis* (unique) right has been introduced by the European Union Database Directive. Database rights operate irrespective of whether a database or any of its contents attracts copyright protection. The introduction of this right was considered necessary to protect non-original databases, which do not attract copyright protection but which are nevertheless valuable and have required substantial economic investment. The *sui generis* right can also protect public documents in a database that would usually not receive copyright protection, in recognition of:

The fact that governments increasingly create databases that are highly valuable and should have the right to recoup the cost of creating such databases.⁵²⁹

This right does not conflict with rights in the contents of the database, or copyright protection of the database.

5.90 The maker of the database is the owner of the database right.⁵³⁰ The maker is:

The person who takes the initiative in obtaining, verifying or presenting the contents of the database and assumes the risk of investing in that obtaining, verifying or presenting.⁵³¹

If several people act together in activities relating to making the database, they will be joint makers of the database and joint owners of the right.⁵³²

5.91 The database right lasts for a 15-year term commencing at the end of the calendar year in which the database was completed. If the database is made available to the public before the end of that period, the term expires 15 years from the end of the calendar year in which the database was first made available to the public.

5.92 The owner of the database right has the right to object to the extraction or re-utilisation of all or a substantial part of the contents of the database.⁵³³ The key terms "extraction", "re-utilisation" and "substantial" are defined in Regulation 12 of the *Copyright and Rights in Databases Regulations 1997* (UK):

- "substantial" in relation to any investment, extraction or reutilisation means substantial in terms of quantity or quality or a combination of both;

⁵²⁸ Berne Convention Article 9(2).

⁵²⁹ B Fitzgerald et al, *Oak Law Project Report No. 1: Creating a Legal Framework for Copyright Management of Open Access Within the Australian Academic Research Sector* (2006)

⁵³⁰ *Copyright and Rights in Databases Regulations 1997* (UK) r 15.

⁵³¹ *Copyright and Rights in Databases Regulations 1997* (UK) r 14(1).

⁵³² *Copyright and Rights in Databases Regulations 1997* (UK) r 14(5).

⁵³³ *Copyright and Rights in Databases Regulations 1997* (UK) r 16(1).

- “extraction” in relation to any contents of a database, means the permanent or temporary transfer of those contents to another medium, by any means or in any form; and
- “re-utilisation” in relation to any contents of a database, means making all or a substantial part of those contents available to the public, by any means including distributing copies, renting, or by online or other forms of transmission.⁵³⁴

5.93 Under the European Union Database Directive, a person must not repeatedly and systematically extract or re-utilize insubstantial parts of the contents of a database, if such acts conflict with the normal exploitation of the database by the database maker or unreasonably prejudice the legitimate interests of the database maker.⁵³⁵

5.94 The database right was considered in *British Horseracing Board Limited v William Hill Organisation Ltd* [2001] 2 CMLR 12. In that case, the respondent used information obtained from a database owned by the British Horseracing Board to provide his own Internet betting service. The question which arose for determination by Laddie J was whether a substantial part of the database had been extracted or re-utilised. Determination of what constituted a substantial part of the database involved not only a comparison between what was taken (or used) from the database and what remained, but also the importance to the appellant of what had been taken. At first instance, Laddie J found that the respondent had infringed the database right, even though the data used had been obtained from a third party.

5.95 The decision was appealed and was eventually referred to the European Court of Justice (“ECJ”). The ECJ drew a distinction between data collected specifically for inclusion in a database (which would be protected by the database right) and data collected or created for some other purpose but which may later be compiled and included in a database (which would not receive protection). The thrust of the ECJ decision seems to be that the purpose for which data is compiled must be for immediate inclusion in a database, or the database right will not be invoked. The decision of the ECJ was endorsed by the English Court of Appeal.

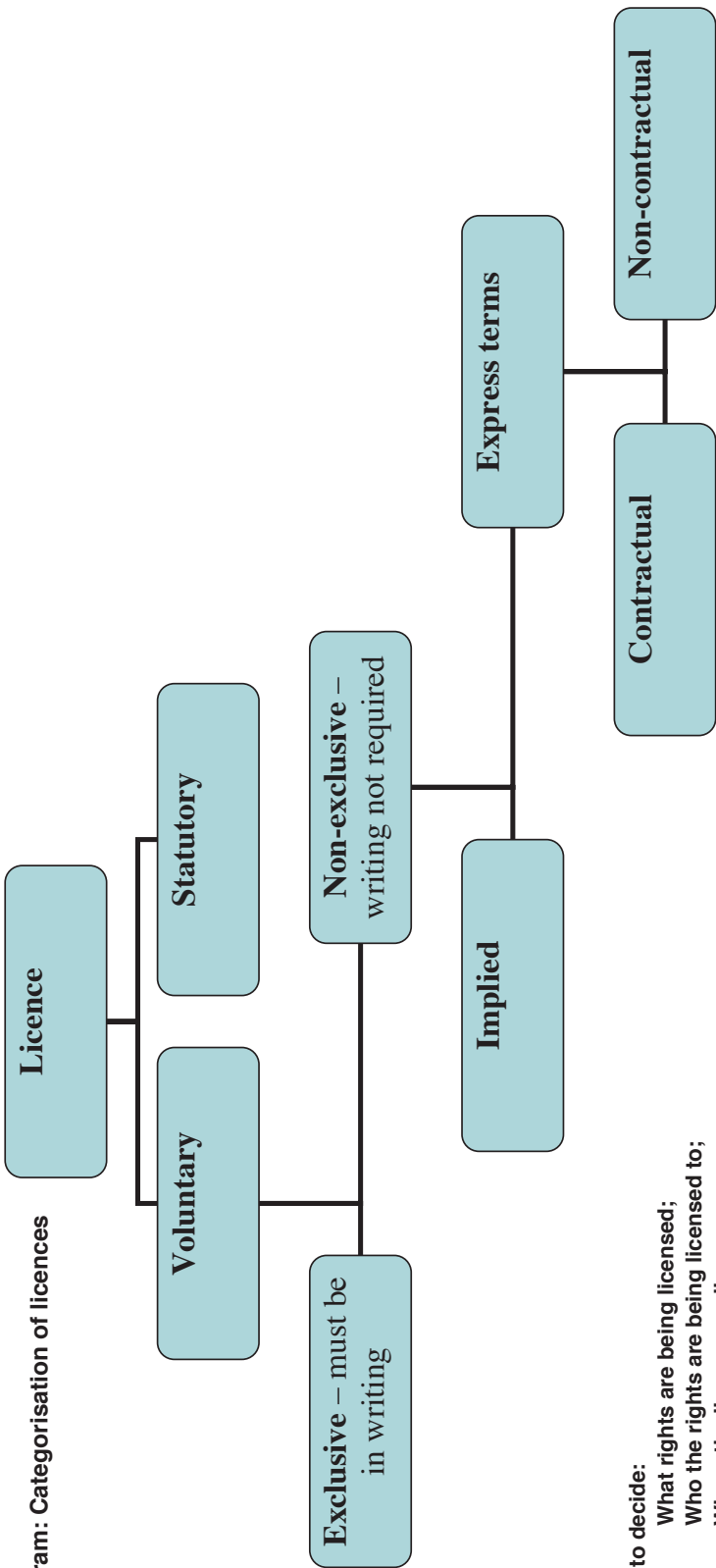
5.96 The database right as introduced by various European jurisdictions has received a mixed reaction, particularly amongst those in the research community who are dependent on bioinformatics databases. While some see the database right as a welcome innovation,⁵³⁶ others have been less enthusiastic and have concluded that copyright provides a better mechanism to balance the interests of bioinformatics database makers and users.⁵³⁷

⁵³⁴ See also European Union Database Directive Article 7(2)(b).

⁵³⁵ See European Union Database Directive Article 7(5).

⁵³⁶ Daniel Pavin, ‘Good News for European Bioinformatics – The Database Right’ (2001) *IP and Tech-Transfer* <<http://www.samedanltd.com/members/archives/EBR/Summer2002/DanielPavin.htm>> at 30 January 2007.

⁵³⁷ Mahesh Madhavan, ‘Copyright versus Database Right of Protection in the UK: The Bioinformatics Bone of Contention’ (2006) 9(1) *Journal of World IP* 81; For information on the operation of Creative Commons licences in relation to the European Database right, see also <<http://sciencecommons.org/resources/faq/databases>>



Need to decide:

- What rights are being licensed;
- Who the rights are being licensed to;
- Where the licence applies;
- The time period for which the licence operates;
- Whether the licence is exclusive or non-exclusive; and
- Any additional limits to be imposed e.g. non-commercial use only.

KEY POINTS

In determining who may access and use research data deposited in a database, it is necessary to ascertain whether copyright applies to the data and/or database and, if so, who owns the copyright. The copyright owner can determine whether the data or database may be accessed and used, and the terms of any permitted access or use.

Copyright may protect:

- *data, where the data takes the form of an original literary or artistic work, or a sound or film recording;*
- *datasets and databases (compilations), as original literary works; and*
- *computer software which enables the database to be searched and data to be retrieved, as an original literary work.*

There may be separate copyright interests in each of these components. However, data that is not protected by copyright will not attain copyright protection simply by being deposited into copyright-protected database.

A work must be original to attract copyright protection. In Australia, copyright may subsist in a compilation (e.g. a database) where substantial labour or expense has been invested in generating, collecting, compiling or organising the material: Desktop Marketing System Pty Ltd v Telstra Corporation Limited (Full Federal Court, 2002).

The owner of copyright in a work or material will usually be the author or creator of that material. It is possible to have more than one copyright owner, particularly where more than one person has participated in the creation of the material. If the material has been created by an employee in the course of their employment, the copyright owner is likely to be the employer or employing institution. An author or creator may also assign copyright to someone else, either entirely or partially, so that the assignee becomes the copyright owner to some extent. Issues may arise in determining ownership of purely computer-generated material.

Copyright confers on the owner the exclusive rights to reproduce, first publish, perform, communicate and adapt the copyright material. Any person who is not the copyright owner and who exercises one of these rights without the permission of the copyright owner engages in an infringing act. A copyright owner may commence legal proceedings against a person for copyright infringement. However, there are some exceptions to infringement that may apply, including fair dealing for research or study.

A copyright owner may use digital rights management (DRM) or electronic rights management information (ERMI), such as encryption or digital watermarking, to protect their digital material. It is unlawful to circumvent, remove or alter DRM or ERMI. It is also unlawful to deal with material where DRM or ERMI has been

removed or altered if the person dealing with the material knows the DRM or ERMI have been altered or removed, and knows or ought to know that this would result in a copyright infringement.

A copyright owner can issue a licence to a person or persons to deal with the copyright material in ways that would otherwise infringe copyright. For example, a licence may allow end-users to access and use data contained in a database for non-commercial purposes. Any person seeking to use the data for commercial purposes would need to obtain further permission from the copyright owner. There are many open content licences available to copyright owners, such as the Creative Commons licences. Additionally, the law may impose statutory licences which permit copyright materials to be used without first obtaining permission (but may require the payment of a royalty or remuneration to the copyright owner), such as licences that allow copying by Commonwealth and State governments, educational institutions and institutions assisting persons with an intellectual or print disability.

It is also important to remember that copyright law in relation to databases may operate differently in other jurisdictions. For example, in the United States the “originality” threshold is higher than in Australia. In the United States, there must be an exercise of creativity in the selection, coordination or arrangement of elements in a compilation before it will be recognised as original (the investment of labour or expense is not enough in itself). In the United Kingdom and Europe, the European Union Database Directive has introduced a sui generis database right that operates separately from and in addition to any copyright that may arise in a database.

CHAPTER 6 - CONFIDENTIAL INFORMATION

*“Various kinds of information which do not fall within the scope of other IP systems may nevertheless be very valuable. The person who possesses such information will want to be able to maintain its secrecy and protect it against public disclosure in order to maintain an edge over competitors.”*⁵³⁸

⁵³⁸ Anne Fitzgerald, *Intellectual Property* (2nd ed, 2002).

CONFIDENTIAL INFORMATION

Aims:

1. Provide an overview of how the law of confidentiality applies to data and data collections;
2. Explain how confidentiality protection is lost when data enters the public domain and how this can be avoided or delayed; and
3. Consider how contract can be used to control access to databases even when data is not confidential.

1. Introduction

6.01 As distinct from original creative works, pure information or raw data will not be protected by copyright.⁵³⁹ Courts have traditionally refused to protect information under copyright law “for fear of restricting innovation, competition and access”.⁵⁴⁰ However, a researcher who has expended considerable time, effort and funds in generating or compiling data, may want to protect this information from others who have not contributed to its production. The action for breach of confidence can be relied upon to protect and control access to such data, provided it is confidential and has not been released to the public.

2. Secret or Confidential Information

6.02 A breach of confidence action will only protect information which has:

The necessary quality of confidence about it, namely, it must not be something which is public property or public knowledge.⁵⁴¹

6.03 Data will only be protected as secret or confidential if it is *not* in the public domain. Information will be in the public domain only if enough people know about it so that it cannot rightly be regarded as secret.

6.04 Information can still be secret even if more than one person knows of it.

The test of secrecy is not absolute and the fact that some other people know that information does not mean that it is not relatively secret enough to be protected.⁵⁴²

⁵³⁹ See Chapter Five.

⁵⁴⁰ See Michael Pendleton, ‘Challenging Law’s Traditional Refusal to Protect Information per se for Fear of Stifling Innovation, Competition and Access’ (2002) 51 *IP Forum* 32.

⁵⁴¹ *Saltman Engineering Co Ltd v Campbell Engineering Co Ltd* (1948) 65 RPC 203, 215; *O’Brien v Komesaroff* (1982) 150 CLR 310, 326-328; *Johns v Australian Securities Commission* (1993) 178 CLR 408 at 432, 438, 461, 475.

⁵⁴² *Franchi v Franchi* [1967] RPC 149.

6.05 Whether information or data is sufficiently secret to be protected will depend on a range of factors, including how many people know about it and what measures the researcher has taken to ensure the data remains secret.

3. The Action for Breach of Confidence

6.06 The basis of the action for breach of confidence rests upon:

[t]he broad principle of equity, that he who receives information in confidence shall not take unfair advantage of it. He must not make use of it to the prejudice of him who gave it, without obtaining his consent.⁵⁴³

6.07 The elements that must be established to succeed in an action for breach of confidence are:

- the information itself must be of a confidential nature;
- the information must have been imparted in circumstances importing an obligation of confidence; and
- there must be an unauthorised use of that information to the detriment of the person claiming the right to maintain her or his confidentiality.⁵⁴⁴

Information Must be of a Confidential Nature

6.08 To be protected as confidential, information does not need to be “commercially valuable” but “the preservation of its confidentiality or secrecy [must be] of substantial concern to the plaintiff”.⁵⁴⁵

6.09 Factors that determine whether information is confidential in the context of employment include that:

- skill and effort was expended to acquire the information;
- information is jealously guarded by the employer, is not readily made available to employees and could not, without considerable effort or risk, be acquired by others;
- it was plainly made known to the employee that the material was regarded by the employer as confidential;
- the usages and practices of the industry support the assertion of confidentiality; and
- the employee in question has been permitted to share the information only by reason of his or her seniority or high responsibility within the employer’s organisation.⁵⁴⁶

6.10 Similar considerations will be relevant to the question of whether data is secret. For example, if a researcher has expended skill and effort in generating or compiling the data and has only allowed a select number of colleagues to access the

⁵⁴³ *Seager v Copydex Ltd* [1967] 2 All ER 415, 417.

⁵⁴⁴ *Coco v A N Clark (Engineers) Ltd* [1969] RPC 41, 47.

⁵⁴⁵ *Moorgate Tobacco Co Ltd v Philip Morris Ltd (No 2)* (1984) 156 CLR 414, 438.

⁵⁴⁶ *Wright v Gasweld Pty Ltd* (1991) 22 NSWLR 317 per Kirby P at 334.

data (after giving them notice that the data is confidential and is to be treated as such), it is likely that the data will be regarded as confidential.

Obligation of Confidence

6.11 The obligation to refrain from disclosing confidential information arises:

[f]rom the circumstances in or through which the information was communicated or obtained.⁵⁴⁷

6.12 There is no need for the obligation of confidentiality to be explicitly or verbally communicated to the other person. It can be determined from all the circumstances, including the relationship between the parties and their dealings with each other. The question is whether a reasonable person would have realised that the information was being given to them in confidence.⁵⁴⁸ If so, then the obligation not to disclose that information will be automatically imposed.

Where information of commercial or industrial value is given on a business like basis and with some avowed common object in mind, such as a joint venture or the manufacture of articles by one party for the other, [it will be difficult for] the recipient [of the information] to repel a contention that he was bound by an obligation of confidence.⁵⁴⁹

6.13 The obligation of confidence will also arise where information has been obtained improperly or dishonestly.⁵⁵⁰ This is because a person who has obtained a researcher's secret data or information by eavesdropping or other improper means should not be able to defend herself by arguing that the obligation of confidence arises only when there is a communication of information from the researcher himself.⁵⁵¹

Unauthorised Use or Disclosure

6.14 The duty of confidence is breached by the unauthorised use or disclosure of all or part of the secret information, even if that use or disclosure is unintentional or even unconscious.⁵⁵² It is uncertain whether the disclosure must cause actual detriment to the person who originally communicated the information in order for an action in breach of confidence to succeed. Where the information disclosed is of a commercial nature, detriment will usually be readily established.

6.15 Third parties can also be held liable for breach of the duty of confidence.⁵⁵³ A third party who receives confidential information innocently and is subsequently notified that the information was originally given in confidence can be restrained from using it.⁵⁵⁴ Failure to refrain from using information after notification may render the third party liable for damages or an account of profits.

⁵⁴⁷ *Moorgate Tobacco Co Ltd v Philip Morris Ltd (No 2)* (1984) 156 CLR 414, 438.

⁵⁴⁸ See *Coco v A N Clark (Engineers) Ltd* [1969] RPC 41, 47.

⁵⁴⁹ *Coco v A N Clark (Engineers) Ltd* [1969] RPC 41, 48.

⁵⁵⁰ See *Ashburton v Pape* [1913] 2 Ch 469; *Franklin v Giddens* [1978] Qd R 72.

⁵⁵¹ See *Franklin v Giddens* [1978] Qd R 72, 79-80.

⁵⁵² See *Seager v Copydex Ltd* (1967) 1 WLR 923.

⁵⁵³ *Stevenson, Jordan & Harrison Ltd v MacDonald* (1952) 69 RPC 10.

⁵⁵⁴ *Fraser v Evans* [1969] 1 All ER 8; *Wheatley v Bell* [1982] 2 NSWLR 544.

6.16 A person who has obtained information in breach of confidence may seek to gain an unfair advantage over the person who originally communicated the information. For example, the recipient of the information may launch a product that would have taken far longer to develop if they had only used information that was legally accessible to them at the time. The so-called “springboard doctrine” prevents a person from doing this. A person who has obtained confidential information in breach of confidence cannot use it as a springboard for activities detrimental to the person who made the confidential communication.⁵⁵⁵ The courts will handicap the person in breach to the extent to which that person has benefited from the breach of confidence.

6.17 In the recent UK House of Lords decision in *Douglas & Ors v Hello! Ltd & Ors* [2007] UKHL 21 the court held (3:2) that a duty of confidentiality will arise, even though the confidential information has been made public at the same time as the unauthorised disclosure, where there is a commercial purpose or contract in existence (prior to the unauthorised disclosure) between third parties for the purpose of keeping that information confidential.⁵⁵⁶

Defences

6.18 Where the person disclosing the information has “just cause or excuse” for doing so, the duty of confidence is not breached.⁵⁵⁷ Just cause or excuse will only really arise where the information is of such a nature that it is in the public interest for it to be disclosed.⁵⁵⁸ For the defence to be successful, the public interest in disclosure on the one hand must outweigh the competing public interest in preserving confidentiality on the other.

6.19 A person will also not be in breach where the confidential information is disclosed under legal compulsion, such as under a court order or in accordance with a requirement imposed by legislation.

Remedies

6.20 A person who is threatening to disclose or use confidential information may be restrained from doing so by a court order (an injunction). If the information has already been disclosed or used to the detriment of the person seeking to maintain its confidentiality, the most appropriate remedy will be monetary compensation. The compensation can be either in the form of:

- damages, which aims to place the wronged person in the financial position that they would have been in if the breach of confidence had not occurred; or
- an account of profits, which requires the person who has used the information to account for any profit that they have made from using the information.

⁵⁵⁵ *Terrapin Ltd v Builder's Supply Co (Hayes) Ltd* [1960] RPC 128.

⁵⁵⁶ *Douglas & Ors v Hello! Ltd & Ors* [2007] UKHL 21.

⁵⁵⁷ *Fraser v Evans* [1969] 1 QB 349.

⁵⁵⁸ *A v Hayden (No 2)* (1984) 156 CLR 532.

4. Information in the Employment Context

6.21 Much of the litigation involving breach of confidence arises from the alleged misuse by employees or ex-employees of information belonging to their employer which the employee has obtained or developed during the course of their employment. The situation is complicated by the fact that in addition to the duty of confidence, the employee owes his or her employer an extensive duty of good faith and fidelity during the course of employment. To understand the extent of an employee's obligations of confidentiality, it is also necessary to consider any express terms in their employment contract or any terms which may be implied into the employment contract.

6.22 The kind of information which an employee may obtain in the course of employment falls into three main categories:

(1) information which is so easily accessible by the public, so trivial or so unimportant that it cannot reasonably be regarded as confidential (for example, a published patent specification);

(2) information which the employee must treat as confidential during the course of his or her employment (either because the employee has expressly been told it is confidential or because it is obviously confidential in character) but which, once learned, remains in his or her memory and becomes part of his skill and knowledge or "know how"; and

(3) specific trade secrets such as "secret processes of manufacture, such as chemical formulae ... or designs or special methods of construction" and other highly confidential information.⁵⁵⁹

6.23 Information in the first category may be freely imparted by an employee both during and after the course of employment.

6.24 The employee's duty to act with good faith towards the employer means that during the course of employment, information in the second category must be treated as confidential. Unauthorised disclosure of such information, as long as the employment continues, will amount to a breach of the employee's duty of confidence and good faith.⁵⁶⁰ Where confidential information has been entrusted to an employee for a limited purpose, it may be used for that purpose only and no other.⁵⁶¹

6.25 The obligation of confidentiality after the termination of employment is more restricted than the general duty of good faith which binds the employee during the course of employment.⁵⁶² "Ordinary" confidential information, that is, information in the second category, may be freely used after the employment relationship terminates.

⁵⁵⁹ *Faccenda Chicken Ltd v Fowler* (1985) 6 IPR 155, 162.

⁵⁶⁰ *Wessex Dairies Ltd v Smith* [1935] 2 KB 80.

⁵⁶¹ *Smith Kline & French Laboratories (Aust) Ltd v Secretary, Department of Community Services & Health* (1990) 22 FCR 73.

⁵⁶² *Faccenda Chicken v Fowler* (1985) 6 IPR 155, 164.

An employee's general duty of loyalty (good faith and fidelity) ceases upon the termination of his or her employment and is replaced by an implied term relating to non-disclosure of especially confidential information.⁵⁶³

6.26 Where the information may properly be regarded as part of the ex-employee's own skill and knowledge "to do as he likes with" and not "a separate part of the employee's stock of knowledge which a man of ordinary honesty and intelligence would recognise to be the property of his old employer" (such as a chemical formula or a list of customers which the employee has memorised), it will not be protected by the doctrine of breach of confidence in a post-employment situation.⁵⁶⁴ An employer wanting to protect information of this kind can do so only by entering into a restrictive covenant with the employee, stipulating that the information will be protected after the termination of employment for a reasonable period of time and within a reasonable geographical territory.⁵⁶⁵

6.27 For information in the third category, the obligation of confidentiality continues to apply after the employment relationship terminates. Factors which will be taken into account in determining whether information falls into the third category include:

- whether the employee was habitually exposed to confidential information in his or her employment;
- whether the information can properly be classed as a trade secret or highly confidential;
- whether the employer impressed on the employee the confidentiality of the information; and
- whether the information can be easily isolated from other information which the employee is free to use or disclose.⁵⁶⁶

6.28 In the absence of an express clause in an employment contract, the traditional view has been that an obligation not to disclose secret processes may be implied into an employment contract.⁵⁶⁷

5. Government Information

6.29 The legal position for secrets held by government differs from the protection afforded to personal and commercial secrets.⁵⁶⁸ Information held by a government

⁵⁶³ *Wimmera Industrial Minerals Pty Ltd v Iluka Midwest Ltd* [2002] FCA 653 (Unreported, Sundberg J, 24 May 2002) [50].

⁵⁶⁴ *Printers & Finishers Ltd v Holloway* [1965] 1 WLR 1; *ANI Corp Ltd v Celtite Australia Pty Ltd* (1990) 19 IPR 506.

⁵⁶⁵ *Faccenda Chicken v Fowler* (1985) 6 IPR 155, 162.

⁵⁶⁶ *Faccenda Chicken v Fowler* (1985) 6 IPR 155, 165-167.

⁵⁶⁷ *Faccenda Chicken v Fowler* (1985) 6 IPR 155, 164; *Burger King Corp v Hungry Jack's Pty Ltd* [2001] NSWCA 187 (Unreported, Sheller, Beazley and Stein JA, 21 June 2001) [165].

⁵⁶⁸ *Minister for Mineral Resources v Newcastle Newspapers Pty Ltd* (1997) 40 IPR 403, 406.

that relates to the workings of government or a matter of public policy is more difficult to keep secret.⁵⁶⁹

6.30 For the government to succeed in a breach of confidence action it must show that disclosure of the information is likely to be detrimental to the public interest. The principles to be applied were set out by Mason J in *Commonwealth v John Fairfax & Sons Ltd* (1980) 147 CLR 39 at 52:

The court will determine the Government's claim to confidentiality by reference to the public interest. Unless disclosure is likely to injure the public interest, it will not be protected. The court will not prevent the publication of information which merely throws light on the past workings of Government, even if it be not public property, so long as it does not prejudice the community in other respects. Then disclosure will itself serve the public interest in keeping the community informed and in promoting discussion of public affairs. If, however, it appears that disclosure will be inimical to the public interest because national security, relations with foreign countries or the ordinary business of government will be prejudiced, disclosure will be restrained. There will be cases in which the conflicting considerations will be finely balanced, where it is difficult to decide whether the public's interest in knowing and in expressing its opinion outweighs the need to protect confidentiality.

6. Information May Lose Its Secrecy

6.31 It is possible for information or data to lose its quality of secrecy and lose any protection accorded by the doctrine of confidentiality. Information or data will no longer be secret if it becomes known to a sufficient number of people such that it passes into the public domain.⁵⁷⁰ Confidentiality will also be lost if the information is independently discovered by someone else whether through observation or by reverse engineering.

6.32 This does not mean that the information has lost any possible protection altogether. Where a researcher controls access to the information in the sense that they have the power to decide who will be allowed to see the information, then the information can be effectively protected through contract. For example, if data is placed in a closed or restricted database that is controlled by the researcher, the researcher may contract with other persons to allow them access to the database provided that those persons only use the data for a particular purpose or on other restrictive conditions.

7. Contracts

6.33 Confidential information and collections of non-confidential information can be protected and controlled through the legal mechanism of contract. Generally, contracts will provide for the disclosure of, or access to, confidential information on condition that the contracting party does not further disclose the information and does not use the information except for the purpose/s set out in the contract. Contracts relating to confidential information are discussed in more detail in Chapter 7.

⁵⁶⁹ *Attorney General (UK) v Heinemann Publishers Australia Pty Ltd* (1988) 10 IPR 153, cf *Attorney General (UK) v Newspaper Publishing Plc* [1988] 1 Ch 333.

⁵⁷⁰ *Attorney General (UK) v Heinemann Publishers Australia Pty Ltd* (1988) 10 IPR 153.

KEY POINTS

Some data generated or compiled in research projects may be confidential or secret. Information is confidential or secret where it is treated as such and has not been released into the public domain. In these circumstances, the action for breach of confidence may be relied upon to maintain the confidentiality of the information and prevent its unauthorised disclosure. Information subject to confidentiality restrictions may be information acquired during the course of employment, information that constitutes a trade secret (such as know-how relating to an advantageous or innovative process used in manufacturing) or which is of value to the holder while it remains confidential. Conversely, information held by the government will very rarely be protected by the doctrine of confidentiality because of the government's duty to be open and accountable to the public.

A researcher who possesses confidential information or data may permit another person to obtain access to the data and view it, on condition that the person agrees to comply with certain contractual terms. Usually, a contract will allow access to data provided that the person accessing the data agrees to maintain the secrecy of the information and to use it only for the specific purpose/s identified in the contract. A contract can also be used to control access to and use of data and information even where that data and information has entered into the public domain and has lost any quality of secrecy.

CHAPTER 7 - CONTRACT

“The stronger the underlying intellectual property right, the more necessary it becomes to devise suitable contractual templates regulating relations ...with a view to ensuring the smooth operation of a contractually reconstructed research commons.”⁵⁷¹

⁵⁷¹ J Reichmann and P Uhlir, ‘A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment’ (2003) 66 *Law and Contemporary Problems* 315, 437
<<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

CONTRACT

Aims:

1. Examine the various contractual mechanisms that can be used to control access to and use of data, including:
 - a. confidentiality agreements;
 - b. access agreements; and
 - c. contractual copyright licences;
2. Explain the importance of contract in protecting, controlling and managing data, the different rights and obligations of parties under contract and the interplay between contract and other areas of the law such as copyright; and
3. Provide diagrammatic examples of the different contractual models that can operate in relation to databases, whether simple or complex, to demonstrate the broad range of considerations to be taken into account when drafting contractual arrangements for data access and reuse. The diagrams provide practical examples illustrating the legal frameworks developed by some of the databases described and analysed in Chapter 4.

1. Introduction

7.01 In practice, the most important legal mechanism used to allocate rights between two or more parties is the contractual agreement. This is very much the case where rights to hold, access, use and reuse research data and information are concerned. The owner of rights in data or information may, by contract, grant others permission to use it in specified ways or, in other words, licence the recipient to use it, subject to certain conditions or restrictions.

7.02 In the case of copyright-protected data and information, the copyright owner may contractually extend the rights they automatically enjoy by virtue of the existence of copyright, and impose additional obligations on the party they permit to use the copyright material. For example, the copyright owner may grant to the recipient of the copyright material the rights to reproduce it and make it available on a web site from which it can be accessed and downloaded by other researchers. Additionally, the copyright owner may also require the recipient to undertake not to hold the copyright owner liable for consequences resulting from any inaccuracies that may be contained in the data supplied.

7.03 Alternatively, a copyright owner may enter into a contract with another party (a research institute) under which they licence to the recipient the rights to make reproductions of the copyright material and to make it available on an internal computer network which can be accessed only by researchers employed in a particular research group. The contract may also impose restrictions on the use of the

information itself, for example that it is to be kept confidential. For certain kinds of information, or where there is perceived to be potential for commercial outcomes flowing from the data and information, obligations may be included in the contract in relation to accounting for commercial benefits that can be attributed to the data or information made available to the recipient.

2. On-Line Contracts (“Click Wrap Agreements”)

7.04 Where access to research data is being provided on the Internet, it is most likely that access rights will be granted to users under click wrap agreements entered into online. This is particularly useful for enabling access to research data where there are numerous individual researchers situated in different geographical locations and access can be granted under standardised licensing terms. Valid contracts can be formed online when they are constructed in a “click wrap” form. A click wrap agreement involves end users first viewing the terms and conditions governing access to and use of the data and information in question, and then clicking an “I accept” or “I agree” button or icon to indicate that they assent to those conditions before they are able to obtain access to the data.

3. Contracts to Protect and Control Use of Information

7.05 Contracts are used to protect and control the use of information in two main contexts:

- firstly, where information is confidential, it may be disclosed to another party under the terms of a contract which require the recipient to maintain the information’s confidentiality; and
- secondly, where information is not confidential (for example where formerly confidential information has passed into the public domain), contracts may be used to control access to and use of a dataset or database, notwithstanding that protection is not based on secrecy of the information.⁵⁷²

Confidentiality Agreements

7.06 A common way of guarding against the dissemination of confidential information - whether commercial or private - is through contractual undertakings between the party disclosing the information and the recipient of the information. Parties may enter into a confidentiality agreement, also referred to as a “non-disclosure agreement”, to treat information about an invention, for example, as “private and confidential.” Such a contractual restraint upon disclosure is not limited to information which, at the time of the alleged breach of the covenant, retains the confidential character it had when first disclosed. However, where the contractual obligation is expressed as being “forever”, it will be rendered invalid by the application of the doctrine of restraint of trade.⁵⁷³

⁵⁷² For more on contracts, see Chapter 7.

⁵⁷³ *Maggbury Pty Ltd v Hafele Australia Pty Ltd* [2001] HCA 70 (Unreported, Gleeson CJ, Gummow, Kirby, Hayne and Callinan JJ, 13 December 2001).

7.07 The contractual and equitable bases for maintaining confidentiality are not mutually exclusive and may be relied upon concurrently to protect information against unauthorised disclosure. An agreement may contain an express declaration that the contractual obligations are intended to be supplementary to, and do not displace, the protection available under the equitable action for breach of confidence.

7.08 Confidentiality agreements and confidentiality clauses in agreements typically:

- identify the owner of rights in relation to the confidential information, who is disclosing it under the agreement (the “provider”);
- identify the information that is to be treated as confidential (often it will be listed in a schedule to a confidentiality agreement);
- impose specific obligations on the person to whom the information is disclosed (the “recipient”) to maintain the secrecy of the information, and set out any exceptions to those obligations;
- define the scope of the permitted use of the information; and
- provide for the consequences of a failure to comply with the confidentiality obligations.

7.09 For example, a confidentiality agreement between the provider and recipient of confidential data might state that:

- (a) the data is only to be used for the purposes of a particular research project;
- (b) the data is to be kept secret by the recipient who may only disclose it to persons directly working as researchers or assistants on a particular research project;
- (c) the recipient is to impose confidentiality obligations on all researchers and assistants to whom it provides the information for the purposes of conducting the research project, in the same terms as the confidentiality obligations imposed on the recipient by the provider;
- (d) the recipient is required to adopt appropriate mechanisms (for example, password codes or other means of restricting access and verifying the identity of persons accessing the data) in order to minimise the likelihood that researchers and assistants will disclose the data; and
- (e) that failure to comply with the obligations imposed under the agreement will attract specified penalties (such as payment of damages, removal of all copies of the provider’s data from the recipient’s computers and return of all copies of the database held by the recipient).

7.10 Where the recipient of confidential data is permitted to disclose the data to the recipient’s employees, assistants or researchers, the recipient will typically be

responsible under the terms of the confidentiality agreement for ensuring that its employees, assistants or researchers comply with the confidentiality obligations imposed by the provider and may be liable for any unauthorised disclosures made by these persons. In circumstances where the recipient of the confidential data is a research institute or organisation, it is common practice for confidentiality obligations to extend to the recipient's employees, assistants or fellow researchers requiring them to maintain the confidentiality of the data, in accordance with the obligations imposed on the recipient by the provider. Confidentiality obligations may be expressly stated in an employment or consultancy contract. The confidentiality obligations of employees are discussed at greater length in Chapter 6.

7.11 In the absence of an express confidentiality term, an obligation to keep information confidential may be implied if it is reasonable and necessary to give "business efficacy" to the arrangements.⁵⁷⁴ The courts will regard the implication of a term as necessary if, without it:

[t]he enjoyment of the rights conferred by the contract would or could be rendered nugatory, worthless, or perhaps be seriously undermined.⁵⁷⁵

Contracts Controlling Access To and Use of Information

7.12 Where a researcher or research organisation has control over who may access the data generated or compiled through their research (for example, they control the database in which the data is deposited), they may wish to retain that control by requiring persons interested in accessing the data to enter into an Access Agreement.

7.13 This Access Agreement may:

- identify the data to be accessed under the agreement;
- identify the person/s or class of persons who are permitted to have access to the data;
-
- state that access rights cannot be transferred onto third parties;
- limit the purposes for which the data may be used; and
- provide for the consequences of a failure to comply with the agreement.

7.14 For example, the contract may allow the data to be accessed by a relevant person for non-commercial purposes only, or may provide that if that person uses the data commercially they must account back to the researcher for a proportion of the profits.

⁵⁷⁴ *Byrne & Frew v Australian Airlines Ltd* (1995) 131 ALR 422; *Castlemaine Tooheys Ltd v Carlton & United Breweries Ltd* (1987) 10 NSWLR 468.

⁵⁷⁵ *Byrne & Frew v Australian Airlines Ltd* (1995) 131 ALR 422, 450.

7.15 Additionally, an Access Agreement may contain a disclaimer to the effect that the researcher takes no responsibility for any inaccuracies in the information to which they are providing access.

4. Contracts and Copyright

Assignment

7.16 A copyright owner of a work (such as a database) can choose to transfer all or some of their rights to another, through a contractual agreement called an assignment. Assignments must be in writing to be effective, and must be signed by or on behalf of the copyright owner. The usual terms of an assignment will be that the copyright owner assigns either all copyright or some rights (for example, the right to publish the work) to the contracting party, in return for some sort of consideration, most commonly money.

Licence

7.17 A copyright owner can give another person permission to deal with their work in a way that would usually infringe copyright (for example, reproduce the work) through a licence. Licences can be contractual or non-contractual. Contractual licences will contain terms relating to:

- what work the licence applies to;
- what acts the licence permits;
- any restrictions imposed upon the party acting under the licence;
- the consideration provided for the licence; and
- whether or not the licence is revocable.

7.18 For example, the owner of a database (A) may grant a licence to another party (B) to reproduce A's database, on the condition that B reproduces the database only for non-commercial purposes.

7.19 A licence only has to be in writing if it is an exclusive licence. An exclusive licence is one which grants the relevant rights to the licensee only (to the exclusion of the copyright owner) and to no one else. In contrast, a non-exclusive licence allows the copyright owner to licence the same right to multiple different persons.

5. Diagrammatic Examples

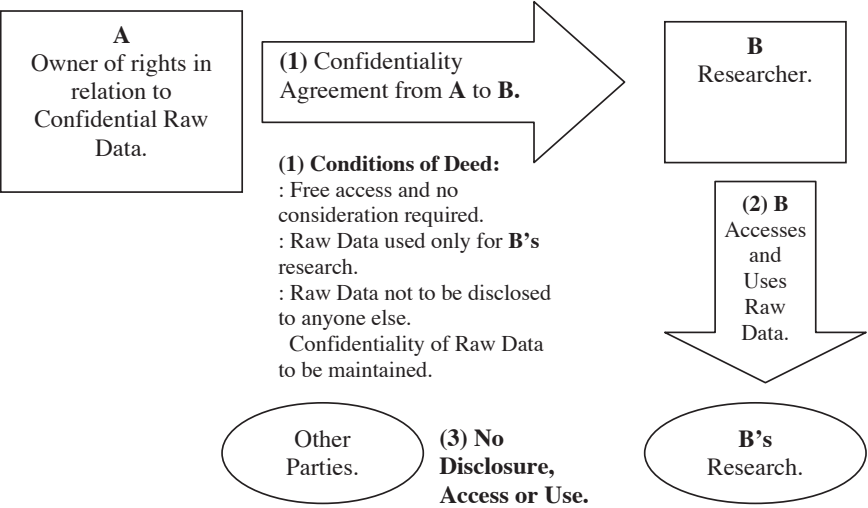
7.20 The following examples demonstrate situations in which contracts, contractual deeds and licenses are used to allocate rights to use research data and information.

Example 1: Basic Confidentiality Agreement for Use of Research Data

Parties and Roles:

- **A** is the owner of rights in relation to a confidential collection of raw data. The raw data is not protected by copyright, because it does not meet the originality threshold required for copyright protection to apply.
- **B** is a researcher who wishes to use **A**'s raw data in **B**'s research.
- **A** and **B** agree to let **B** have access to **A**'s data without charge, on the condition that confidentiality is maintained.

Basic Confidentiality Agreement for Use of Research Data Diagram:



Example 1 - Steps Taken:

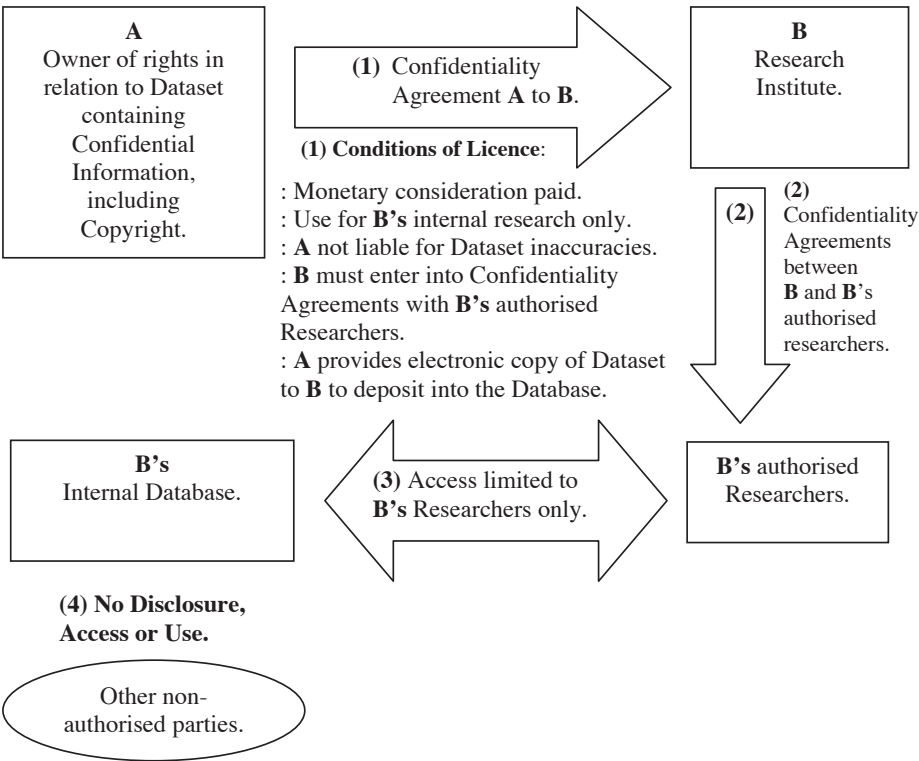
- **(1) A** grants **B** the right under a Confidentiality Agreement to use the raw data in **B**'s research. **A**'s access and use is free, the confidentiality agreement is made as an enforceable deed. The conditions are that **B** only uses the raw data for certain research purposes, that **B** does not disclose the raw data to anyone else and that **B** keeps the raw data confidential.
- **(2) B** accesses and uses **A**'s raw data in **B**'s research and ensures that there is no disclosure, access to or use of the data by other parties **(3)**.

Example 2: Internal Confidentiality Agreement

Parties and Roles:

- **A** is the owner of rights, including copyright, in relation to a dataset containing confidential data.
- **B** is a research institute and is the managing entity of an internal research database which is only accessible by authorised researchers within the **B** organisation and is not accessible by outside parties. **B**'s researchers are either employees or contractors of **B** and wish to access **A**'s dataset through **B**'s database to use it for **B**'s specific research projects.
- **A** agrees that **A**'s dataset will be made available without charge for access on **B**'s internal database, subject to obligations of confidentiality imposed by **B** on **B**'s authorised researchers.

Internal Confidentiality Agreement Diagram:



Example 2 - Steps Taken:

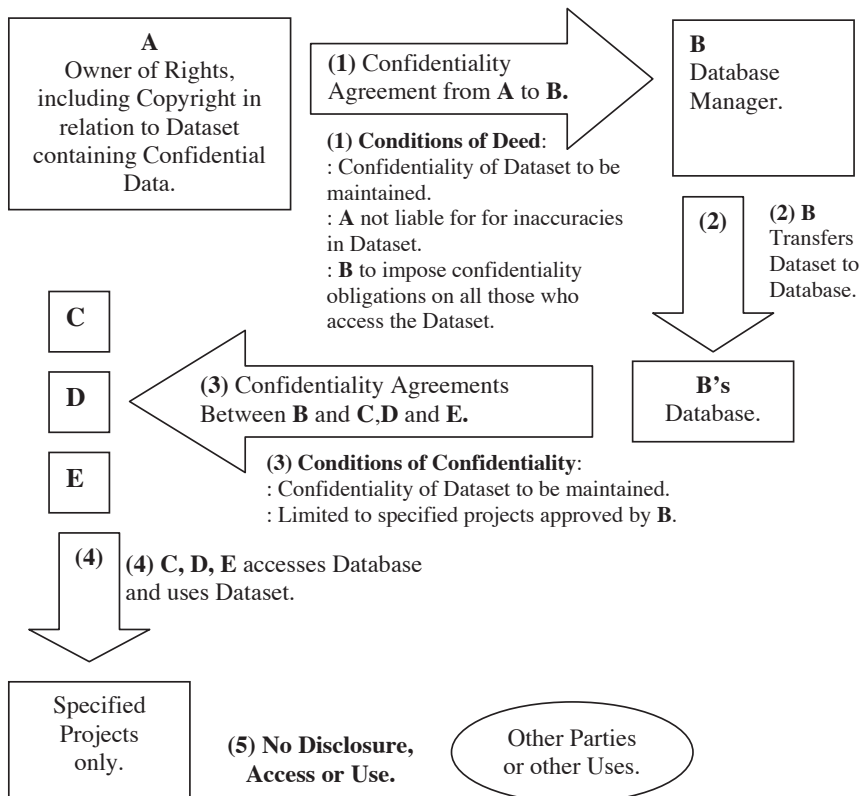
- **(1) A** grants **B** the right under a confidentiality agreement to make copies of the dataset for restricted distribution to **B**'s researchers only. As the right to reproduce the data is given for monetary consideration, the confidentiality agreement is in the form of a contractual licence.
- **A** imposes conditions on **B** that the dataset can only be used for internal research purposes, that **A** is not liable for any inaccuracies in the data provided and that **B** must enter into confidentiality agreements with its authorised researchers, so that the researchers are obliged to **B** not to disclose the dataset or technology derived from the dataset to other parties.
- **A** provides an electronic copy of the dataset to **B** and **B** transfers this copy into **B**'s internal database.
- **(2) B** enters into confidentiality agreements in the form of deeds with each of the researchers whom **B** has authorised to have access to the dataset. Other non-authorised employees, contractors and persons outside of the **B** organisation will not have access.
- **(3) B's** researchers access the dataset on the internal database operating within **B**'s institute and ensure that there is no disclosure, access to or use of the data by non-authorised parties **(4)**.

Example 3: Confidentiality Agreement – Conditional User Access

Parties and Roles:

- **A** is the owner of rights, including copyright, in relation to a dataset containing confidential data.
- **B** is the managing entity of a database which is accessible by registered users on a controlled access basis. **B** is obligated to maintain the confidentiality of data submitted to the database.
- **C, D** and **E** are members of the research community who wish to access **A**'s dataset through **B**'s database and to use it for each of their own research projects.
- **A** agrees that **A**'s dataset will be made available for free access on **B**'s database, subject to obligations of confidentiality imposed by **B** on users.

Confidentiality Agreement - Conditional User Access Diagram:



Example 3 - Steps Taken:

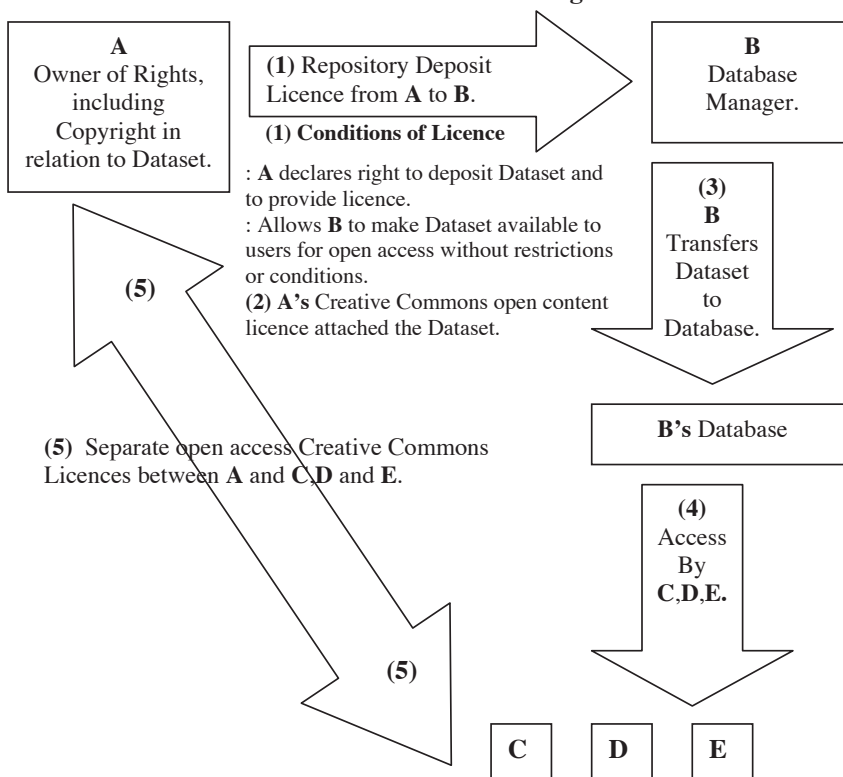
- **(1)** **A** enters into a confidentiality agreement with database manager **B** to reproduce the dataset in a database. As the reproduction is free, the confidentiality agreement is made as an enforceable contractual deed.
- Under that confidentiality agreement, **B's** right to reproduce data is conditional upon **B** maintaining the confidentiality of the dataset. **A** requires that **B** enters into confidentiality agreements with all persons who access the dataset from the database.
- **(2)** **B** transfers the dataset to the database.
- **(3)** **C**, **D**, and **E** access the dataset and automatically enter into confidentiality agreements with **B** and **B** limits their use of the dataset to specified projects approved by **B**.
- **(4)** **C**, **D**, and **E** use the dataset in their own individual specific projects and each of them ensure that there is no disclosure, access by or use of the data by other parties or for uses outside of their specified projects **(5)**.

Example 4: Access Content Licence - Owner and User

Parties and Roles:

- **A** is the owner of rights, including copyright, in relation to a dataset.
- **B** is the managing entity of an open access database which is accessible by users on a worldwide basis without charge. **B** is not obliged to maintain the confidentiality of submitted data or to enter into licences with users for their access and use of the data.
- **C, D** and **E** are members of the research community who wish to access **A**'s dataset from **B**'s database and to use it for each of their own research projects.
- **A** agrees that **A**'s dataset will be deposited into **B**'s database and reproduced on the database for access by any users. **A** does not impose any restrictions or conditions on access and use of the data, other than those contained in a Creative Commons open content licence attached to the dataset by **A** before providing it to **B**.

Open Access Content Licence – Owner and User Diagram:



Example 4 - Steps Taken:

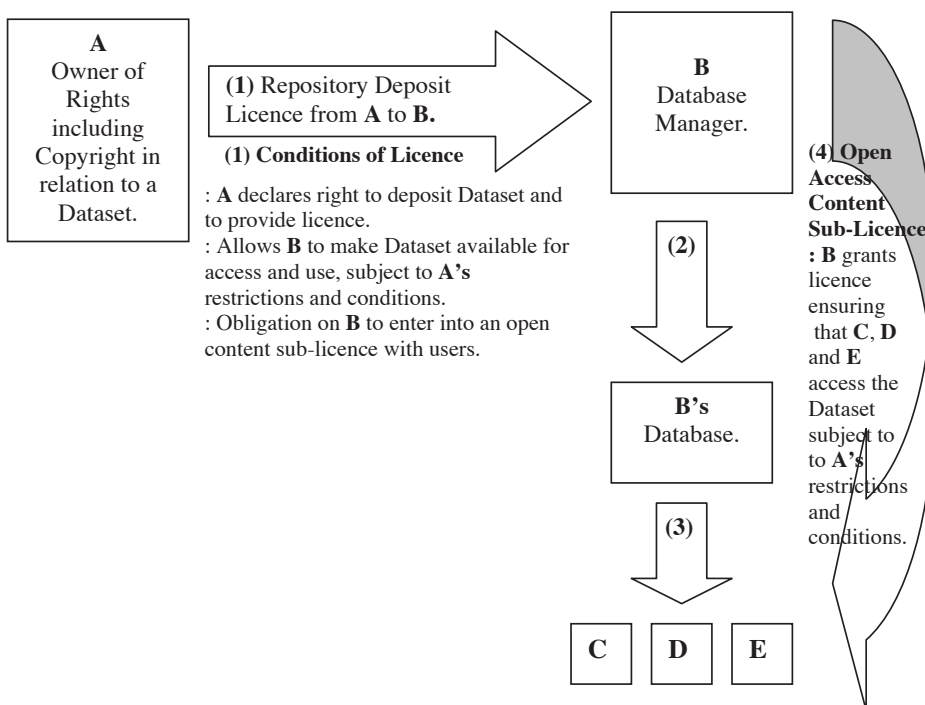
- **(1)** **A** deposits its dataset into **B's** database under a repository licence to database manager **B** without charge. This repository licence allows **B** to reproduce and communicate the dataset to the public through **B's** database for open access by users on a worldwide basis without restrictions or conditions. **A** does not impose any obligations on **B** to maintain confidentiality or to licence the data to users.
- **(2)** **A** has attached a Creative Commons licence to their dataset before the dataset is provided to **B's** database. This permission by **A** extends to all users who access the dataset on **B's** database and creates an open content licence between each user and **A** when **A's** dataset is accessed by users.
- **(3)** **B** transfers the dataset to **B's** database.
- **(4)** **C,D** and **E** access **A's** dataset in **B's** database.
- **(5)** **C,D** and **E** are automatically granted permission by **A** to access and use the data through the Creative Commons open content licence. Access and use is then regulated by the terms of that Creative Commons licence.

Example 5: Open Access Content Sub-Licence - Database and User

Parties and Roles:

- **A** is the owner of rights, including copyright in relation to a dataset.
- **B** is the managing entity of a database which is accessible by users on a worldwide basis without charge. **B** is also obligated to provide and to regulate permission to users who wish to access and use data through the database.
- **C, D** and **E** are members of the research community who wish to access **A**'s dataset through **B**'s database and to use it for each of their own research projects.
- **A** agrees that **A**'s dataset will be deposited into **B**'s database and reproduced on the database for access without charge. This is subject to certain restrictions and conditions imposed by **A** on access and use of the data, which will be enforced by **B** in open access content sub-licences between **B** and users.

Open Access Content Permission – Database and User Diagram:



Example 5 - Steps Taken:

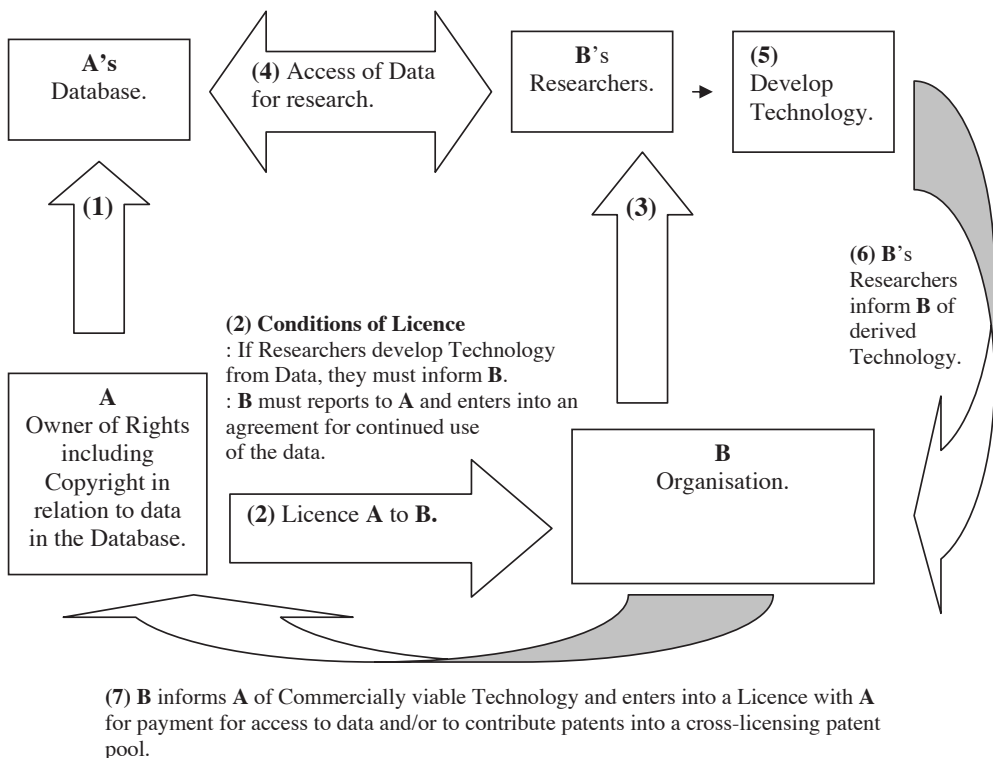
- **(1)** **A** deposits its dataset into **B**'s database under a repository licence to database manager **B** without charge. This repository licence allows **B** to reproduce and communicate the dataset to the public through **B**'s database for access by users on a worldwide basis.
- In the licence, **A** imposes restrictions and conditions on access and use of the data. **A** grants **B** the right to sub-licence the dataset to users, making those users subject to the restrictions and condition on access and use.
- **(2)** **B** transfers the dataset to the database and **B** attaches an access content sub- licence to the dataset. This permission by **B** extends to all users who access **A**'s dataset on **B**'s database. It creates an open access content sub-licence between each user and **B**.
- **(3)** **C,D** and **E** access **A**'s dataset through **B**'s database.
- **(4)** **C,D** and **E** are automatically granted permission by **B** to access and use **A**'s dataset through **B**'s open access content sub-licence. The conditions and restrictions in this sub-licence replicate the conditions and restrictions on access and use imposed by **A** under the licence between **A** and **B** **(1)**.

Example 6: Open Access Content Licence – Commercial Restrictions

Parties and Roles:

- **A** is the managing entity of a database which is accessible by users on a worldwide basis without charge for non-commercial purposes only. **A** is also the owner of rights, including copyright in relation to data contained in the database.
- **B** wants its researchers to evaluate the data in the database for commercial use. **B**'s researchers are either employees of **B** or are contracted by **B** to undertake specific research projects. If **B**'s researchers are able to develop commercial technology and/or a patentable invention from the data, then **B** will want to commercialise that technology.
- Where there is commercial use of the data **A** imposes further conditions on all users of **A**'s database. Users can access and use data for a commercial purpose, provided that such users immediately inform **A** regarding any discovery by them of commercially viable technology derived from such data and that they enter into an agreement with **A** for the commercial use of that data.

Open Access Content Licence – Commercial Restrictions Diagram:



Example 6 - Steps Taken:

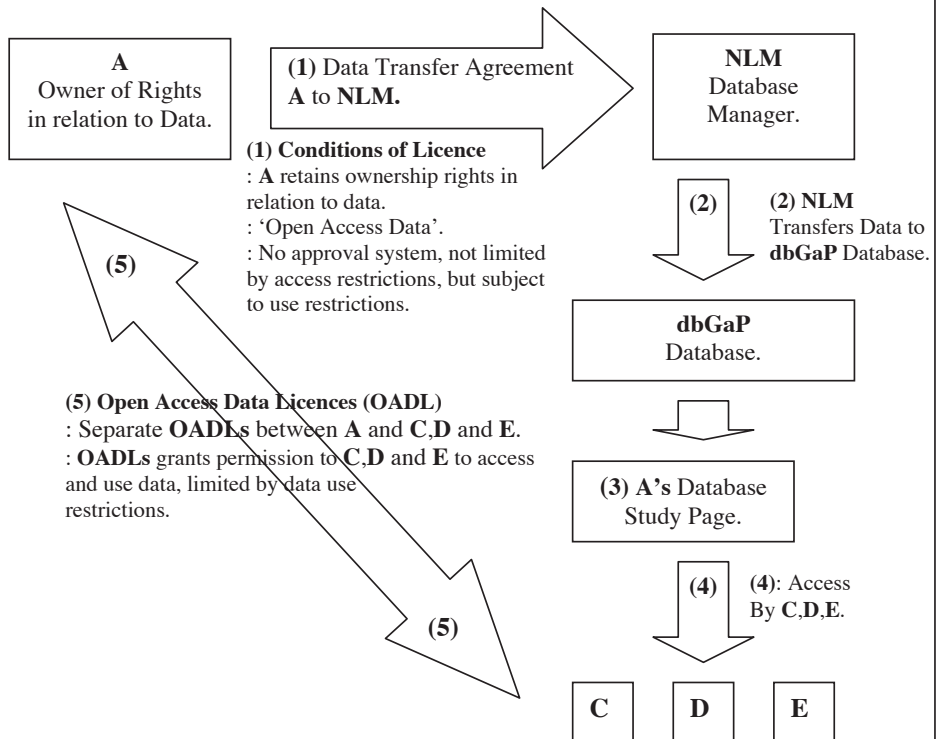
- (1) **A** has transferred data into the database and is the owner of rights, including copyright in relation to data contained in **A**'s database.
- (2) **A** attaches an Open Access Content Licence to the data (in relation to which it holds copyright). It is provided on condition that any further reproduction or communication to the public is for non-commercial purposes only.
- This Open Access Content Licence is subject to the condition that if the data is used by **B**'s researchers (or any other party) in developing commercially valuable technologies or patentable inventions, then **B** (or any other party) must enter into a further licence with **A**. Under that licence, **B** can make payments to **A** for continued use of the accessed data and/or will contribute any patents that **B** develops into an agreed 'patent pool'.
- (3) **B** instructs its researchers to access the database and that they must notify **B** if they derive commercially valuable technologies or patentable inventions from the accessed data.
- (4) **B**'s researchers access the database and begin research. (5) **B**'s researchers derive commercial technology and /or a patentable invention from the accessed data. (6) **B**'s researchers notify **B** of this derived technology.
- (7) **B** notifies **A** of the development of the commercial technology and/or a patentable invention. **B** then enters into a licence with **A** for payment for **B**'s continued access to and use of the data and/or **B** is obliged to licence **B**'s patents to members of a patent pool nominated by **A**.
- In a patent pool, **B** agrees to licence its data derived patent technology for little or no fee to **A** and/or other users of the database. In return, **A** and any other users agree to licence technology which they derived from that same data (if any) to **B**. The parties will need to obtain legal advice on the structuring of such arrangements to avoid any competition law limitations on patent pooling.

Example 7: dbGaP – Data Transfer Agreement - ‘Open Access to Data Licence’

Parties and Roles:

- **A** owns rights in relation to the database (presumably copyright exists).
- **NLM** is the managing entity of the **dbGaP** database which is accessible by users on a worldwide basis without charge. **NLM** categorises certain data as ‘open access data’ (**OAD**). **OAD** is not subject to any restrictions on access, but may be subject to restrictions on use of the data. **NLM** is not obliged to maintain the confidentiality of **OAD** or to licence users for its access and use.
- **C, D** and **E** are researchers who wish to access **A**’s data from the **dbGaP** database and to use it for each of their own research projects.
- **A** agrees that **A**’s data will be reproduced on the **dbGaP** database and categorised as publicly available **OAD**.

dbGaP Data Transfer Agreement – ‘Open Access to Data’ Diagram:



Example 7 - Steps Taken:

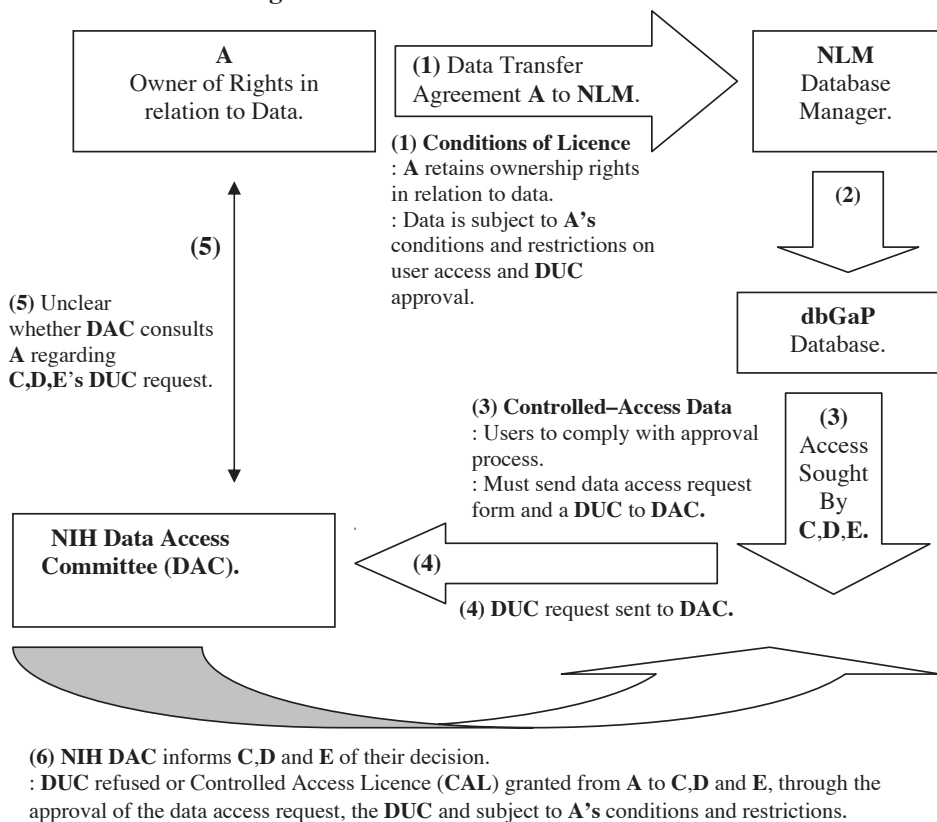
- **(1) A** enters into a ‘data transfer agreement’ with the **NLM** database manager. The agreement states that **A** retains ownership rights in relation to the data, that the data will be categorised as **OAD** and that it will not be subject to restrictions on access, but can only be used for research purposes, cannot be further distributed nor used to determine the identity of any test subject.
- **(2) NLM** transfers the data to the database.
- **(3) NLM** drafts the open-access data study page which is publicly available, categorising the data as **OAD**. There is no restriction on its access but the study page does set out “Use Restrictions” as listed above.
- **(4) C,D** and **E** access **A’s OAD** through **A’s** study page on the **dbGaP** database website.
- **(5)** By accessing the data, **C, D** and **E** are each automatically granted permission under an open access to data licence (**OADL**) given by **A** to access and use that **OAD**, subject to the use restrictions.

Example 8: dbGaP – Data Transfer Agreement – ‘Controlled Access to Data Licence’

Parties and Roles:

- A is the owner of rights in relation to data.
- **NLM** is the managing entity of the **dbGaP** database which is accessible by users on a worldwide basis. In relation to data categorised as controlled-access data, the owners of rights in relation to such data impose conditions and restrictions on its access and use and **NLM** is obliged to maintain the confidentiality of this submitted controlled access data.
- **C, D** and **E** are members of the research community who wish to access **A**'s controlled-access data from the **dbGaP** database and to use it for each of their own research projects.
- **A** agrees that **A**'s dataset will be reproduced on the **dbGaP** database and categorised as controlled access data. Potential users must submit a 'Data Use Certificate' (**DUC**) to the NIH Data Access Committee (**DAC**) for permission to access and use this data.

dbGaP Data Transfer Agreement – Controlled Access to Data:



Steps Taken:

- (1) **A** enters into a ‘data transfer agreement’ with the **NLM** database manager. The agreement states that **A** retains ownership rights in relation to the data, that the data will be categorised as controlled-access data, that access and use is subject to user’s adhering to **A**’s “Access Restrictions” and that access and use is subject to the approval of a data access request and a **DUC** submitted by potential users to the **DAC**.
- (2) **NLM** transfers data to the database.
- (3) **NLM** drafts and locates the controlled-access data study page under the “Authorized Access” webpage, categorising the data as controlled access data. This webpage then provides a link to the request procedures for potential users. These procedures set out the requirements that that potential users must comply with (see chapter 4), including the submission of a data access request form and a **DUC** to the **DAC** for approval before they can access and use this data.
- (4) **C,D** and **E** each want to access and use **A**’s controlled access data and each comply with those requirements and submit a data access request form and a **DUC** to the **DAC**.
- (5) The **DAC** receives the data access request forms and the **DUCs**, although it is unclear whether the **DAC** and consults with **A** regarding the requests.
- (6) The **DAC** informs **C,D** and **E** of the decision. The **DUC** is either refused or controlled access licences (**CAL**’s) are granted from **A** to each of **C,D** and **E** through the approval of the data access request and the **DUC** to access and use the controlled access data, in accordance with **A**’s “Access Restrictions”, the submitted **DUC** and/or with additional restraints and conditions that may be imposed.

KEY POINTS

In practice, contracts play an important role in the management of rights to access and reuse research data.

Rights to access and reuse research data may be granted under a variety of contractual arrangements including the following, which may be used individually or in combination:

- *confidentiality agreements, which impose obligations to maintain the secrecy of confidential information and prevent its dissemination to unauthorised parties;*
- *copyright assignments, which transfer ownership of copyright entirely or partially to another party;*
- *copyright licences, which authorise another party to exercise some or all of the rights included within the scope of the copyright owner's exclusive rights; and*
- *general contracts which provide for access to and use of a database and the data contained in it, subject to certain conditions determined by the owner of the data or the database (Access Agreements).*

CHAPTER 8 - DATA SHARING FRAMEWORKS

“The scientific process runs on data; without data, science cannot progress, science education stymies, and science cannot provide the underpinnings for the economic progress, health, and welfare of society. Despite this underlying importance...data are not always readily accessible. The dilemma is that...data, whether collected with public or private funds, are not always “open” (made available to the public)...”⁵⁷⁶

⁵⁷⁶ Geological Society of America (GSA) Position Statement, Adopted in May 2005, available at <http://www.geosociety.org/aboutus/pos7_data.pdf> at 24 May 2007

DEVELOPING DATA SHARING FRAMEWORKS – ESTABLISHING ACCESS AND USE PRINCIPLES

Aims:

1. Provide an overview of open access policies and principles relating to data at different levels of authority:
 - a. international open access statements;
 - b. statements of public sector research funding organisations in Australia, the United States and the United Kingdom; and
 - c. statements of private sector research organisations; and
2. Consider how these open access policies and principles can inform research practice and assist in the development of frameworks for management of research data and, specifically, the arrangements for access to and reuse of data.

1. Introduction

8.01 A significant shift has occurred during the last decade in attitudes about and practices relating to access to and reuse of research data. In Australia, there have been some important recent developments which are indicative of a shift in approach towards a greater degree of accessibility to the results of publicly funded research. In 2004, the Federal Government established the Accessibility Framework to manage research information, outputs and infrastructure in order to make them more discoverable, accessible and shareable.⁵⁷⁷ In 2006, ARC and NHMRC announced that they were adopting policies that require open access to be provided to data and publications resulting from funded research projects.

8.02 If the open access objectives described in these documents are to be given effect, it will be necessary to work on the development of:

- clearly enunciated statements of principles to give effect to open access policies; and
- practical guidelines to assist institutions and researchers in implementing open access policies and principles in specific research contexts.

8.03 Work is required to develop high-level principles relating to availability of, access to and use of data and information generated through publicly funded research activities. In developing data access policies, principles and guidelines for application in the Australian research and academic sector, it is necessary to understand

⁵⁷⁷

<http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/accessibility_framework/> at 16 January 2007.

developments that have occurred not only domestically but also in other key jurisdictions and international organisations.

8.04 This chapter provides an overview of open access policies and principles, as they relate to data generated through research. Summaries are provided of key international statements on open access, including the Bermuda Principles and the Berlin Declaration. The chapter also considers the open access policies and statements of public sector research funding organisations including the ARC and the NHMRC in Australia, the NIH in the United States and the various United Kingdom Research Councils. Open access policies and position statements developed by private sector research organisations, such as the Wellcome Trust in the UK, are also examined.

2. International Organisations

Bermuda Principles (1996)⁵⁷⁸

8.05 The Bermuda Principles were developed by scientists involved in the International Human Genome Sequencing Consortium and their funding agencies. Agreement on the Bermuda Principles was reached at International Strategy Meetings on Human Genome Sequencing held in Bermuda in 1996 and 1997. The principles represent an agreement among researchers about the need to establish a basis for the rapid and open sharing of pre-publication data on gene sequences.⁵⁷⁹

8.06 The Bermuda Principles require automatic release of sequence assemblies larger than 1kb (preferably within 24 hours) and immediate publication of finished annotated sequences. They aim to make the entire gene sequence freely available to the public for research and development in order to maximise benefits to society.

8.07 The introduction of requirements for rapid deposit of genome sequence information into the public international DNA sequence databases (GenBank, EMBL and DDBJ) was a strategic response to attempts by both public and private sector organisations to obtain patents for nucleotide sequences generated under the Human Genome Mapping project.⁵⁸⁰ Attempts at patenting were widely criticised and vehemently opposed by the genetic sequencing community. The response was:

[T]he publicly funded sequencing laboratories agreed on a collective patent-defeating strategy, requiring academic researchers to deposit their sequence data to publicly available repositories on the Internet within hours of the data generation. By making the data publicly accessible, the academic researchers intended not only to implement the norm of communality but to defeat the novelty of patent applications on the sequences. Agreement to these rules was largely the product of normative pressure by the research community in addition to some formal enforcement by scientific journals that required sequence disclosure as a stipulation for article publication.⁵⁸¹

⁵⁷⁸ Available at <<http://www.gene.ucl.ac.uk/hugo/bermuda.htm>>.

⁵⁷⁹ <http://en.wikipedia.org/wiki/Bermuda_Principles>.

⁵⁸⁰ See J Boyle, 'Enclosing the Genome: What squabbles over genetic patents could teach us, in F S Keiff (ed), *Perspectives on Properties of the Human Genome Project* (2003) 97.

⁵⁸¹ Dan L Burk, *Intellectual Property in the Context of e-Science*, Minnesota Legal Studies Research Paper No. 06-47, (2006) 7–8 <<http://ssrn.com/abstract=929479>>; see also article 13 (2007) 12(2) *Journal of Computer-Mediated Communication* <<http://jcmc.indiana.edu/vol12/issue2/burk.html>>

8.08 The significance of the shift in attitude and practice evidenced by the Bermuda Principles should not be underestimated. Typical practice at the time is demonstrated by the arrangement GlaxoSmithKline entered into in 1993 with Human Genome Sciences to develop drugs based on genome data, and the fact that in 1998, Bayer spent US\$465 million to obtain access to the genetic library being assembled by Millennium Pharmaceuticals. Neither collaboration led to a marketed drug.⁵⁸²

8.09 The Bermuda Principles were reaffirmed at a meeting of an international group of data producers, users, database personnel, journal editors and funding agency representatives convened by the Wellcome Trust at Fort Lauderdale in January 2003.⁵⁸³ The group stated that pre-publication release of sequence data by the International Genome Sequencing Consortium and other sequence producers had greatly benefited the scientific research community.⁵⁸⁴ It recommended that the Bermuda Principles should be extended beyond their initial application to sequence assemblies of a minimum size from BAC-based sequence projects so that they apply to rapid release of all sequence data, including both the raw traces submitted to the Trace Repositories at NCBI and Ensembl and whole genome assemblies, subject only to the data meeting appropriate quality assessment standards.

8.10 In 2003, the US National Human Genome Research Institute (NHGRI), the focal point of NIH's research into the genetics of human disease,⁵⁸⁵ reaffirmed and extended its:

Commitment to the Bermuda Principles for all types of large-scale DNA sequence data sets, including those that were not considered when the Bermuda Principles were originally devised.⁵⁸⁶

8.11 The NHGRI policy requires sequence assemblies larger than 2kb to be deposited into the public databases within 24 hours and raw shotgun sequences to be deposited within a week. The policy reminds users of the database that they:

Are expected to acknowledge the source of the sequence data through the use of appropriate citations" and urged to recognize that producers have a legitimate interest in publishing their own data.⁵⁸⁷

At the same time, depositors are reminded that:

Even if the sequence data are occasionally used in ways that violate normal standards of scientific etiquette, unconditional release is a necessary risk because the benefits of immediate release are "considerable".⁵⁸⁸

⁵⁸² Matthew Herper and Robert Langreth, 'Biology Goes Open Source', *Forbes.com* 12 February 2007, <http://www.forbes.com/2007/02/12/novartis-genes-diabetes-research-biz-cz_mh_0212novartis_print.html>.

⁵⁸³ The Wellcome Trust, 'Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility' (Report of a meeting organised by the Wellcome Trust, Fort Lauderdale USA, 14-15 January 2003) <<http://www.wellcome.ac.uk/assets/wtd003207.pdf>>.

⁵⁸⁴ The NHGRI policy arose from this meeting.

⁵⁸⁵ See <<http://www.genome.gov>>.

⁵⁸⁶ *Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects* (2003) <<http://www.genome.gov/10506537>>; see also Tabitha M Powledge, 'Revisiting Bermuda' (2003) *The Scientist*, <<http://www.the-scientists.com/news/20030311/03/>>.

⁵⁸⁷ Ibid.

⁵⁸⁸ Ibid.

Budapest Open Access Initiative (2002)

8.12 The Budapest Open Access Initiative (BOAI) arose out of a meeting convened in Budapest by the Open Society Institute in 2002.⁵⁸⁹ The purpose of the meeting was to accelerate progress in making research articles in all academic fields freely available on the internet. In contrast to the Bermuda Principles, the Budapest statement is specific to scholarly journal literature.

8.13 The BOAI has been significant in advancing the open access movement generally, articulating the importance of open access in:

Removing access barriers...[It] will accelerate research, enrich education, share learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.⁵⁹⁰

Bethesda Statement on Open Access Publishing (2003)

8.14 Following the Budapest Open Access Initiative in 2002, there was a strong surge in the open access movement worldwide. 2003 saw the development of more well-defined, thorough and progressive open access statements, adopted by funding bodies and academic institutions alike. Among these was the Bethesda Statement on Open Access Publishing, released on 20 June 2003.⁵⁹¹ Like the BOAI, the Bethesda Statement was strongly focused on published research literature as opposed to data. Yet the Bethesda Statement was broader in its terms, supporting the position that:

[A] complete version of the [published] work and *all supplemental materials* (our emphasis)...in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository ...⁵⁹²

Although “supplemental material” is not defined in the statement, it could conceivably encompass the data on which the published work is based, or at least the metadata supporting the deposited work.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003)

8.15 Of the early international statements on open access, the Berlin Declaration is perhaps the most significant because it is the statement most referred to and relied upon by institutions creating open access policies for their repositories.⁵⁹³ The goal of the Berlin Declaration was to support “the open access paradigm via the Internet” and to:

Promote the Internet as a fundamental instrument for a global scientific knowledge base.⁵⁹⁴

⁵⁸⁹ See <<http://www.soros.org/openaccess/>>.

⁵⁹⁰ <<http://www.soros.org/openaccess/read.shtml>>.

⁵⁹¹ <<http://www.earlham.edu/~7Epeters/fos/bethesda.htm>>

⁵⁹² <<http://www.earlham.edu/~peters/fos/bethesda.htm>>

⁵⁹³ For example, see Biomed Central, *Open Access Now*

<<http://www.biomedcentral.com/openaccess/www/?issue=10>> and the Italian Universities for Open Access: towards open access for scholarly literature, *Messina Declaration* <<http://www.aepic.it/conf/viewappendix.php?id=49&ap=1&cf=1>> among others.

8.16 The Berlin Declaration drew strongly on the Bethesda Statement, with the addition of an inclusive definition for open access contributions. The Berlin Declaration defined open access contributions to include scientific research results, raw data and metadata, thereby extending the scope of open access principles to cover more than just literature and publications. The move away from the restrictive policy of relating open access principles to only literature was an important one, particularly in the context of the growth of scientific and research databases that include raw data and information as well as reports and articles.

World Summit on the Information Society (WSIS) Declaration of Principles (2003)

8.17 The WSIS was a United Nations summit held in two phases – the first in Geneva in 2003, and the second in Tunis in 2005. The Geneva phase established the foundations for an Information Society to represent all attendees in matters of information and research.

8.18 The WSIS Declaration of Principles sets out some overarching principles, including:

- everyone has the right to freedom of opinion and expression, including the right to receive and impart information through any media regardless of frontiers;
- the sharing and strengthening of global knowledge for development can be enhanced by removing barriers to equitable access to information for economic, social, political, health, cultural, education and scientific activities; and
- the wide dissemination, diffusion and sharing of knowledge is important to encourage innovation and creativity.⁵⁹⁵

8.19 To this end, WSIS attendees stated:

We strive to promote universal access with equal opportunities for all to scientific knowledge and the creation and dissemination of scientific and technical information, including open access initiatives for scientific publishing.⁵⁹⁶

8.20 The WSIS Declaration committed the attendees to promote:

- the long-term systematic and efficient collection, dissemination and preservation of essential scientific digital data, for example, population and meteorological data in all countries; and
- principles and metadata standards to facilitate cooperation and effective use of collected scientific information and data as appropriate to conduct scientific research.⁵⁹⁷

⁵⁹⁴ <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.

⁵⁹⁵ <<http://www.itu.int/wsisis/docs/geneva/official/dop.html>> at 13 February 2007

⁵⁹⁶ <<http://www.wsisis-si.org/UNESCO/C7.e-science-texts.html>>.

8.21 Reference to “scientific knowledge” in the WSIS Declaration is reminiscent of the Berlin Declaration and the broad approach taken in determining what constitutes open access information.

8.22 Another focus of the WSIS Declaration was to turn “the digital divide into a digital opportunity” for developing countries in particular, where the benefits of information technology are unevenly distributed between rich and poor.⁵⁹⁸

8.23 The Tunis phase of the summit reaffirmed the principles established at the Geneva phase and reiterated that:

Access to information and sharing and creation of knowledge contributes significantly to strengthening economic, social and cultural development.

8.24 The attendees recognised that open access to information could not be achieved without first creating a trustworthy, transparent and non-discriminatory legal and policy environment. They placed some reliance on governments to develop and adopt regulatory frameworks, stating:

We are convinced that our goals can be accomplished through the involvement, cooperation and partnership of governments and other stakeholders, i.e. the private sector, civil society and international organizations.

8.25 The next WSIS sessions will take place in Geneva in May 2007. At these meetings, the attendees intend to develop “Action Lines.” These are action plans to implement the principles and goals set out at the previous WSIS sessions in 2003 and 2005.

Organisation for Economic Co-operation and Development (OECD) Declaration on Access to Research Data from Public Funding (2004)

8.26 The OECD is a group of 30 member countries who aim to facilitate and promote good governance⁵⁹⁹ The OECD:

Produces internationally agreed instruments, decisions and recommendations to promote rules of the game in areas where multilateral agreement is necessary for individual countries to make progress in a globalised economy.⁶⁰⁰

8.27 At a meeting in Paris in January 2004, the OECD member countries adopted the Ministerial Declaration on Access to Research Data from Public Funding.⁶⁰¹ “Research data from public funding” is defined as:

The research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government.⁶⁰²

⁵⁹⁷ Ibid.

⁵⁹⁸ See <<http://www.itu.int/wsis/docs/geneva/official/dop.html>> at 13 February 2007

⁵⁹⁹ <http://www.oecd.org/about/0,2337,en_2649_201185_1_1_1_1_1,00.html>.

⁶⁰⁰ Ibid.

⁶⁰¹ OECD, Committee on Scientific and Technological Policy, *Ministerial Declaration on Access to Research Data from Public Funding* C(2004)31/REV1

<<http://webdomino1.oecd.org/horizontal/oecdacts.nsf/Display/3A5FB1397B5ADFB7C12572980053C9D3?OpenDocument>> at 23 April 2007.

8.28 The Declaration recognised that:

An optimum international exchange of data, information and knowledge contributes decisively to the advancement of scientific research and innovation.⁶⁰³

8.29 OECD member countries declared their commitment to establishing open access regimes for digital research data derived from publicly funded projects, in accordance with specific objectives and principles, including:

Openness: Balancing the interests of open access to data to increase the quality and efficiency of research and innovation with the need for restriction of access in some instances to protect social, scientific and economic interests;

Legal conformity: Paying due attention, in the design of access regimes for digital research data, to national legal requirements concerning national security, privacy and trade secrets;

Protection of IP: Describing ways to obtain open access under the different legal regimes of copyright or other IP law applicable to databases as well as trade secrets.⁶⁰⁴

8.30 The 2004 Declaration was followed in May 2006 by a Draft Recommendation Concerning Access to Research Data from Public Funding. The Draft Recommendation reiterated the OECD member countries' commitment to open access and called on them to increase their efforts to develop policies and good practices relating to the accessibility, use and management of research data.⁶⁰⁵

8.31 In December 2006, the OECD Council agreed to a recommendation to member countries to take the Principles and Guidelines on Access to Research Data from Public Funding into consideration and:

apply them, as appropriate for each Member country, to develop policies and good practices related to the accessibility, use and management of research data.⁶⁰⁶

8.32 The recommendation instructs the OECD Committee for Scientific and Technological Policy to review its implementation as necessary and to review the Principles and Guidelines on Access to Research Data from Public Funding and, as appropriate, to:

take into account advances in technology and research practices, with the intention of further fostering international co-operation.⁶⁰⁷

8.33 The relevance of the OECD's 2004 Declaration to the development of strategic frameworks for scientific data management was acknowledged by PMSEIC Data Working Group in its 2006 report, *From Data to Wisdom: Pathways to*

⁶⁰² Ibid.

⁶⁰³ <<http://www.codataweb.org/UNESCOmtg/dryden-declaration.pdf>>.

⁶⁰⁴ Ibid.

⁶⁰⁵ <http://www7.nationalacademies.org/usnc-codata/OECD_Principles_and_Guidelines.pdf>.

⁶⁰⁶ OECD, *Recommendation of the Council concerning Access to Research Data from Public Funding* (2006)184

<<http://webdomino1.oecd.org/horizontal/oecdacts.nsf/Display/3A5FB1397B5ADFB7C12572980053C9D3?OpenDocument>> at 23 April 2007.

⁶⁰⁷ Ibid.

Successful Data Management for Australian Science (2006).⁶⁰⁸ The PMSEIC Data Working Group recommended that the OECD guidelines should be taken into account in the development of a strategic framework for management of research data in Australia.⁶⁰⁹

8.34 In relation to “open educational resources”, the OECD has released a 2007 report which discusses open licensing and frameworks for the free sharing of digital educational resources in the creation of virtual learning environments.⁶¹⁰ The report also advocates the use of Creative Commons licences to facilitate use of open educational resources.

OECD Working Party on the Information Economy (31 May 2006)

8.35 A useful set of general principles for public sector information policy, based on analyses of national and regional policies, were presented by the (late) Peter Weiss of the US Department of Commerce and Professor Burkert at the OECD Working Party on the Information Economy’s Workshop on Public Sector Information held in May 2006. The principles were summarised as follows:

1. **Inventory principle** - Public sector institutions (PSI) should make an inventory of their information holdings, update it regularly and actively make it generally and easily accessible.
2. **Access principle** - PSI holdings should be subject to a regime of access principles. These comprise the right of anyone to obtain PSI. Exemptions should only be based on consideration of personal privacy, preservation of significant private commercial interests where explicitly protected by copyright, or legitimate national security concerns.
3. **Quality principle** - PSI holdings should be provided in the same quality as they have been provided by the public sector.
4. **Cost principle** - The costs chargeable to any requester should not exceed marginal costs of distribution; there should be the possibility to waive such costs in cases where requesters can show a specific public interest is involved.
5. **Choice principle** - If available (or if easily transformable) Information should be provided in the requested format. The requester may be charged with transformation costs, provided administrative costs of recovering them do not exceed those.
6. **IP rights and control of origin principles** - PSI holdings should be exempted from IPR and also copyright and data-base protection regimes. The public sector should, however, be entitled to ensure through minimal regulation that responsibilities for any changes to the original information after its transfer are made appropriately transparent.
7. **Legitimate improvement principle** - Public sector institutions may extend, improve quality and format of their information provided they do so after a transparent procedure and in order to improve quality and/or extent of their services. Public bodies should not “feel compelled to discontinue a service that is to the public benefit simply because a commercial vendor chooses

⁶⁰⁸ Prime Minister’s Science, Engineering and Innovation Council, Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science*, (2006) <http://www.dest.gov.au/sectors/science_innovation/publications_resources/profiles/Presentation_Data_for_Science.htm> at 26 March 2007.

⁶⁰⁹ Ibid Recommendation 9.

⁶¹⁰ OECD, *Giving Knowledge for Free The Emergence of Open Educational Resources*, 2007 OECD Publishing <<http://www.oecd.org/dataoecd/35/7/38654317.pdf>> at 26 June 2007

to duplicate it. [...] Information vital to the public interest should not be “captured” by any entity, particularly in the private sector, which has economic reasons for controlling access”.⁶¹¹

8. **Continuity of obligations principle** - PSI activities even if transferred to the private sector are subject – to the extent of their privilege – to the same principles as PSI holdings.⁶¹¹

International Council for Science (ICSU) Scientific Data and Information Report (2004)

8.36 The ICSU is a global organisation representing both national scientific bodies and international scientific councils.⁶¹² It is actively involved in promoting freedom of access to scientific data and information, which it advocates through discussion forums, conferences and symposiums. In December 2004, the ICSU released its *Scientific Data and Information: A Report of the CSPR Assessment Panel*, which focused on open access to scientific data.⁶¹³ The Report was strongly influenced by the Berlin Declaration, the Bethesda Statement on Open Access Publishing, and the open access endorsements of both the OECD and the WSIS. The basis of the Report was expressed as follows:

Scientific progress relies on full and open access to data and on the open disclosure of research results in the scientific literature. A strong public domain for scientific data and information promotes greater return from the public investment in research by stimulating innovation and more-informed decision making. Principles of open access to scientific data and information can be applied to research data, metadata, or scientific publications, although the specific issues vary with each.⁶¹⁴

8.37 The ICSU made several important recommendations to promote and strengthen scientific data management:

- financial support for data and information management should be routinely included in research budgets and the criteria for evaluating research funding proposals should include evaluation of data management;⁶¹⁵
- efforts should be made to raise awareness of the increasingly important role played by institutional repositories in relation to the management and preservation of scientific information and the need to ensure that such repositories are properly resourced, developed and maintained;

⁶¹¹ H Burkert and P Weiss, ‘Towards a Blueprint for a Policy on PSI’ in G Aichholzer and H. Burkert (eds), *Public Sector Information in the Digital Age* (2004) as cited in DSTI/ICCP/IE(2005)2/FINAL, reproduced in OECD, Working Party on the Information Economy, *OECD Workshop on Public Sector Information: Summary* DSTI/ICCP/IE(2006)14, 27

<<http://www.oecd.org/dataoecd/34/42/37865140.pdf>>. For other documents presented at the workshop <<http://www.oecd.org/sti/digitalcontent>>.

⁶¹² See <http://www.icsu.org/5_abouticsu/INTRO.php>

⁶¹³

<http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf>.

⁶¹⁴ Ibid, 24.

⁶¹⁵ The importance of appropriate funding was also emphasised by the UK Office of Science and Innovation (OSI) e-Infrastructure Working Group in its report *Developing the UK’s e-infrastructure for science and innovation* <<http://www.nesc.ac.uk/documents/OSI/report.pdf>> at February 2007.

- all scientists should receive training in data management as part of their undergraduate and post-graduate education;
- guidelines should be developed for data management by employed scientists and their institutions; and
- collaborative development (involving information technology specialists, librarians, research scientists, government data producers, and donors) of standards and curricula for professional training for scientific data managers should be implemented.⁶¹⁶

8.38 A range of activities need to be coordinated if research data and information are to be effectively managed. These include:

- preservation of data and information, so that it will be available in digital formats and on media that can be used in the future;
- use of common metadata standards, to facilitate the identification, re-use and integration of scientific data and to provide information about data quality;
- permanent archiving of scientific data and information and compliance with institutional data archiving obligations where applicable;
- promotion of interoperability between systems and metadata standards to facilitate cooperation and effective use of data and information;
- ensuring data security and integrity;
- compliance with legal requirements, including obligations to protect personal privacy and to maintain confidentiality; and
- ensuring that IP laws as they relate to scientific data and information recognise the importance of full and open access to data for scientific research and educational purposes.⁶¹⁷

8.39 The ICSU Report is one of the few international reports to specifically examine the role of metadata. The ICSU encourages data repositories to ensure that standard metadata is available for all databases and records. It also takes the view that metadata should contain information on the legalities of, and the security and integrity measures employed in, the collection and management of relevant data. Finally, and most importantly, it advocates that metadata should be openly and freely accessible to all, through multidisciplinary metadata catalogues.

3. Governments and Public Sector Research Funding Bodies

8.40 There is an increasing amount of activity directed at the development and implementation of systems to enable information and content generated or held by

⁶¹⁶ ICSU, *Scientific Data and Information: A report of the CSPR Assessment Panel* (2004) 21.

⁶¹⁷ ICSU, *Scientific Data and Information: A report of the CSPR Assessment Panel*, (2004) 9-11.

public sector institutions and publicly-funded universities and research institutes to be more readily accessed and re-used.

8.41 Statements of the principles underlying access and re-use are central to the European Union's *Directive on the Re-use of Public Sector Information* (2003).⁶¹⁸ The importance of basing an access regime on an appropriate policy and guiding principles has been recognised in the Intrallect-University of Edinburgh report commissioned by the UK's Common Information Environment, *The Common Information Environment and Creative Commons* (2005) and in the Australian Government's Office of Spatial Data Management (OSDM) *Spatial Data Access and Pricing Policy* (2001).⁶¹⁹ In the US, scientific organisations have long had commitments to data sharing, through their ethical codes⁶²⁰ and publication policies.⁶²¹ For example, the National Science Foundation (NSF) Economics Program has long had a requirement that the data underlying an article produced as a result of an NSF project is to be deposited in a public archive.

United States

NIH Data Sharing Policy (2003)

8.42 NIH is the largest funder of basic biomedical research in the world, spending US\$27 billion in the 2005 financial year. It spends about US\$3 billion per year on biomedical research it carries out, as well as providing about US\$23 billion annually to fund external research.

8.43 NIH has been dealing with issues of data access, IP and licensing for many years. Its current policy on data sharing was published in February 2003.⁶²² NIH believes that data sharing promotes many of its research goals and is viewed as particularly important for unique data that cannot be readily replicated.⁶²³ Data sharing allows scientists to expedite the translation of research results into knowledge, products, and procedures to improve human health. NIH takes the view that all data should be considered for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data.

8.44 To facilitate data sharing, investigators submitting a research funding application to NIH for \$500,000 or more in any single year are expected to include a plan for sharing final research data for research purposes, or state why data sharing is

⁶¹⁸ Directive 2003/98/EC, 17 November 2003, OJ L345/90, 31 December 2003, available at <http://europa.eu.int/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf>

⁶¹⁹ Commonwealth Interdepartmental Committee on Data Access and Pricing, *A Proposal for a Commonwealth Policy on Spatial Data Access and Pricing* (2001) <<http://www.osdm.gov.au/osdm/policy/accessPricing/SDAP.pdf>>; summary available at <http://www.osdm.gov.au/osdm/policy/accessPricing_summary.html> at 14 July 2006.

⁶²⁰ For example, the American Sociological Association.

⁶²¹ For example, the American Psychological Association.

⁶²² See NIH *Data Sharing Policy* <http://grants.nih.gov/grants/policy/data_sharing/index.htm>; NIH *Statement on Sharing Research Data* <<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>> and NIH *Data Sharing Policy and Implementation Guidance* <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm>.

⁶²³ See <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave>.

not possible.⁶²⁴ In Australia, the ARC has now adopted a very similar strategy to essentially require (without actually mandating) Australian researchers to deposit their research results and reports into an open access repository.

8.45 The NIH policy on data sharing applies to:

- the sharing of *final research data* for research purposes; and
- basic research, clinical studies, surveys, and other types of research supported by NIH, with particular importance attached to the sharing of *unique data* that cannot be readily replicated.

“Final research data” is defined as:

Recorded factual material commonly accepted in the scientific community as necessary to document, support, and validate research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based. For most studies, final research data will be a computerized dataset. For example, the final research data for a clinical study would include the computerized dataset upon which the accepted publication was based, not the underlying pathology reports and other clinical source documents. For some but not all scientific areas, the final dataset might include both raw data and derived variables, which would be described in the documentation associated with the dataset.

“Unique data” is defined as:

Data that cannot be readily replicated. Examples of studies producing unique data include: large surveys that are too expensive to replicate; studies of unique populations, such as centenarians; studies conducted at unique times, such as a natural disaster; studies of rare phenomena, such as rare metabolic diseases.

8.46 NIH uses the term “Restricted Data” to refer to datasets that cannot be distributed to the general public for various reasons including participant confidentiality concerns, third party licensing agreements or national security considerations.

8.47 Recognising the breadth and variety of science that NIH supports, neither the precise content for the data documentation nor the formatting, presentation, or transport mode for data is stipulated. NIH recognises that a sensible and practical approach in one field or one study may not be appropriate for others. NIH encourages members of multiple disciplines and their professional societies to discuss data sharing, determine what standards and best practices should be proposed, and create a social environment that supports data sharing.

8.48 When applicants sign the face page of an NIH application, they are assuring compliance with policies and regulations governing research awards. NIH expects grantees to follow these rules in conducting the work described in the application. Where an application describes a data-sharing plan, the agency expects that plan to be enacted. Progress made with a data-sharing plan must be acknowledged in the progress report. In the final progress report, if not sooner, the grantee must state what

⁶²⁴ This requirement is in place for all applications on or after 1 October 2003.

steps have been taken with respect to the data-sharing plan. In the case of non-compliance (depending on its severity and duration), NIH can take various actions to protect the US Federal Government's interests in providing the funding. In some instances, NIH may make data sharing an explicit condition of subsequent awards.

8.49 NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. The specific time will be influenced by the nature of the data collected. Data from small studies can be analyzed and submitted for publication relatively quickly. If data from large epidemiologic or longitudinal studies are collected over several discrete time periods or waves, it is reasonable to expect that the data would be released in waves as data become available or main findings from waves of the data are published. NIH recognizes that the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort. NIH continues to expect that the initial investigators may benefit from first and continuing use, but not from prolonged exclusive use.

Canada

Canadian Institutes of Health Research (CIHR)

8.50 In 2006, CIHR released a draft policy on access to CIHR-funded research outputs.⁶²⁵ The policy applies to all research outputs that have been financially supported in whole or in part by CIHR, including industry-partnered research. A research output is conceptual or practical knowledge, data, information, or a physical or biological object developed during the course of CIHR-funded research, and is essential for reproducing results and for furthering research discoveries.

8.51 Research outputs covered under the draft policy are:

- peer-reviewed journal publications;
- research materials; and
- final research data.

8.52 The CIHR draft policy provides:

3. Research Data

- Final research data refers to the factual information that is necessary to replicate and verify research results. Data can include original data sets, data sets that are too large to be included in the peer-reviewed publication, and any other data sets supporting the research publication. Research data is typically an electronic data set, and may include interview transcripts and survey results provided confidential data and subject privacy is protected. Research data does not include lab books and unpublished research protocols, or physical objects like tissue samples.
- Grant and award holders should strive to make final data sets, generally in electronic form, available upon request after the publication date of a peer-reviewed publication.

⁶²⁵ Available at <<http://www.cihr-irsc.gc.ca/e/32326.html>>.

- For effective sharing of data, grant and award holders should ensure the quality of the data and have accompanying metadata (i.e., information that describes the characteristics of the data set) or codebooks.
- CIHR requires grant and award holders to deposit bioinformatics, atomic and molecular coordinate data, experimental data, as already required by most journals, into the appropriate public database immediately upon publication of research results.
- CIHR also requires grant and award holders to retain original data sets arising from CIHR-funded research for a minimum of five years after the last date of the "Authority to Use Funds" period of the grant. This applies to all data, whether published or not.
- The grant or award holder's institution and research ethics board may have additional policies and practices regarding the preservation, retention, and protection of research data that must be respected.

GeoConnections Canada

8.53 GeoConnections Canada is a national partnership launched in 1999, directed to build the Canadian Geospatial Data Infrastructure (CGDI). After successfully commissioning the CGDI, GeoConnections received further funding to evolve and expand the CGDI in response to community feedback. Over the course of their work they have released a number of reports on different aspects of geospatial data infrastructure.

The Dissemination of Government Geographic Data in Canada: Guide to Best Practices (2005)⁶²⁶

8.54 This report found that a “myriad of licensing terms” were currently being used for data sharing and that:

The complexity of the current government geographic data dissemination environment in Canada stems from the lack of a coordinated approach to dissemination and licensing of such data.⁶²⁷

8.55 The report suggested an integrated framework for government geographic data, which would help to provide easy access to the data. This framework would need to be developed logically, keeping in mind the various types of licensing arrangements commonly used by the government:⁶²⁸

In order to reduce the complexity of government geographic data licensing practices, concerted effort should be expended to develop an integrated approach to the distribution of government geographic data. Such an integrated framework cannot, however, be designed haphazardly. Its development needs to be achieved through an articulation of the values and objectives for the various types of licensing arrangements in common use across government, and must reflect an appreciation of the continuum of objectives underlying government geographic data dissemination policy.⁶²⁹

⁶²⁶

<cgdi-dev.ccrs.nrcan.gc.ca/publications/Best_practices_guide/Guide_to_Best_Practices_v12_finale_e.pdf>.

⁶²⁷ Ibid.

⁶²⁸ Ibid, 22.

⁶²⁹ Ibid, 22.

8.56 One of the key areas highlighted by the report is the need to expressly deal with copyright issues surrounding derivative datasets. The government retains copyright in their datasets but licences them out on certain terms. GeoConnections considered this to be a key issue in promoting innovation. The report states that:

Licensees of government geographic data should generally have the right to develop, manufacture and distribute commercially derived products they have or have caused to be created, which incorporate or derive from the licensed geographic data.⁶³⁰

8.57 The report also highlights that any licence that data is taken under must address the legal risks to the repository or database in providing the data, most likely through disclaimers of liability for use of the data. Additionally, any data accessed must be cited as licensed government data.

8.58 The Guide provides a quick reference chart to relevant issues:⁶³¹

- **Indemnification:** Indemnification acts to further manage the risk of licensor liability, to the extent possible, by requiring that the licensee warrant that they will not hold the licensor liable for damages arising from the use of the licensed data. Indemnification of the government data provider may not be enforceable in agreements struck between a data licensee and their own sub-licensees, to the extent that it can be managed. However, this does not materially elevate risk of government liability, as the sub-licensee would have to demonstrate that the government:
 - 1) knew or ought to have known that the data delivered was not accurate;
 - 2) owed a duty of care to the sub-licensee; and
 - 3) that the sub-licensee incurred damages as a result of the government's negligence.
- **Acknowledgement:** Acknowledgement of government data source is expressed as a requirement to cite the data source and include any government furnished metadata in any downstream products or services containing the originally licensed government data.
- **Derived Products:** A derived product is any product or service made operational through use of, or derived from the licensed government data. A key objective is to promote the development of derived products and services as a stimulus to socioeconomic growth. The end-user model restricts the further distribution only of derived products containing the originally licensed data.
- **IP:** The IP rights in the licensed data are never transferred as a result of the use or modification of the licensed data. Conversely, the IP rights resulting from any extension or enhancement of the licensed data, or development of derived products or services rest with the party giving rise to those enhancements and developments. The integrated framework recommends the use of automatically repeating terms of a fixed length. Termination of an agreement is affected automatically by breach of contract, by request from the licensee, or by mutual agreement.
- **Other Supporting Components:** The integrated framework employs the use of other standard clauses that reinforce and clarify various contractual aspects of each dissemination model for the benefit of the data producer and user. Common approaches are employed for preamble, most definitions, entirety of agreement, dispute resolution, etc.⁶³²

⁶³⁰ <cgdi-dev.ccrs.nrcan.gc.ca/publications/Best_practices_guide/Guide_to_Best_Practices_v12_finale_e.pdf>

⁶³¹ Ibid 70.

⁶³² Page 70-1 <cgdi-dev.ccrs.nrcan.gc.ca/publications/Best_practices_guide/Guide_to_Best_Practices_v12_finale_e.pdf>.

CGDI Roadmap

8.59 This roadmap takes into account a complex environment, which is grouped into seven thematic areas: technological, governance, legislative, legal, human resources, financial and government priorities. CGDI originally laid down seven guiding principles, called “founding principles”, and has extended this to ten in consultation with stakeholders by adding “building principles”.⁶³³

Founding Principles:

- **Open:** The CGDI will be based on open and interoperable standards and specifications for operational transactions and information exchange. “Open and shared” in this context means that the specifications are available for the world to take, use, and modify for other purposes. These specifications will be based on national and international standards where available.
- **Transparent:** The CGDI will allow users to access data and services seamlessly in a manner that removes the complexities of the underlying technology and information infrastructure. “Seamless” implies the elimination or hiding of artificial boundaries introduced by jurisdictions or by technical issues such as scale or quality of information.
- **Cooperative:** The CGDI will help organizations from the private sector, government and academia collaborate. The CGDI will define common technologies and standards rather than prescribe single or proprietary implementation solutions.
- **Evolving:** The network of participating organizations will continue to encompass new requirements and business applications for information and service delivery to their respective users. The CGDI will evolve to meet these changing requirements and developments.
- **Timely:** The CGDI will define and recommend technologies and services that will support timely or real-time access to information.
- **Self-sustaining:** The CGDI will be sustained through the contributions of the participating organizations and the broad user-community and through being relevant to these groups.
- **Self-organizing:** The CGDI will enable various levels of participating organizations to contribute geospatial information, metadata, services and applications without the requirement for centralized administration, access, and data warehousing.

Building Principles:

- **User-driven:** The CGDI will emphasize the nurturing of and service to a broad user community. This approach will include user-driven developments, services, and enhancements that facilitate policy and decision making.
- **Closest to Source:** The CGDI will build upon its principle of self organization to encourage organizations that are closest to source to provide data. This emphasis will increase quality and efficiency by eliminating duplication and overlap. The CGDI will need to be developed further through partnerships with municipal, provincial and territorial governments; other federal departments and agencies; as well as international sources.

⁶³³ Page 26

<http://www.geoconnections.org/publications/tvip/Roadmap_E/CGDI_Roadmap_final_E.pdf>.

- **Secure:** The CGDI recognizes the importance of openness but realizes that a need exists to secure sensitive or proprietary data. This need for security is augmented by the requirement for high stability and data reliability.

GeoConnections Policy Advisory Node ⁶³⁴

8.60 The Policy Advisory Node of the GeoConnections team has focused on five major areas of interest:

- roles and responsibilities – this involves defining the area of operation, considering the partnerships that may be forged between different actors, especially public and private sector;
- access – this is the general idea that the data compiled is a public good and should be as broadly available as possible;
- copyright and licensing – this is concerned with the constraints on the use of data, and includes the setting of licence fees, tracking of data usage and management of legal risk;
- pricing; and
- financing – this deals with the economic reality of finding and justifying funding for open access resources.

8.61 The GeoConnections Policy Advisory Node's report, *Archiving, Management and Preservation of Geospatial Data: Summary Report and Recommendations* (2005),⁶³⁵ recommends that:

Organisations should define and implement policies and practices for the creation, use, retention, dissemination, preservation and disposition of geospatial data.⁶³⁶

8.62 The report also suggests that organisations adopt a custodianship model where a custodian is responsible for compliance with policy and procedures, recommending that:

Organisations must establish authoritative responsibility centres that empower individuals with the ability to define and apply the information management principles required to ensure the integrity of an organisation's geospatial data holdings.⁶³⁷

United Kingdom and Europe

8.63 Internationally, one of the most significant initiatives in relation to access to information generated by public sector entities has been the European Union's

⁶³⁴ <<http://www.geoconnections.org/CGDI.cfm/fuseaction/policy.pgmOverview/id/392/gcs.cfm>>.

⁶³⁵ <<http://www.geoconnections.org/CGDI.cfm/fuseaction/policy.pgmOverview/id/392/gcs.cfm>>.

⁶³⁶ Page 10

<www.geoconnections.org/.../proCom_policy/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf>.

⁶³⁷ Ibid.

*Directive on the Re-use of Public Sector Information*⁶³⁸ (EU Directive) which was adopted by the European Parliament and Council on 17 November 2003.

8.64 Research Councils which provide public funding for research have progressively announced that they are adopting open access policies which require research results to be deposited in open access repositories. In February 2007, the European Commission announced that it plans to support more cost-free access to the results of scientific research.

Research Councils UK

8.65 The UK Medical Research Council (MRC) has a data sharing policy which requires published research results to provide links to the associated data. All applicants submitting funding proposals to the MRC are expected to include a statement explaining their strategy for data preservation and sharing.⁶³⁹

Research Information Network (RIN)

8.66 The Research Information Network (RIN) was established in early 2005 by a consortium of UK sponsors - the four Higher Education funding bodies, the three National Libraries and the seven Research Councils.⁶⁴⁰

8.67 The mission of RIN is:

To lead and co-ordinate new developments in the collaborative provisions of research information for the benefit of researchers in the UK.⁶⁴¹

8.68 In April 2007, RIN released a draft policy statement, *Stewardship of digital research data: a framework of principles and guidelines*.⁶⁴² The statement sets out a policy framework for the management of data. In explaining why a policy framework for data management is necessary, the RIN statement provides:

In order to produce high-quality research, researchers must have access to as wide a range as possible of the data and information produced by other researchers, as well as relevant information produced by other agencies in the UK and overseas. Similarly, successful dissemination and exploitation of research depends on effective flows of information between researchers and other individuals and organisations who have an interest in its results. A successful research and innovation system thus depends on the open exchange of ideas, information and knowledge.

The essential goals we are seeking to achieve are thus to facilitate the advancement of research and innovation, to enhance the efficiency and effectiveness of research, and to maximise the value of public and private investment in research. In pursuance of those goals, the fundamental policy objective is to ensure that:

⁶³⁸ *Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of the public sector information* [2003] OJ L 345/90

<http://europa.eu.int/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf

⁶³⁹ UK Medical Research Council, *MRC Guidance on Open Access to Published Research* (2006) <<http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc002548.pdf>>.

⁶⁴⁰ <<http://www.rin.ac.uk/about>> at 30 May 2007, see also <<http://www.rin.ac.uk/origins>> and <<http://www.rin.ac.uk/sponsors>> at 30 May 2007

⁶⁴¹ <<http://www.rin.ac.uk/about>> at 30 May 2007

⁶⁴² See <<http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20-%20published%20draft%20for%20consultation.pdf>> at 30 May 2007

*Ideas and knowledge derived from publicly-funded research should be made available and accessible for public use, interrogation, and scrutiny, as widely, rapidly and effectively as practicable.*⁶⁴³

8.69 RIN advocates a collaborative management strategy between institutions to facilitate transparent, equitable and consistent systems of information governance.⁶⁴⁴ The RIN policy statement sets out five principles for data management:

- the roles and responsibilities of researchers, research institutions and funders should be clearly defined and codes of practice should be established so that all parties, including creators and users of data, fulfil their responsibilities in relation to the care and management of data;
- digital research data should be created and collected in accordance with applicable international standards and there should be procedures for proper quality assurance of data;
- digital research data should be easily accessible and provided in an environment that maximises ease of use and protects the rights of those with legitimate interests in the data;
- models and mechanisms for managing and providing access to data should be efficient and cost-effective in the use of public and other funds; and
- data of long term values should be preserved and remain accessible for current and future generations.⁶⁴⁵

European Research Council

8.70 The Scientific Council of the European Research Council issued its *Statement on Open Access* in December 2006, stating:

5. [I]t is the firm intention of the ERC Scientific Council to issue specific guidelines for the mandatory deposit in open access repositories of research results – that is, publications, data and primary materials – obtained thanks to ERC grants, as soon as pertinent repositories become operational.

6. The ERC Scientific Council moreover hopes that research funders across Europe will join forces in establishing common open-access rules and in building European open access repositories that will help make these rules operational. To facilitate this process for EU funded research, it recommends that the European Commission sets up a task force including representatives from the various FP7 programmes (Cooperation, Ideas, People, ...) to develop an operational FP7 policy on open access by the end of 2007 (which takes in particular into account disciplinary differences and technological constraints).⁶⁴⁶

⁶⁴³

<<http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20-%20published%20draft%20for%20consultation.pdf>> at 30 May 2007, p3

⁶⁴⁴ Ibid, p4

⁶⁴⁵

<<http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20-%20published%20draft%20for%20consultation.pdf>> at 30 May 2007

⁶⁴⁶ Available at <<http://erc.europa.eu/pdf/open-access.pdf>>

European Union's Directive on the Re-use of Public Sector Information (EU Directive)⁶⁴⁷

8.71 The EU Directive represents the culmination of efforts that began in the late 1980s to facilitate the development of European data products based on public sector information.⁶⁴⁸ With a lack of clear policies or uniform practices in relation to access to and re-use of public sector information, European content firms engaging in the aggregation of information resources into value-added information products were perceived to be at a competitive disadvantage in comparison to their US counterparts. The lack of harmonisation of policies and practices regarding public sector information resources among the EU Member States was regarded as a barrier to the establishment of European information products based on information obtained from different countries.⁶⁴⁹

8.72 EU Member States were required to bring their national laws into conformity with the EU Directive by 1 July 2005,⁶⁵⁰ and to review the application of the EU Directive by 1 July 2008.⁶⁵¹ By 15 December 2005, 12 countries (including France, Ireland, Italy, Sweden, the Netherlands and the UK) had notified the European Commission that they had given effect to the Directive.⁶⁵² In the UK, the Directive has been given effect by the *Re-use of Public Sector Information Regulations 2005*,⁶⁵³ which came into force on 1 July 2005. In May 2005, the UK government established

⁶⁴⁷ Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of the public sector information [2003] OJ L 345/90

<http://europa.eu.int/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf>.

⁶⁴⁸ See, in particular the Commission of the European Communities', *Guidelines for Improving the Synergy Between the Public and Private Sectors in the Information Market* (1989)

<http://europa.eu.int/information_society/policy/psi/docs/pdfs/1989_public_sector_guidelines_en.pdf>. The agenda appears to have been reinvigorated by a major policy conference on public sector information, sponsored by the European Commission (Stockholm, 1996). See *History*

<http://europa.eu.int/information_society/policy/psi/history/index_en.htm>.

⁶⁴⁹ Problems were identified with response times to requests for information, pricing, existing exclusive deals and the overall lack of transparency: see *Background*

<http://europa.eu.int/information_society/policy/psi/directive/index_en.htm>. Such problems had been identified as far back as the 1980s, in the Introduction to the Commission of the European Communities, *Guidelines for Improving the Synergy Between the Public and Private Sectors in the Information Market* (1989) 5:

Governments and public sector bodies collect large amounts of data and information, as part of their routine functions, which could be made available to the private sector for the construction and marketing of electronic database services. The private sector is well placed to combine information from a variety of government sources, and its prime function is to produce and distribute information products oriented to the needs of the market. In order to develop and strengthen the information industry, a positive initiative is required from governments, to encourage the use and exploitation of public sector data and information. However, there are few convergent policies or guidelines within Member States relating to the role of the public sector in this area. In addition, if there are different policies operating in the different Member States, then it will be very difficult to develop the market. It is therefore desirable that national policies, as far as they exist, be coordinated at the Community level in order to allow the majority of the EC countries not yet having such a policy to follow these orientations on a national level.

⁶⁵⁰ Article 12.1

⁶⁵¹ Article 13

⁶⁵² See *Implementation*

<http://europa.eu.int/information_society/policy/psi/implementation/index_en.htm

⁶⁵³ Statutory Instrument 1515 of 2005 <<http://www.opsi.gov.uk/si/si2005/20051515.htm>>.

an Office of Public Sector Information with responsibility for the coordination of policy standards on the re-use of public sector information.⁶⁵⁴

8.73 The EU Directive establishes a minimum set of rules governing the re-use and the practical means of facilitating re-use of existing documents held by public sector bodies of EU Member States.⁶⁵⁵

8.74 The principles set out in the EU Directive may be summarised as follows:

- where re-use of public sector documents is permitted, the documents will be re-usable for commercial or non-commercial purposes (in accordance with the conditions in Articles 5 – 11) and, where possible, will be made available through electronic means;⁶⁵⁶
- public sector bodies are - through electronic means where possible and appropriate - to process requests for re-use and make documents available for re-use to applicants or, if a licence is needed, to finalise the licence offer to the applicant within a reasonable time;⁶⁵⁷
- public sector bodies shall make documents available in any pre-existing format or language, through electronic means where possible and appropriate;⁶⁵⁸
- where charges are made:

The total income from supplying and allowing re-use of documents shall not exceed the cost of collection, production, reproduction and dissemination, together with a reasonable return on investment;⁶⁵⁹

- any conditions and charges applying to the re-use of documents must be transparent, i.e. they must be pre-established and published, through electronic means where possible and appropriate;⁶⁶⁰
- public sector bodies may allow for re-use of documents without conditions or may impose conditions, where appropriate, in the form of a licence, and such

⁶⁵⁴ See the Office of Public Sector Information's website <<http://www.opsi.gov.uk/>>. The OPSI, attached to the Cabinet Office, will advise on and regulate the operation of the re-use of public sector information, and will set standards and provide a practical framework to increase transparency and remove obstacles to re-use.

⁶⁵⁵ Note that 'public sector body' is defined in Article 2 as 'the State, regional or local authorities, bodies governed by public law and associations formed by one or several such authorities or one or several such bodies governed by public law'.

Article 1.1. The Directive is required to be implemented by EU Member States. An earlier attempt at providing voluntary guidelines for exploitation by the private sector of public sector information had little impact. See Commission of the European Communities', *Guidelines for Improving the Synergy Between the Public and Private Sectors in the Information Market* (1989)

<http://europa.eu.int/information_society/policy/psi/docs/pdfs/1989_public_sector_guidelines_en.pdf>.

⁶⁵⁶ Article 3.

⁶⁵⁷ Article 4.

⁶⁵⁸ Article 5.

⁶⁵⁹ Article 6.

⁶⁶⁰ Article 7.

conditions are not to unnecessarily restrict possibilities for re-use and are not to be used to restrict competition;⁶⁶¹

- licences can be adapted to meet particular licence applications and are available in electronic form so that they can be processed electronically and Member States are to encourage all public sector bodies to use the standard licences;⁶⁶²
- practical arrangements must be in place to:

Facilitate the search for documents available for re-use, such as asset lists, accessible preferably online, of main documents and portal sites that are linked to decentralized asset lists;⁶⁶³

- any conditions on the re-use of documents must be “non-discriminatory for comparable categories of re-use”;⁶⁶⁴
- where documents are re-used by a public sector body as input for commercial activities falling outside the scope of its public tasks, the same charges and other conditions are to apply to the supply of the documents for these activities as apply to other users;⁶⁶⁵
- the re-use of documents is to be open to all potential actors in the market:

Contracts or other arrangements between the public sector bodies holding the documents and third parties shall not grant exclusive rights”;⁶⁶⁶ and

- exclusive arrangements established after the entry into force of the Directive are to be transparent and made public and all existing exclusive arrangements that do not qualify for the exception in Art 11.2 are to be terminated at the end of the contract or not later than 31 December 2008.⁶⁶⁷

8.75 While the EU Directive has established an EU-wide legal framework governing policies and practices relating to re-use of public sector information, it should be viewed alongside other EU initiatives designed to make digital content in Europe more accessible, usable and exploitable. Of particular relevance is eContentplus, a 4 year program (2005-08) established by the European Commission’s Directorate-Generale for Information Society and Media, with a budget of €149 million to:

⁶⁶¹ Article 8.1.

⁶⁶² Article 8.2.

⁶⁶³ Article 9.

⁶⁶⁴ Article 10.1.

⁶⁶⁵ Article 10.2.

⁶⁶⁶ Article 11 of the *Directive*. There is a public interest exception to this prohibition in Article 11(2): ‘Where an exclusive right is necessary for the provision of a service in the public interest, the validity of the reason for granting such an exclusive right shall be subject to regular review, and shall, in any event, be reviewed every three years. The exclusive arrangements established after the entry into force of this Directive shall be transparent and made public.’

⁶⁶⁷ Article 11.3.

Tackle organisational barriers and promote take up of leading-edge technical solutions to improve accessibility and usability of digital material in a multilingual environment.⁶⁶⁸

8.76 The eContentplus program has targeted three specific areas where development has been slow and the program can have a maximum impact:

- geographic information;
- educational content; and
- cultural, scientific and scholarly content.⁶⁶⁹

8.77 The aim of the program is to facilitate:

access to digital content, its use and exploitation, enhancing quality of content with well-defined metadata, and reinforcing cooperation between digital content stakeholders.

Australia

Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC)

8.78 In December 2006, the ARC and NHMRC released their policies for research projects for which they will provide funding in 2008,⁶⁷⁰ creating a link between funding of research projects and accessibility of the resultant outputs in appropriate subject or institutional repositories.

8.79 The funding rules for ARC Discovery Projects for 2007 (projects for which funding will commence in 2008)⁶⁷¹ effectively mandate that data resulting from funded research is to be made available in accordance with open access principles. The ARC policy does not expressly state that research outputs **must** be deposited in an open access repository, but simply encourages researchers to do so. However, if researchers do not intend to deposit the data from a funded project into a repository within a six-month period, they are required to explain their reasons for not doing so in the project's Final Report. This requirement applies to data but not to publications. The obligation to justify non-compliance means that, at least as far as research data is concerned, it is little short of a direct mandate. The ARC funding rules for Discovery Projects 2007 state:⁶⁷²

⁶⁶⁸ See the eContentplus program website at

<http://europa.eu.int/information_society/activities/econtentplus/index_en.htm>. The Decision No. 456/2005/EC of the European Parliament and of the Council, of 9 March 2005, establishing a multilateral Community programme to make digital content in Europe more accessible, usable and exploitable, Official Journal of the European Union L79/1, 24 March 2005, is available at <http://europa.eu.int/information_society/activities/econtentplus/docs/prog_decision_2005/econtentplus_decision_en.pdf>

⁶⁶⁹ For details of the eContentplus Work Program across these areas of content, see <http://europa.eu.int/information_society/activities/econtentplus/docs/call_2005/ecp_work_programme_2005.pdf>

⁶⁷⁰ See Bernard Lane, 'ARC sold on open access to research', *The Australian* 13 December 2006, <<http://www.theaustralian.news.com.au/story/0,20867,20917528-12332,00.html>>.

⁶⁷¹ Australian Research Council, *Discovery Projects Funding Rules for Funding Commencing in 2008* <http://www.arc.gov.au/pdf/DP08_FundingRules.pdf>.

⁶⁷² Ibid 13.

1.4.5.1. The Australian Government makes a major investment in research to support its essential role in improving the wellbeing of our society. To maximise the benefits from research, findings need to be disseminated as broadly as possible to allow access by other researchers and the wider community.

1.4.5.2. The ARC acknowledges that researchers take into account a wide range of factors in deciding on the best outlets for publications arising from their research. Such considerations include the status and reputation of a journal or publisher, the peer review process of evaluating their research outputs, access by other stakeholders to their work, the likely impact of their work on users of research and the further dissemination and production of knowledge. Taking heed of these considerations, the ARC wants to ensure the widest possible dissemination of the research supported under its funding, in the most effective manner and at the earliest opportunity.

1.4.5.3. The ARC therefore encourages researchers to consider the benefits of depositing their data and any publications arising from a research project in an appropriate subject and/or institutional repository wherever such a repository is available to the researcher(s). If a researcher is not intending to deposit the data from a project in a repository within a six-month period, he/she should include the reasons in the project's Final Report. Any research outputs that have been or will be deposited in appropriate repositories should be identified in the Final Report.

8.80 In December 2006, the NHMRC released its Project Grants funding policy for funding commencing in 2008.⁶⁷³ The NHMRC policy encourages open access for data and publications resulting from the NHMRC funds and requires research outputs that have been or will be deposited in appropriate repositories to be identified in the project's Final Report. However, the NHMRC policy does not go as far as the ARC policy in requiring an explanation of reasons why research data will not be deposited in a repository within six months. The NHMRC policy states:

16.2 Dissemination of Scientific Results

To maximise the benefits from research, findings need to be disseminated as broadly as possible to allow access by other researchers and the wider community. The NHMRC encourages researchers to consider the benefits of depositing their data and any publications arising from a research project in an appropriate subject and/or institutional repository wherever such a repository is available to the researcher(s). Any research outputs that have been or will be deposited in appropriate repositories should be identified in the Final Report.⁶⁷⁴

8.81 In December 2006, the ARC made a submission on the Productivity Commission's draft research report on Public Support for Science and Innovation,⁶⁷⁵ which had suggested (in draft finding 5.1) that:

Published papers and data from ARC and NHMRC-funded projects should be freely and publicly available.

8.82 In its response to the Productivity Commission's draft research report, the ARC addressed the issue of the costs of funding open access, recommending that consideration be given to funding of institutional open access repositories.⁶⁷⁶

⁶⁷³ Available at <http://www.nhmrc.gov.au/publications/_files/profundingpol.pdf>.

⁶⁷⁴ NHMRC *Project Grants Funding Policy – for projects commencing in 2009* 21.

⁶⁷⁵ Productivity Commission, *Public Support for Science and Innovation, Draft Research Report*, (November 2006) <<http://www.pc.gov.au/study/science/draftreport/index.html>>.

⁶⁷⁶ Australian Research Council, *Response to the Productivity Commission Draft Research Report – Public Support for Science and Innovation*, (December 2006)

Office of Spatial Data Management (OSDM) Spatial Data Access and Pricing Policy (2001)

8.83 In 2001, the Australian Government released the reports: *Positioning for Growth – the Spatial Information Industry Action Agenda*⁶⁷⁷ and *A Proposal for a Commonwealth Policy on Spatial Data Access and Pricing*.⁶⁷⁸ The Office of Spatial Data Management (OSDM) is charged with implementing the Australian Government's Spatial Data Access and Pricing Policy.

8.84 The Policy establishes the following principles as the basis for spatial data access and re-use:

- fundamental spatial data will be provided free of charge over the Internet, or where that is not possible, at no more than the marginal cost of transfer for packaged products or at the full cost of transfer for customised services;
- there will be no restrictions on commercial value-adding to the listed fundamental spatial datasets; and
- spatial data will be provided subject to a licence setting out the conditions of transfer.

8.85 The policy encourages the two-way exchange of spatial data between Commonwealth departments and agencies. The Commonwealth must be able to enter into whole-of-government, multi-lateral arrangements with State and Territory governments, for the exchange for spatial data to maximise mutual benefit. Additionally, the efficiency and effectiveness of the Commonwealth's investment in spatial data is increased when data conforms to common standards and there is a consistent approach to access and pricing across all government agencies.

8.86 The policy goals are met through a licence that provides access to data. The licence sets out the conditions under which data can be used, and the rights and responsibilities of both the data provider and the data receiver. A licence enables the Commonwealth to retain copyright in any spatial dataset that is made available to another party. It requires the user to acknowledge that copyright in the fundamental data is vested in the Commonwealth, and to absolve the Commonwealth from any liability arising out of the subsequent use of the data or a product developed from the data. The licence also requires licensees to report their value-adding activity to enable the Government to assess the effectiveness of this policy.

8.87 OSDM's pricing policy is summarised as follows:⁶⁷⁹

All fundamental spatial data should be freely available at no more than marginal cost of transfer in order to maximise the net economic and social benefits arising from its use. As user

<http://www.arc.gov.au/pdf/response_PCdraftresearchreport_06.pdf>.

⁶⁷⁷ Available at

<http://www.industry.gov.au/assets/documents/itrinternet/SIAA_Positioning_20050606100443.pdf>.

⁶⁷⁸ Available at <<http://www.osdm.gov.au/osdm/policy/accessPricing/SDAP.pdf>>.

⁶⁷⁹ Summary is available on the OSDM web site

<http://www.osdm.gov.au/osdm/policy/accessPricing_summary.html>.

requirements and technology trends converge, all agencies will make fundamental spatial data available through their web sites. This is consistent with the Access Policy and the broader Government Online initiatives. As datasets become accessible over the Internet, the marginal cost of transfer approaches zero. Therefore, all fundamental spatial data will eventually be made available free of charge.

The basic elements of this Pricing Policy are:

- custodians of fundamental spatial data will make that data freely available through the Internet at no cost, as soon as appropriate technology becomes available within the custodian agency;
- fundamental spatial data distributed as packaged products will be made available at a price not exceeding the marginal cost of transfer;
- fundamental spatial data distributed as customised products will be made available at a price not exceeding the full cost of transfer;
- there will be no restrictions on commercial use or value-added activities related to fundamental spatial data, as defined in the Schedule to the Policy, although copyright may be reserved by the Commonwealth.

The cost of providing fundamental spatial data as packaged products (e.g. CDs) or customised products (e.g. significant staff time and other resources to generate) is a legitimate charge to users - hence these may be made available at a price. However, data accessed through these mechanisms will also be available free over the Internet, as each agency develops this capability.

8.88 Metadata relating to all data under the pricing policy is maintained in a special database, which is made available at no cost as a way of meeting accessibility and transparency principles.

8.89 Both the access and pricing policies are based on the following principles:

- Community Access: That all sectors of the community have a right to easy, efficient and equitable access to government information under conditions that ensure that technology, data formats, institutional arrangements, location, costs and conditions do not inhibit their use;
- Access and pricing arrangements must support the objectives of the Commonwealth in relation to online service delivery;
- Exchange of Data: This policy encourages the two-way exchange of spatial data between Commonwealth departments and agencies. The Commonwealth must be able to enter into whole-of-government, multi-lateral arrangements with State and Territory governments, for the exchange for spatial data to maximise mutual benefit;
- Net Benefits: Pricing decisions must be based on maximising net benefits to the community arising from better decision making and ready community access to quality spatial data;
- Efficiency of Commonwealth Spatial Data Programs: To maximise the efficiency and effectiveness of the Commonwealth's investment in spatial data, it is necessary to adopt a whole-of-government approach to sharing data and avoidance of duplication of effort and expenditure;
- Uniformity and Consistency: The efficiency and effectiveness of the Commonwealth's investment in spatial data is increased when data conforms to common standards and there is a consistent approach to access and pricing across all government agencies;

- **Copyright and Use:** The Commonwealth Government will retain copyright in any spatial dataset that is made available to another party, even if no licence fees or royalties are sought, and even if the other party is extended very liberal rights in the use of the data. Commonwealth copyright should be explicitly noted in relation to the making available of datasets;
- **Data Licence:** All provision of spatial data under this policy shall be accompanied by a licence clearly setting out the conditions under which the data may be used, the rights and responsibilities of the data provider, and the rights and responsibilities of the data receiver. The licence will require the user to acknowledge that copyright over the fundamental data is vested in the Commonwealth and to absolve the Commonwealth from any liability arising out for the subsequent use of the data or a product developed from the data, and should seek to encourage accuracy in reproduction. The licence should also require licensees to report their value-adding activity to enable the Government to assess the effectiveness of this policy;
- **Custodianship:** Agencies must adopt best practices in managing data and must follow the principles of custodianship established by ANLIC;
- **Standards:** Fundamental spatial data must conform to international and national standards, as identified by the CSDMG;
- **Industry Development:** Access and pricing arrangements must facilitate development of an innovative and competitive spatial information industry in Australia; and
- **Rights and Obligations:** The rights of the individual and the Commonwealth in relation to confidentiality, privacy, security and IP must be preserved. Government legislation and Australia's international obligations must be complied with.

OSDM Data Custodianship Guidelines

8.90 The Custodianship Guidelines outline what OSDM considers to be the rights and responsibilities of custodians of spatial data:

Custodianship reinforces the concept of one agency being ultimately responsible for a dataset which other agencies might use. This gives users of that data confidence in the level of integrity, precision and completeness of the data.

The coordination of custodianship, through mechanisms such as the OSDM, assists in avoiding duplication of effort and unnecessary cost to the government, thus achieving maximum benefit from investment.⁶⁸⁰

8.91 The overarching principle is that data custodianship is necessary to ensure:

Accountability for the care, maintenance and credibility of information.⁶⁸¹

8.92 A custodian has various rights and responsibilities with respect to a particular dataset, including:

- storing data;
- maintaining and revising data;
- ensuring data security;
- providing metadata;
- promoting data use;

⁶⁸⁰ <www.osdm.gov.au/policy/custodianship/CustodianshipGuidelines.pdf>.

⁶⁸¹ Page 36 <www.osdm.gov.au/policy/custodianship/CustodianshipGuidelines.pdf>.

- facilitating data access;
- administering data distribution;
- charging for data or recovering costs associated with data supply, consistent with agency and jurisdictional policies;
- consulting with users;
- preserving data over time; and
- complying with legislation, policies and guidelines.⁶⁸²

The custodian also has an obligation to periodically survey the present and future needs of user groups and to address issues affecting existing and potential users of the data.⁶⁸³

8.93 Some of these responsibilities clearly relate more closely to general spatial data than academic research data. The custodian of an academic data repository will not determine the priorities for data capture; this will be conducted by the primary researchers that the repository serves. Responsibility for acquisition and integration will fall upon the researchers who use or reuse the data in the repository. The responsibilities set out by OSDM are all issues to be considered when establishing a data repository.

8.94 The Custodianship Guidelines leave the issue of copyright ownership of derived datasets to the individual custodians, but do deal with the issues facing data custodians where derivative datasets are created from entries in the repository. The guidelines address this issue as follows:

When a dataset is modified, by enhancement or integration with other data, and especially through a series of progressively more derived information products (a 'value chain'), then the responsibility of the original custodians will be attenuated through each step in the chain. The rights and responsibilities of custodians of original data in relation to reporting, accounting, liability, copyright enforcement, acknowledgement, etc., with progressively derived data should be resolved on a case-by-case basis...It is desirable that, regardless of any copyright, ownership or custodianship rules, agreements (or arrangements between agencies which form part of the Australian Government) relating to the modification of datasets define a clear point at which pedigree of copyright (and custodianship responsibility) will be deemed abandoned.⁶⁸⁴

8.95 As part of the requirement to distribute the data in a repository, the Custodianship Guidelines highlight the importance of "Data Quality Statements". These provide users with information as to the source, reliability, accuracy, completeness and currency of any spatial datasets supplied. Additionally, data users are encouraged to take note of any limitations imposed on the data, and not use the data for purposes beyond its originally intended use.⁶⁸⁵ The requirement to provide a Data Quality Statement links directly to questions of liability and warranty. The information in the Data Quality Statement will come from the metadata record, which should be completed by the depositor of the data and checked by the custodian before the data is accepted (as part of the data intake process).

⁶⁸² Page 37 <www.osdm.gov.au/policy/custodianship/CustodianshipGuidelines.pdf>.

⁶⁸³ Ibid.

⁶⁸⁴ Page 39 <www.osdm.gov.au/policy/custodianship/CustodianshipGuidelines.pdf>.

⁶⁸⁵ Page 43 <www.osdm.gov.au/policy/custodianship/CustodianshipGuidelines.pdf>.

8.96 The principle of custodianship has the clear benefit of ensuring that responsibility for compliance is managed centrally.

The custodian is the named agency which is responsible for ensuring that, in capturing, storing, using and supplying the data, any relevant legislation is complied with. Thus the custodian is obligated to keep abreast of all developments in legislation. Acts which are relevant include those on Freedom of Information, Privacy, Copyright and Archives.⁶⁸⁶

8.97 The OSDM Custodianship Guidelines suggest that custodians should be responsible for metadata compliance. This is a common theme in spatial data reports produced by the Commonwealth Government. However, metadata creation and verification in an academic repository is most efficiently handled by the creators of the dataset, not the custodian. Anything else results in a perception that the data may be “thrown over the fence” to the repository and from that point on ignored. The custodian should be responsible for ensuring compliance with metadata technical and quality standards rather than creation.

ANZLIC Model Data Access and Management Agreement

8.98 This agreement (described internally as a “protocol”) was developed from the Data Access and Management Agreement which was signed by all Australian jurisdictions in September 2001 as part of the Australian National Land and Water Resources Audit. The protocol explains that all jurisdictions in Australia have agreed on the national principles for spatial data management. The protocol aims to maximise access to data and minimise duplication and overlap in data collection. It:

Defines access, ownership, custodianship, archiving and updating arrangements for the data collected, developed for and used in partnership projects.

8.99 The goals of the protocol clearly draw on the OSDM Spatial Data Access and Pricing Policy (2001) above. The guiding principles for the protocol are:

- Efficiency: Unnecessary duplication of data collection and management between governments, agencies, industry groups and the community should be minimised;
- Simplicity: Data access and management arrangements should be simple to understand and designed to minimise compliance costs, preferably through the application of consistent and Australia-wide standard approaches such as this agreement; and
- Standards: Arrangements should be consistent across jurisdictions and be delivered within the framework for development of the Australian Spatial Data Infrastructure.

8.100 The relevant principles of the protocol are:

- Data is developed and maintained to meet agreed international or national guidelines or standards for the management of spatial information as endorsed by ANZLIC or through national coordination arrangements. This will ensure the data are comparable and consistent where required and can be used for various applications;
- Data is documented in the Australian Spatial Data Directory. The documentation must be current and provide enough information for users to determine whether the data are suitable

⁶⁸⁶ Page 44 <www.osdm.gov.au/policy/custodianship/CustodianshipGuidelines.pdf>.

for their purpose. This will ensure that users can easily find out whether suitable data already exist and the limitations on use of those data sets;

- There is easy, efficient and equitable access to data for all sectors of the community in format, location, cost and under conditions that does not inhibit their use. This will ensure that users can obtain the data;
- Data is accompanied by a licence when transferred, clearly setting out the conditions under which the data may be used, the rights and responsibilities of the data provider, and the rights and responsibilities of the data receiver. Licence arrangements are required to ensure that the spatial information is accessible, while protecting copyright, IP, privacy and confidentiality. The rights of the individual and governments in relation to confidentiality, privacy, security and IP must be preserved. This will ensure that the rights of all parties are protected and understood; and
- Before funding data collection, organisations and jurisdictions should actively identify and exploit the many existing opportunities for cooperation and sharing of fundamental spatial data to avoid duplication and to maximise benefits of investment in data collection.

Australian eResearch Sustainability Survey

8.101 The Australian Partnership for Sustainable Repositories (APSR) project, funded by DEST, published the results of its Australian eResearch Sustainability Survey in October 2006 in the report *Sustainability Issues for Australian Research Data*.⁶⁸⁷

8.102 The APSR Sustainability Report stated:

The development of standards is critical for the sharing of data. Research communities with effective data sharing have developed standards based structures for sharing data.

8.103 Standards for data sharing do not exist for some research communities, while for others there is more than one relevant standard and no advice available about which standard is most appropriate. Research communities without standards need assistance in this process and can be guided by intermediaries such as APSR, who can bring together researchers who have been through this experience with those still to do so. The APSR Report suggested collaborative effort, mediated by outside agencies (such as APSR), to develop or select standards for data deposit. However, it did note that standardisation of formats can take a long time to achieve considering the international and fragmented nature of research communities.⁶⁸⁸

8.104 The APSR Report considered how researchers make value decisions about what data to retain and what to discard, observing that:

Value is based on uniqueness, time slice, cost and ease of reproduction, but not always with an appreciation of the opportunities for re-use outside of their project or even their domain. Value decisions are also made by omission as data collections are stored but not migrated across media or software.⁶⁸⁹

8.105 The survey results on compliance for preservation of data were also not encouraging, being that:

⁶⁸⁷ <<http://www.apsr.edu.au/currentprojects/currentprojects.htm>>.

⁶⁸⁸ Page 5 <<http://www.apsr.edu.au/aeres/>>.

⁶⁸⁹ Page 12 <<http://www.apsr.edu.au/aeres/>>.

The existing compliance environment for the preservation of data was not seen as strong and the current requirements were typically not being observed. This is related to a number of factors for researchers – funding, time, culture, rewards and infrastructure. There is no clear responsibility for the enforcement and auditing of compliance. This weakens the existing system and reduces the chances of success of a new system.⁶⁹⁰

8.106 The report recommends a more formal approach for the sustainability of data which would address the threats to sustainability and deliver benefits for future research.⁶⁹¹ Such an approach needs to provide cohesion, coordination, collaboration and compliance.⁶⁹² The report referred to this approach as “a national data stewardship framework.” The key features would be:

- A distinct administrative home for the task of data sustainability;
- The use of an existing layer of repositories and data centres for the provision of data storage and sustainability services for data no longer actively required by those who generated it;
- Data reviewed for sustainability by the appropriate research community;
- A level of certification for the repository structure which allows clear understanding by all parties of the range and depth of services to be provided by individual repositories;
- The requirement that institutions receiving significant research grants develop data management plans which include certified repositories;
- A level of certification linked to research funds received – the more research funds received the higher the level of certification required;
- The linking of certification with the provision of advisory services by repositories about data management;
- The mandating of a data management plan for grant applications;
- The review of data management plans by grant review panels;
- The provision of funds in addition to the research funds, where appropriate, for data management;
- Consideration of data creation and management as some part of research metrics for researchers;
- The addition to these metrics if data is used by others;
- The provision of funding to existing repositories by institutional grants, based on research quantum and later to be based on research data preserved, for use in skill development, maintenance and projection;
- Ongoing external assessment of system costs and benefits; and

⁶⁹⁰ Page 14 <<http://www.apsr.edu.au/aeres/>>.

⁶⁹¹ Page 50 <<http://www.apsr.edu.au/aeres/>>.

⁶⁹² Page 50 <<http://www.apsr.edu.au/aeres/>>.

- An administrative apparatus that audits and amends the national system.

Australian Government Information Interoperability Framework⁶⁹³

8.107 In April 2006, the Australian Government Information Management Office released a report by the Information Interoperability Working Group. Information interoperability is:

The ability to transfer and use information as a uniform and efficient manner across multiple organisation and information technology systems.

8.108 The information management principles underpinning the framework are:⁶⁹⁴

1. Manage information as an asset and a strategic resource: The importance of regarding information as an asset and a strategic resource should be promoted. Expenditure on information management should be treated as an investment, not a liability. Information should be managed according to its value to the government and its agencies, with a focus on high value information assets;
2. Standardise information management practices: Information management practices should be standardised across government to share and improve processes for accessing information. Information should be managed according to lifecycle management protocols and be transferable across organisations, subject to the requirements of privacy, confidentiality, IP and associated security standards;
3. Generate information to support decision making: Accurate, timely and relevant information should be available to share with others who have an appropriate business requirement. This principle is based on the need to continually work towards optimised agency and whole-of-government service delivery and organisational capability, supported by evidence-based decisions;
4. Collect quality information: The collection of information should aim to be accurate, relevant, timely, reliable and cost-effective. The impost of information collection on the Australian public should be minimised as far as possible. Duplication and rework for staff should be minimised;
5. Re-use information from single authoritative source: Information should be collected in a consistent manner and represent a single authoritative government perspective. The principle of re-use, where information is created once and is available to be used for different purposes with confidence, is fundamental;
6. Promote trust and confidence, rights and responsibilities: The ethical use of information is paramount. Information management practices should be transparent, respect rights and enforce responsibilities. Access to and use of information should promote trust and confidence through adherence to privacy, confidentiality, and IP and security requirements; and
7. Achieve a net social benefit: A net social benefit should be derived from whole-of-government and agency-specific information holdings. This should reflect a balance between compliance and service delivery and satisfy the important goals of service improvement and value creation.

⁶⁹³ <http://www.agimo.gov.au/__data/assets/pdf_file/50725/Information_Interoperability_Framework.pdf>.

⁶⁹⁴ Ibid 17.

Productivity Commission Review of Access to and Pricing of Government Information 2002⁶⁹⁵

8.109 The Productivity Commission Review proposed charging for some forms of data, which is not generally appropriate for an institutional repository. However, some of the issues canvassed about cost recovery are relevant, especially the recommendation that the cost structures must be reviewed periodically:

Such activities include:

- Regularly collecting and assessing information to check compliance with standards, such as the Australian Prudential Regulation Authority's regular reviews of the financial statements of institutions to monitor their liquidity levels;
- Conducting random audits to monitor compliance, such as the Civil Aviation Safety Authority's monitoring of airlines' compliance with aircraft safety standards; and
- Managing complaints handling mechanisms, such as the Australian Securities and Investments Commission's oversight of the operation of the Superannuation Complaints Tribunal.⁶⁹⁶

8.110 The review should be extended as a principle to all policies and activities of an institutional repository. Standards compliance and complaint investigation and resolution policies must be implemented.

8.111 The review highlights that legal authority is necessary for any charges to have validity, accountability and transparency.

Prime Minister's Science, Engineering and Innovation Council (PMSEIC) Report (2006)

8.112 The PMSEIC Report considered issues of access to and management of scientific data in government, universities and research institutions and centres. The PMSEIC Report is also relevant to funders of scientific research that produces data.

8.113 In the Report, the PMSEIC Working Group recommends the implementation of national strategic frameworks for scientific data and associated digital repositories, in order to facilitate sharing and collaboration:

Among the working group's recommendations are that:

- Australia's government, science, research and business communities establish a nationally supported long-term strategic framework for scientific data management, including guiding principles, policies, best practices and infrastructure;
- the necessary policy and programmes be implemented with a view to establishing a sustainable publicly-funded national network of 'federated' digital repositories;
- the principle of open equitable access to publicly-funded scientific data be adopted wherever possible and that this principle be taken into consideration in the development of data for science policy and programmes. As part of this strategy, and to enable current and future data and information resources to be shared, mechanisms to enable the discovery of, and access to, data and information resources must be encouraged; and
- funding agencies offer incentives to encourage researchers and institutions to:
 - develop data management plans for each research grant application involving data collection and generation, and that standards be made freely available and widely disseminated so as to encourage best practice in data management;

⁶⁹⁵ <<http://www.pc.gov.au/inquiry/costrecovery/finalreport/index.html>>.

⁶⁹⁶ 'Proposed regulatory agency guidelines' page 15
<<http://www.pc.gov.au/inquiry/costrecovery/finalreport/costrecovery2.pdf>>.

- introduce policies and practices to encourage collaboration and sharing of data across Australia's scientific research institutions and across agencies; and
- analyse and re-use existing data.⁶⁹⁷

National Collaborative Research Infrastructure Strategy (NCRIS) - Platforms for Collaboration

8.114 NCRIS was established by the Australian Government in 2004 as part of the initiative, *Backing Australia's Ability – Building our Future through Science and Innovation*.⁶⁹⁸ NCRIS aims to provide researchers with world-class research facilities and supporting infrastructure and networks.⁶⁹⁹

8.115 One of the so-called “capabilities” of NCRIS is the Platforms for Collaboration.⁷⁰⁰ The intent is to provide dynamic and evolving technological and management platforms to enhance the quantity, quality and productivity of research.⁷⁰¹

8.116 Platforms for Collaboration include a series of “inter-related components”:

- Data storage management, access, discovery and curation to improve interaction and collaboration;
- Grid enabled technologies and infrastructure to enable seamless access to the facilities and services required in various research fields;
- Support skills to assist researchers in developing and using this infrastructure effectively;
- High performance computing to allow analysis, modelling and simulation; and
- High quality network access through high capacity bandwidth to permit interaction with diverse data and computing resources.⁷⁰²

8.117 The NCRIS Strategic Roadmap, issued in February 2006, contains the following statement of principle:

Ideally, investment in platforms for collaboration should provide researchers with the ability to: gain access to information relevant to their field from a variety of sources seamlessly; exchange information collaboratively with colleagues; annotate their datasets or publications; and to manage and disseminate the results of their research through supported repositories.⁷⁰³

8.118 In April 2007, the NCRIS Committee accepted the Investment Plan prepared for the Platforms for Collaboration.⁷⁰⁴ The Investment Plan addressed the formation of an Australian e-Research Infrastructure Council to develop nationwide

⁶⁹⁷ PMSEIC Working Group on Data for Science (2006) 11-12 cited Productivity Commission, *Public Support for Science and Innovation*, Research Report (2007) 230.

⁶⁹⁸ See <<http://www.ncris.dest.gov.au/>> at 31 May 2007

⁶⁹⁹ Ibid

⁷⁰⁰ See <http://www.ncris.dest.gov.au/capabilities/collaborative_investment_plan_platforms.htm> at 31 May 2007

⁷⁰¹ Ibid

⁷⁰² National Collaborative Research Infrastructure Strategy (NCRIS) Strategic Roadmap, February 2006, available at <<http://www.ncris.dest.gov.au/NR/rdonlyres/91C5DFB3-10E5-4A09-A861-6973B2912417/9519/NCRISStrategicRoadmap.pdf>> at 31 May 2007, p49

⁷⁰³ National Collaborative Research Infrastructure Strategy (NCRIS) Strategic Roadmap, February 2006, available at <<http://www.ncris.dest.gov.au/NR/rdonlyres/91C5DFB3-10E5-4A09-A861-6973B2912417/9519/NCRISStrategicRoadmap.pdf>> at 31 May 2007, p49

⁷⁰⁴ See <http://www.ncris.dest.gov.au/capabilities/collaborative_investment_plan_platforms.htm> at 31 May 2007

infrastructure for the strategic motivation and promotion of e-research.⁷⁰⁵ This infrastructure will include:

- national data management infrastructure to improve management and use of research data;
- national high end computation facilities; and
- collaboration infrastructure for interoperable national services across shared research resources.⁷⁰⁶

4. Private Sector Organisations

Wellcome Trust Position Statements (2003)

8.119 The Wellcome Trust is a UK independent charity, established in 1936, with the aim of fostering and promoting research to improve human and animal health. It is the UK's largest non-governmental source of funds for biomedical research,⁷⁰⁷ awarding grants for research in the areas of human and animal health and biology.

8.120 It is a condition of Wellcome Trust research funding grants that:

A copy of the final manuscripts of all research papers supported in whole or in part by the Grant must be deposited into PubMed Central (or UK PubMed Central once established) upon acceptance for publication, to be made freely available as soon as possible and in any event within six months of the journal publisher's official date of final publication.⁷⁰⁸

8.121 PubMed Central is a database operated by the National Center for Biotechnology Information within the US National Library of Medicine which provides free access to biomedical and life science journal literature.

8.122 The grant condition requiring deposit of research papers in PubMed Central is supported by the Wellcome Trust Position Statement in Support of Open and Unrestricted Access to Published Research. The Position Statement provides:

With recent advances in internet publishing, the Wellcome Trust seeks to encourage initiatives that broaden the range of opportunities for quality research to be widely disseminated and freely accessed. The Wellcome Trust therefore supports unrestricted access to the published output of research as a fundamental part of its charitable mission and a public benefit to be encouraged wherever possible.⁷⁰⁹

8.123 As well imposing an obligation on researchers to deposit their research papers, the Wellcome Trust Position Statement states that the Wellcome Trust:

Expects authors of research papers to maximise the opportunities to make their results available for free and, where possible, to retain their copyright.⁷¹⁰

⁷⁰⁵ Ibid, see also *NCRIS Final Investment Plan for the Research Capability – Platforms for Collaboration*, available at <<http://www.ncris.dest.gov.au/NR/rdonlyres/2DFC56E4-83B4-4CFC-BC74-1CA4145A2657/16382/PfCInvestmentPlanFinalwebversion1.pdf>> at 31 May 2007, p6

⁷⁰⁶ Ibid

⁷⁰⁷ For more information see <<http://www.wellcome.ac.uk/aboutus/>>.

⁷⁰⁸ <<http://www.wellcome.ac.uk/assets/wtx026668.pdf>>.

⁷⁰⁹ <http://www.wellcome.ac.uk/doc_WTD002766.html>.

⁷¹⁰ Ibid.

8.124 The Wellcome Trust, as a major funder of biomedical research including the Human Genome Project, has taken a keen interest in the accessibility of biomedical data. In support of the Human Genome Project, and in the hope that it will lead to advances in healthcare, the Wellcome Trust has formulated a statement called the Wellcome Trust Statement on Genome Data Release, which aims to make genome sequence data freely available as soon as possible.⁷¹¹ The Wellcome Trust Statement on Genome Data Release endorses the Bermuda Principles and states:

These principles should apply for all human genomic sequence generated by large-scale sequencing centres, funded for the public good, in order to prevent such centres establishing a privileged position in the exploitation and control of human sequence information.⁷¹²

8.125 At a meeting sponsored by the Wellcome Trust in January 2003, the Bermuda Principles were reaffirmed and the attendees recommended that they should be extended to other non-DNA sequence data.⁷¹³ In particular, the meeting concentrated on data from “community resource projects,” being research projects that are:

specifically devised and implemented to create a set of data, reagents or other material whose primary utility will be as a resource for the broad scientific community.⁷¹⁴

8.126 Examples of community resource projects include the International Human Genome Sequencing Consortium, the Mouse Genome Sequencing Consortium, the Mammalian Gene Collection, the SNP Consortium, and the International HapMap Project.⁷¹⁵ Many large-scale datasets and collections of materials are produced as community resources and are becoming increasingly important as the drivers of progress in biomedical research. It has been argued that widespread data availability is the quid pro quo for the large sums of money granted to major research centres to complete these large-scale projects, without having to undergo peer review at each stage of funding.⁷¹⁶

8.127 The 2003 Wellcome Trust meeting considered that the scientific community would be best served if the results of community research projects (for example, large-scale protein structure determination or gene expression analysis) were made immediately available for free and unrestricted use by the scientific community to enable researchers to engage in “creative science.”⁷¹⁷ Consequently, the meeting attendees recommended that the principle of rapid pre-publication release of data should apply to data produced by community resource projects. The Wellcome Trust

⁷¹¹ <http://www.wellcome.ac.uk/doc_WTD002751.html>.

⁷¹² Ibid.

⁷¹³ The Wellcome Trust, ‘Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility’ (Report of a meeting organised by the Wellcome Trust, Fort Lauderdale USA, 14-15 January 2003) <<http://www.wellcome.ac.uk/assets/wtd003207.pdf>>.

⁷¹⁴ Ibid.

⁷¹⁵ Ibid. For more detailed descriptions of these community resource projects and databases, see Chapter 4.

⁷¹⁶ See Rebecca Eisenberg and Arti Rai, “Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California’s Stem Cell Initiative” (2006) 21 Berkeley L.J. 1187 at 1200

⁷¹⁷ The Wellcome Trust, *Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility*, Report of a meeting organised by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA, available at <<http://www.wellcome.ac.uk/assets/wtd003207.pdf>>.

meeting called on funding agencies, as the major sources of support for research projects leading to community resources, to require, as a condition of funding, free and unrestricted release of research data from community resource projects to appropriate central and searchable public databases. Funding bodies should encourage more researchers to serve the community through involvement in community resource projects and ensure that researchers engaged in the preparation of public databases have sufficient support for curation, maintenance and distribution of the data to the community, as well as resources to perform initial analyses using the resources they have generated.⁷¹⁸

8.128 However, aside from community resource projects, much valuable data is produced by research efforts whose primary goal is not resource generation. In such cases, contribution of data to public resource databases is voluntary. Attendees at the Wellcome Trust meeting acknowledged that pre-publication release of data might conflict with the fundamental scientific incentive of being able to publish the first analysis of data that has been generated in the course of research. While this incentive cannot be absolutely guaranteed without applying restrictions that would undermine the rationale for rapid, unrestricted release of data from community resources, to continue to encourage good scientists to work on these projects the scientific community must understand that pre-publication data release requires active, broad-ranging support. Consequently, the contributions and interests of large-scale data producers should be recognised and respected by the users of data, and the ability of data production centres to analyse and publish their own data should be supported by their funding agencies. To obtain the benefits that flow from making research data available in public databases, incentives should be developed by the scientific community to support the voluntary release of data, by recognising and protecting the interests of scientists who wish to share pre-publication data with the community.

Novartis

8.129 In February 2007, Novartis announced that it would provide free and open access on the internet to its genetic research data on type 2 diabetes. Novartis' research has helped uncover, among the 20,000 genes identified by the Human Genome Project, which genes are likely to be associated with diabetes. As it takes enormous resources to interpret research data, Novartis holds the view that it stands to gain more from providing open access rather than holding the research data secret.⁷¹⁹

8.130 In partnership with researchers at Sweden's Lund University and the Broad Institute at Cambridge University, Novartis compared the genomes of 1,500 people who had diabetes with 1,500 who did not, using gene chips that allowed them to track 500,000 places in the genetic code where previous research had shown that there are likely to be differences. The result is a scientific library of genetic differences that are likely to increase a patient's risk of diabetes.

⁷¹⁸ Ibid.

⁷¹⁹ Matthew Herper and Robert Langreth, *Biology Goes Open Source*, *Forbes.com* 12 February 2007, <http://www.forbes.com/2007/02/12/novartis-genes-diabetes-research-biz-cz_mh_0212novartis_print.html>.

8.131 Since the researchers do not know what most of these genes do or why diabetics are more likely to have them, Novartis has decided to release all of its raw data (but not three years' worth of analyses of the data) on the Internet. Novartis acknowledges that translating the study's identification of diabetes-related genes into the invention of new treatments will require a global effort. By making its raw data freely available Novartis hopes to tap into a global research community to dramatically scale and speed up its early-stage Research and Development activities.⁷²⁰

University and Industry Innovation Summit (2005): Open Collaboration Principles

8.132 The University and Industry Innovation Summit held in August 2005 and sponsored by IBM and the Ewing Marion Kauffman Foundation (a private foundation focusing on the advancement of innovation and entrepreneurship), brought together leaders from the IT industry, academics and government to discuss a variety of research relationships and complex IP practices that affect innovation. Summit participants considered collaborative innovation between industry and universities in the fields of computer science, IT software and IT services, with the aim of accelerating collaborative research for open source software.

8.133 The result of the Summit was the Open Collaboration Principles which provide a foundation for addressing IP created through open source software research. Although the principles are specific to software, they are instructive in the development of principles for open access to data.⁷²¹ The Open Collaboration principles are:⁷²²

- Open access should apply to material for both commercial and academic use;⁷²³
- Where it is necessary to have access to a patent in order to use a contribution, the patent should be licensed to the world at large for no charge; and
- There should be a requirement that persons taking data from a repository are not allowed to use the data in ways or for means that are contrary to open access principles or the Open Collaboration Project.⁷²⁴ For example, research data should not be taken from a repository to support a patent claim which

⁷²⁰ Don Tapscott and Anthony Williams, 'The New Science of Sharing', *BusinessWeek.com* 2 March 2007,

<http://www.businessweek.com/innovate/content/mar2007/id20070302_219704.htm?chan=technology_technology+index+page_more+of+today's+top+stories> at 12 March 2007.

⁷²¹ <http://www-304.ibm.com/jct09002c/university/scholars/news/open_collaboration.html>.

⁷²² The *Open Collaboration Principles* document with explanatory footnotes is available at <http://www-304.ibm.com/jct09002c/university/scholars/news/open_collaboration.html> and <www.kauffman.org/items.cfm?itemID=662>.

⁷²³ Page 2

<http://www-304.ibm.com/jct09002c/university/scholars/downloads/Open_Collaboration_Principles_overview.pdf>.

⁷²⁴ <http://www-304.ibm.com/jct09002c/university/scholars/downloads/Open_Collaboration_Principles_overview.pdf>.

would negatively impact on the ability of others to use the repository or data in the repository for innovation.

KEY POINTS

Development of a policy and principles for management of research data should be based upon and informed by a thorough understanding of the policies, principles and guidelines on open access to data and information that have been formulated in Australia and internationally.

These existing policies, principles and guidelines, which operate at the international, national and organisational level, have been developed by a broad range of entities, including:

- *International organisations – among these policy statements are the Bermuda Principles, the Budapest Open Access Initiative, the Berlin Declaration on Open Access to Knowledge in the Science and the Humanities, the World Summit on the Information Society (WSIS) Declaration of Principles, and the Organisation for Economic Co-operation and Development (OECD) Declaration of Access to Research Data from Public Funding;*
- *Governments and public sector research funding bodies – including the National Institutes of Health (NIH) Data Sharing Policy, the European Union's Directive on the re-use of public sector information, the Australian Research Council (ARC) and the National Health and Medical Research Council (NHRMC) funding policies, and the Office of Spatial Data Management's (OSDM) Spatial Data Access and Pricing Policy; and*
- *Private sector organisations – such as the Wellcome Trust Position Statement.*

The policies developed and adopted by these international, public sector and private sector organisations can inform practices and protocols regarding ownership, use, access and control of research data. They will be particularly relevant to the development of legal frameworks governing access to and reuse of data in research databases (see Chapter 4). When examining the existing policies, particular attention should be paid to those relating to specific fields of research activity, for example the policies relating to genome and genetic data.

CHAPTER 9 - DATA SHARING INFRASTRUCTURE

“The scientific community can and must assert greater control over the management of its own data supplies.”⁷²⁵

⁷²⁵ J Reichmann and P Uhler, ‘A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment’ (2003) 66 *Law and Contemporary Problems* 315, 324
<<http://heinonline.org.ezp02.library.qut.edu.au/HOL/Page?handle=hein.journals/lcp66&id=323&collection=top30&index=journals/lcp>>.

DEVELOPING DATA SHARING INFRASTRUCTURE – PUTTING IN PLACE THE LEGAL FRAMEWORK

Aims:

1. Examine how data can be practically managed in the Australian research sector, taking into account legal and regulatory frameworks for data management;
2. Describe how data access policies and principles can be formulated and why they are important; and
3. Consider the matters that should be addressed in a data management plan and a data management toolkit and how they can be developed to accord with current practices and policies.

1. Introduction

9.01 This chapter considers how researchers can be assisted in managing the legal issues inherent in data sharing and proposes that the following steps be taken:

- formulate a data access policy and principles;
- identify the specific kinds of data to be made available for access and use;
- ascertain conditions for the access and use of data under contract and licence; and
- adopt mechanisms for practical data management, such as a data management plan for database managers and a database management toolkit for researchers.

2. Formulate a Data Access Policy and Principles

9.02 Research collaboration is directly effected by legal requirements and informal procedures. Open collaborative research and open access can best be achieved through:

Development of a network of repositories and services, replication and collaboration between them, longer-term funding frameworks and definition of different types of repository, roles, and responsibilities over the lifecycle of research information.⁷²⁶

9.03 Paul David has observed that it is important for institutional arrangements to:

Minimize the extent to which the law becomes an impediment to cooperation among researchers, whether directly or indirectly by undermining informal mechanisms of trust and

⁷²⁶ Neil Beagrie, *e-Infrastructure Strategy for Research: Final Report from the OSI Preservation and Curation Working Group* (2007) 11.

dispute resolution. The challenge in designing appropriate legal arrangements for collaborative e-Science is ...to construct agreements that are adequately clear and determinative without damaging the trust and informal norms essential to the day-to-day conduct of collaborative research; and to provide processes for constructing those agreements that involve the scientists without unduly burdening them with negotiations over legal complexities.⁷²⁷

9.04 The APSR report, *Sustainable paths for Data-intensive Research Communities at the University of Melbourne* (2006), found that:

There is a lack of best practice guidelines and policy statements available to support researchers with their data management and decision making processes. The lack of shared language and terminology around many aspects of data and its management suggests the importance for all policies and guidelines to include clear definitions of concepts and terms used.

Areas of need include:

- Implementation of research record keeping principles and requirements.
- Data management for short terms sustainability and long term preservation.
- Metadata standards, principles and systems:
 - Across the discipline divide.
 - For raw and processed research data.
 - For web presentations.
 - For other scholarly works.
- Authentication and authorisation standards and systems for access and storage of scholarly IP.⁷²⁸

9.05 A fundamental step in the development of effective infrastructure for data management is the formulation of a clear policy and principles on data access and reuse. The data management policy should contain:

- high level statements about the approach to be adopted in relation to making data available for access and use; and
- principles expanding upon the high level policy statements and indicating how they are to be applied.

9.06 The development of data management policies and principles will be useful to a range of entities, including funding bodies, research institutes, participants in scientific collaborations, industry bodies and private sector research organisations.

9.07 Data can be generated in a range of different contexts, including:

- major scientific collaborations involving national and international coordination of research, management and data infrastructure;
- public and private sector research organisations and institutes; and

⁷²⁷ Paul A David, *Towards a Cyberinfrastructure for Enhanced Scientific Collaboration: Providing Its 'Soft' Foundations May Be The Hardest Part*, Oxford Internet Institute Research Report No.4 (2004) 14, 16.

⁷²⁸ Anna Shadbolt, Dirk van der Kniff, Eve Young and Lyle Winton, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories*, August 21 2006, <http://www.apsr.edu.au/aeres/sustainable_paths.pdf> at 18 June 2007, p38-39

- health research communities collecting medical and health-related patient data.

9.08 The development of a data management policy and principles should occur in the context of existing policies and principles (described in Chapter 8) and be based on knowledge about current practices and researchers' attitudes to data sharing (described in Chapter 4). Consideration needs to be given to formulating a policy which ensures that researchers' objectives, needs and responsibilities in each research situation are appropriately addressed.

9.09 A range of factors will determine the data access policies and principles adopted by various research groups:

- the research discipline and special practices in that discipline;
- the kind of data generated;
- how the data is to be deposited and the time at which deposit is to occur;
- the basis on which the data is to be made available for access by other researchers;⁷²⁹
- the funding arrangements for the research;
- the legal obligations imposed on the researcher or research group; and
- time limits for the public release of data under contractual or confidentiality obligations.

9.10 For many major government or publicly funded research projects (such as the genomic community research projects), policy and practice strongly support immediate and open access to research data. Often, these policies will reflect public interest in publicly funded research data being made freely available to the public that has funded the research.

9.11 However, different considerations arise in situations involving private sector research organisations or funding bodies. Eisenberg and Rai caution that a public domain approach to data access:

Imposes some costs of its own [and that] aggressive versions of a public domain approach may undermine the types of small firms that tend to provide specialised research inputs in the marketplace.⁷³⁰

9.12 Where significant funding has been provided by private sector collaborators, the policy will need to take into account the private sector partner's interest in

⁷²⁹ See further below.

⁷³⁰ R Eisenberg and A Rai, "Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California's Stem Cell Initiative" (2006) 21 Berkeley L.J. 1187 at 1202.

recouping its investment.⁷³¹ Access restrictions may be imposed at least until the private collaborators can recover their expenditure through commercial application of the research outcomes.⁷³²

9.13 Policy and principles may apply differently to various kinds of research data and information. While certain kinds of research data will be made available for indefinite access, other data will only be made accessible for a limited period of time.

9.14 The policy must state how it is to be given effect. For example, a government-funded research project may have a policy that states that research data must be deposited into an open access repository in accordance with a data management and sharing plan prepared and submitted with the funding application. This requirement is increasingly common with large, government-funded community resource projects.⁷³³

9.15 In view of the powerful factors that cause researchers to withhold access to data, the policy needs to be based on an understanding of incentives to encourage researchers to comply with objectives and expectations relating to data access and sharing. The importance of incentives has been explained by Eisenberg and Rai:

[M]ost rewards in research science, including academic appointments, promotion and grant funding, depend on a record of frequent publication. Scientists may perceive sharing data, even after an initial publication, as providing advantages to competitors in the race to generate further publications. Scientists may also be reluctant to share data because of involvement in commercial activities. Sharing may imperil patent applications or destroy trade secrecy. Emerging evidence reveals that some research communities in the life sciences are reluctant to share data even after publication.⁷³⁴

9.16 It is important that raw data and information are, in themselves, valued as reliable scientific resources whether or not they have been published in a journal. This may require not only the establishment of database infrastructure, but a change in researchers' attitudes towards the importance of sharing and preserving data collected during the course of research:

Critical for the exchange and sharing of data is not only the infrastructure for storage but also the cultural change which will reward the deposit of data in appropriate repositories and recognise the impact of this data. Currently there is no universal standard for citing data. Indeed, citing unpublished data (data that has not been published in the scientific literature) in the references, even when electronically archived and made available through the Internet, is not allowed by some journals.⁷³⁵

⁷³¹ For example, under the GAIN project, contributing researchers are given a period of nine months during which they have exclusive rights to submit publications based on the data they have contributed.

⁷³² See Productivity Commission, *Draft Report on Public Support for Science and Innovation*, (2006) [5.11].

⁷³³ The data produced by such projects is what NCRIS describes as 'data derived from significant public or shared investment' (as opposed to 'private' data which is usually held by a researcher through self generation, or by acquisition from surveys, an instrument or device").

⁷³⁴ See Rebecca Eisenberg and Arti Rai, "Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California's Stem Cell Initiative" (2006) 21 Berkeley L.J. 1187 at 1200.

⁷³⁵ Data Information Creation Working Group, *20/20 Vision: an e-Infrastructure for the Next Decade*, Report of the Data and Information creation Working Group to the e-Infrastructure Steering Group (2006) 17.

9.17 Determining incentives can be difficult and complex, but clear incentives for contribution are essential if data release policies are to be effectively implemented. Incentives for individual researchers and groups will vary according to context, but may include:

- pre-publication rights for researchers who contribute data to open access databases prior to publication of their analysis of the data;
- attribution of contributors' data deposits in a citable format which enables the sources of data to be referenced, as proposed by the Wellcome Trust in *Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility* (2003)⁷³⁶;
- acknowledgement of researchers whose data is frequently downloaded from a database;
- rewards for compliance, for example, privileged access to data analysis tools for those who contribute to a database;
- sanctions for non-compliance, for example loss of continued funding;
- opportunities to build collaborative relationships with other researchers, research institutions, funding bodies and commercial partners within a field of research; and
- normative pressures to comply with data disclosure requirements, for example the strong pressures to rapidly disclose data in the Human Genome Project.

3. Identify Data to be Made Available for Access and Use

9.18 Having formulated a data access policy and principles, it is necessary to identify the kinds of research data and information to which it will apply.⁷³⁷

9.19 Identification of specific data and datasets to be deposited in accordance with the data access policy and principles must be based on:

- a practical understanding of the legal context in which the data is collected or generated;
- any legal or administrative constraints that may apply to dissemination and reuse of the data; and

⁷³⁶ The Wellcome Trust, *Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility*, Report of a meeting organised by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA, available at <http://www.wellcome.ac.uk/assets/wtd003207.pdf>. See further in Chapter 8.

⁷³⁷ See Rebecca Eisenberg and Arti Rai, 'Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California's Stem Cell Initiative' (2006) 21 *Berkeley Law Journal* 1187.

- any relevant technical requirements that impact upon the kinds of data that can be made available, such as data, metadata and formatting standards nominated by the data custodian.

9.20 Each research organisation should develop a clear and comprehensive listing of all restrictions applying to the management, dissemination and reuse of the different kinds of data that may be generated and used.

9.21 Not all data generated in a research project will be suitable for access and reuse, due to privacy, confidentiality, commercial or legal restrictions. It is also necessary to consider any IP rights (copyright and patents) that may exist in the data, datasets, or database.

9.22 Researchers must identify any data received and the basis on which they have been provided with access to that data. Researchers need to ensure that in redistributing the data or re-using the data to create new datasets:

- they are not in breach of the terms of any licence under which they have been provided with the data (bearing in mind that the licence may relate to both copyright and other IP rights and may also impose additional contractual restrictions, usually relating to commercial use of the data); and
- they are not infringing any other legal obligations, such as privacy obligations imposed by legislation or administrative standards.

9.23 Where IP rights are licensed, the terms and conditions of the licence must be sufficiently extensive to enable the research organisation to carry out its intended activities. This may include the ability to distribute the data to other researchers, which may involve depositing the data in an institutional or disciplinary repository. If there are no administrative or legal prohibitions or restraints, data can be made available in accordance with the relevant data access policy and principles.

4. Ascertaining the Conditions of Access and Use

9.24 Researchers operating under an open access regime want to contribute their data into an institutional or disciplinary repository so that it is available to other researchers for reuse in their projects. To be able to make their research data available for open access, the researchers must:

- (a) have all necessary rights or permissions to be able to deposit the data or datasets into a repository; and
- (b) be able to grant to the database repository manager or end users all appropriate rights and powers to deal with the data.

The grant of rights to the database repository manager and end users will typically occur under a Repository Deposit Licence (or Agreement).

9.25 At this stage, the following matters need to be considered:

- the original sources of data (where the research data is based on data derived from another source);

- the rights in the data held by the contributing researcher or research organisation;
- all other ownership rights in the data;
- the terms under which the researcher or organisation is licensing and depositing the research data into a repository;⁷³⁸
- the rights to access and reuse the data granted to other researchers; and
- whether different access and use rights are to be granted to different researchers or research groups and whether different restrictions, limitations or conditions are to apply to access to and reuse of data according to the purpose for which it is to be used.

9.26 If an open access approach is adopted, data may be made available without any restrictions imposed on who can obtain access, what data can be accessed or how it can be used.

9.27 In other cases, restrictions may apply to data access and reuse. Common restrictions are:

- access is limited to certain categories of researchers (for example, non-profit, publicly funded researchers or researchers engaged in a particular field of research);⁷³⁹
- the data is to be used for a specified purpose (usually a select research project) only;
- any use of the data or any publications referring to the data must attribute the researchers who produced the data and the body that provided funding for the research project;
- the data can be used for non-commercial purposes only and any commercial use of the data requires the permission or involvement of the data owners or the researchers who produced the data; and

⁷³⁸ See R Eisenberg and A Rai, *Harnessing and Sharing the Benefits of State-Sponsored Research: IP Rights and Data Sharing in California's Stem Cell Initiative* (2006) 21 Berkeley L.J. 1187.

⁷³⁹ Eisenberg and Rai, *ibid* at 1206, describe a possible tiered-access approach to data access in the context of the California Stem Cell Initiative (CIRM): "[CIRM] might, for example, permit access by: (1) CIRM-funded nonprofit researchers only; (2) all CIRM-funded research-owners; (3) all California researchers; (4) all stem cell researchers who had contributed their own data (and/or agreed to contribute their own annotations/improvements to the database); or (5) all stem cell researchers. Certain categories of researchers could be excluded altogether or could be given access under restrictive conditions. CIRM could require for-profit organisations, or non-California institutions, to pay for access. Non-price methods of tiering, such as early access by certain favoured categories of researchers, could favour preferred groups while still permitting broad access."

- the data should not be the subject of a restrictive patent application or be used in such a way as to restrict future researchers' access to the data.

5. Creative Commons Licenses

9.28 Creative Commons open content licences may operate where:

- open access to data is supported by an organisation's data access policy; and
- the data or datasets are protected by copyright.

9.29 Creative Commons licences grant the right to copy, distribute, display, digitally perform and make verbatim copies of a work into another format, subject to certain conditions as imposed by the copyright owner. All Creative Commons licences contain the condition that the copyright owner must be attributed. Other conditions may include:

- the uses of the work must be non-commercial;
- only exact copies of the work may be made (no derivatives); or
- the work (or any derivatives) may only be distributed under a Creative Commons licence identical to the one attaching to the original work.

A more detailed explanation of Creative Commons licences is provided in Chapter 5.

6. Adopting Mechanisms for Practical Data Management

9.30 Researchers need assistance in managing the legal issues that are inherent in generating, accessing and using research data. The objective is to design mechanisms for use in the research environment which ensure that these legal issues become integrated into standard operational data management practices.

9.31 The kind of mechanisms that provide practical guidance for researchers include:

- Data Management Plans (DMP); and
- Data Management Toolkits (DMT).

9.32 These mechanisms must cover the data management activities carried out at different stages of an e-Research project. The UK Office of Science and Innovation's e-Infrastructure Working Group, in its report *Developing the UK's e-Infrastructure for Science and Innovation* (2007), identified six stages of the data lifecycle:

- Data and information creation: The creation of digital data;
- Preservation and curation: Data management and preservation for long-term use;
- Search and navigation: The means to ensure the effective use of that information through highly complex search and navigation facilities;

- Virtual Research Communities: To ensure collaboration across institutional. Disciplinary and even national boundaries;
- Networks, computers and data storage: Massive network and computational capacity to support distributed ubiquitous and 24/7 access to resources of all kinds; and
- Authentication, authorisation and accounting, middleware and digital rights management: Services to ensure that access is safe, secure and legally compliant.⁷⁴⁰

7. Data Management Plans

9.33 It is clear that it is not appropriate to simply embark on research projects without first considering the issues of access to and use of data. These issues should be considered even before the data is generated. Where data access requirements are imposed by research funding bodies or under institutional policies and principles, arrangements for data management should be included in funding proposals.

9.34 For example, the Medical Research Council (MRC), a large funding body in the UK, states in its Guidance of Open Access to Published Research (updated 1 October 2006) that:

From 1 January 2006, all applicants submitting funding proposals to the MRC are expected to include a statement explaining their strategy for data preservation and sharing. MRC data sharing policy indicates that, where possible, published results should provide links to the associated data.⁷⁴¹

9.35 An effective Data Management Plan (DMP) based on an understanding of existing database models and data management schemes for the storage and management of data should be in place from the conception and commencement of a research project.⁷⁴² A DMP must address technological requirements and how data is collected or generated in the legal context of the research operations and should describe:

- how data is generated or collected;
- how data is integrated and stored in the database;
- data ownership and legal controls;
- how data will be managed and disseminated;

⁷⁴⁰ UK Office of Science and Innovation (OSI) e-Infrastructure Working Group, *Developing the UK's e-Infrastructure for Science and Innovation* <<http://www.nesc.ac.uk/documents/OSI/report.pdf>> at February 2007.

⁷⁴¹ UK Medical Research Council, *MRC Guidance on Open Access to Published Research* (2006) <<http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc002548.pdf>> at 21 March 2007.

⁷⁴² See Chapter 4 and Chapter 8. See also Alan Tonge and Peter Morgan, *Project SPECTRA: Submission, Preservation and Exposure of Chemistry Teaching and Research Data*, JISC Final Report (2007) <http://www.lib.cam.ac.uk/spectra/documents/SPECTRA_Final_Report_v10.doc> who state: '[T]he ownership of scientific data and licensing arrangements for data re-use need clear guidelines that can be adopted uniformly across the research community to ensure consistent practice.'

- regulatory mechanisms and controls that apply to the data;
- quality assurance and control measures, including provisions for the auditing of compliance with data management and access requirements;
- data security measures;
- how access will be provided to the data and how the data will be disseminated;
- the timeline for distribution of data;
- roles and responsibilities of researchers and database managers;
- database infrastructure;
- sustainability of data; and
- expenditures.

9.36 A DMP should relate to all parties involved in a research project, including:

- the creators of the data;
- the collectors of the data;
- the compilers of the data;
- the consumers or users of the data who will seek access to the database in which the data is stored;
- the parties that have funded the data collection; and
- the managers of the database or repository housing the data.

9.37 The rights, roles and responsibilities of the parties involved are numerous. Database managers have a particularly important role, being responsible for the organisation and maintenance of the database and for ensuring continued access to the data.⁷⁴³ The level of responsibility undertaken by a party, especially a database manager, may depend on the kind of data being collected and stored and the discipline in which the research is being undertaken. For example, where the data is personal or health information, an important responsibility of the database manager or other officer will be ensuring that privacy standards are maintained.

⁷⁴³ For an explanation of the responsibilities regarding preservation (ensuring continuing access to data) and curation (adding value to and maintaining digital research assets over time), see Neil Beagrie, *e-Infrastructure Strategy for Research: Final Report from the OSI Preservation and Curation Working Group* (2007) 5.

How Data is Generated or Collected

9.38 At the very outset of a research project, the scope of the kind of data required (as ascertained or capable of being ascertained at the commencement of the research project) and what is to be done with the data must be determined. This will affect how the data is to be collected and what protocols need to be followed during the data collection process.

9.39 It is also important to resolve how individual researchers will contribute data to the project and what incentives can be offered to encourage them to contribute.

9.40 A DMP should address:

- technical implications in terms of information technology requirements;
- the kind of data that will be collected or generated;
- what existing data is required from other sources;
- the research discipline and special practices in that discipline; and
- the various parties contributing to the data collection process.

9.41 Where data is collected under a contractual arrangement, such as an agreement between a research project and a funding organisation, it is important that the data is collected in accordance with the terms of the contract.

How Data is Integrated and Stored in the Database

9.42 Once data is collected, it should be carefully stored in a physical and/or digital database. The person or body with physical custody of the data will not always be the person or body formally responsible for the management of the data.

9.43 The DMP should describe:

- where the data is to be stored;
- who is responsible for the data in the database;
- arrangements for the integration of data that is collected or generated by the research project (original data); and
- arrangements for the integration of data that is acquired from other parties (non-original data).

9.44 Researchers should be able to identify non-original data included in the database and the terms under which they can use and further distribute the non-original data, so that legal obligations to other parties are not breached.

Data Ownership

9.45 Data may be owned by more than one party, especially where the research project generating the data is jointly funded. Parties may also claim rights in data even where they do not actually own the data. For example, a consumer who uses the data will have certain rights to possess and use the data, even though the data remains in the ownership of someone else:

The important point is that all parties who may have a claim to ownership or part ownership of a dataset should agree at the outset as to how it is to be exploited, and how the benefits are to be shared.⁷⁴⁴

9.46 Intellectual property rights are closely connected with ownership. An owner of data or data compilations may also have the right to apply for and obtain a patent for the data itself or for a patentable invention derived from the data. For copyright protected data, the copyright owner will hold a number of exclusive rights including the right to reproduce the data and the right to electronically communicate the data online. These rights can be assigned or licensed to other parties through contract. Contracts can also control access to and use of data, especially where the data is confidential data that has not yet been released into the public domain.

9.47 A DMP should define the owner of data as:

- the person who collects, creates or generates the data (the researcher);
- the employer of the researcher, under the terms of the researcher's employment contract;
- the funder of the research, under the terms of the funding agreement; or
- the database owner or provider, if ownership in the data is assigned upon inclusion in the database.

9.48 A DMP should highlight whether the data is jointly-owned by more than one party.

How Data will be Managed

9.49 Data must be carefully managed in accordance with legislative, administrative, contractual and legal requirements.

9.50 Legislative requirements must balance two public interests:

- protecting sensitive, personal and health information; and
- providing wide access to government and public information.

⁷⁴⁴ Natural Environment Research Council, *NERC Data Policy Handbook* (2002) [3.2] <http://badc.nerc.ac.uk/data/NERC_Handbookv2.2.pdf> at 21 March 2007.

9.51 Increasingly, strong arguments are being made for data derived from publicly funded research projects to be made freely available to the public that has funded the research.

9.52 Specific legislation may apply in relation to research projects operating in a particular field or area. For example, specific legislation will apply to environmental data, or where data is collected about children.

9.53 A DMP should explain:

- the kind of data dealt with in a research project; and
- whether legislation applies to restrict or allow access to the data.

9.54 Administrative requirements, usually in the form of government standards or codes (for example, the Queensland Information Standards), will also apply to supplement legislative requirements relating to the collection and use of data in some fields. The most important non-legislative controls imposed upon data management are:

- requirements stipulated by funders of research projects; and
- licensing arrangements concerning the release of data and datasets.

9.55 A DMP should describe:

- the conditions under which the research project is funded; and
- the proposed arrangements (both legal and technical) for providing access to data in accordance with the funding conditions.

9.56 Different funding arrangements will require different policies and principles to be applied to data management. Where publicly funded research projects are concerned, it is increasingly important for researchers to be able to explain how they will be managing and distributing their data, either under a DMP or otherwise.

9.57 There may be other considerations relevant to how data is managed, including cultural protocols and the general practices and procedures followed by researchers and data managers in the particular research discipline.

Regulatory Mechanisms and Controls

9.58 When collecting and managing data, researchers must follow any contractual restrictions or conditions imposed by funding organisations or other interested parties. Contractual terms may control how data is to be collected, stored and managed, and may also provide that access to and use of data is to be restricted in certain circumstances or for certain purposes.

9.59 A DMP should describe the contracts that apply to data collected or generated:

- confidentiality agreements for secret information;
- copyright assignments and licences, especially where the data is non-original data obtained from a third party;
- deposit agreements, where data is to be included in a database or digital repository; and
- access agreements to provide users with access to the data subject to certain conditions (that are usually imposed to protect commercial interests).

9.60 Where copyright is concerned, a DMP should address all the different licensing arrangements that are likely to affect the research project. Where non-original data has been acquired from a third party, it may have been provided under a licence that controls how the data can be used and redistributed. Licences will be important where copyright attaches to the data, so that outside researchers cannot reproduce the data without the copyright owner's permission.

9.61 A DMP should describe:

- the different licensing frameworks that may apply, including statutory licences and open source licences such as Creative Commons licences; and
- the licence conditions, to ensure that researchers are using the data conscientiously and not breaching the terms of the licence.

Quality Assurance and Control Measures

9.62 A DMP should describe:

- the data standards and quality assurance arrangements to be put in place for the research project;
- how quality is assessed;
- whether there is a minimum standard of quality that applied to the data included in the database;
- whether data must be authenticated or verified by other researchers or industry professionals before it is released to the public; and
- provisions for the regular auditing of the various parties' compliance with data management and access requirements.

Data Security Measures

9.63 Different levels of security may apply depending on:

- the type of data;
- the reason why it needs to be protected; and
- the level of protection required.

9.64 For confidential data, access may be completely blocked for all except a select number of researchers. For personal or health data, the data should be de-identified before it is able to be viewed. Where the data is protected by copyright or other intellectual property rights, restrictions may be placed on how the data is used and disseminated, through licensing conditions that must be agreed to before the data can be accessed.

9.65 A DMP should explain:

- the different levels of security applicable to the different levels of data;
- how the security measures are to be implemented; and
- the data that is:
 - secret data protected by the doctrine of confidentiality;
 - sensitive, personal or health information subject to privacy legislation;
 - subject to cultural protocols;
 - data that may form the basis of a patent application;
 - protected by copyright; and
 - data that researchers do not want to disclose for the time being, either for purposes of competition with other research groups or so that they can be the first to commercially exploit it.

9.66 Technological measures that can provide a level of security to data and its use (particularly to copyrighted data) are digital rights management (DRM) and electronic rights management information (ERMI). DRM uses technology to express rights in digital materials and regulate the exercise of those rights. Examples of DRM include encryption, digital watermarking and embedding rights management information. ERMI is electronic information that is attached to or embedded in the data. It can include the name of the owner of the data and the terms and conditions for use of the data. It also allows the data to be monitored and tracked, so that the owner can be informed of every instance of access and use.

9.67 A DMP must set out in detail the time limits for when data collected or generated in a research project must be released by. Data should be released as soon as possible. However, there may be situations where it is not appropriate to release data immediately after acquisition. These situations include where the data is subject to confidentiality restrictions or contractual restrictions, or where the researchers want

to delay release so that they can verify the data or be the first to publish the results of the research project.

9.68 A DMP should:

- allow for any restrictions on release that may apply in these situations;
- ensure that the length of the restrictions is appropriate in the circumstances; and
- explain the levels of access that will apply during the restriction period and after the restriction period has expired.

Roles and Responsibilities of Researchers and Database Managers

9.69 There will be many different roles and responsibilities within the entire data management process.

9.70 The level of responsibility assigned will depend upon the stage of the research project and who is involved at that particular stage. Researchers who have collected data may be responsible for its initial management and for depositing the data into the database. Responsibility may then shift to database managers, who will manage and maintain the database and monitor access and use.

9.71 Responsibilities will also be allocated according to each party's expertise, including their legal, technical and management skills and obligations and their field of research. For example, a medical science research project may make someone responsible for ensuring that personal and health data is de-identified and remains private.

9.72 A DMP should set out formal levels of responsibility for data and database management and maintenance for:

- researchers;
- curators and database managers; and
- other relevant parties.

Database Infrastructure

9.73 A DMP will need to describe the technological infrastructure of the database. The hardware and the software must be adapted to the kind of database being provided. It may be necessary to ensure interoperability between multiple databases, for associated projects or joint ventures.

9.74 The database must be sustainable in the long term and should be able to be updated regularly. Databases should:

- have data organised logically within the database;

- have appropriate controls imposed upon the data (where relevant); and
- have a system for inputting metadata, so that results are searchable.

Sustainability of Data

9.75 Emphasis is often placed on the long-term preservation of data that could be useful in future research projects, even where this usefulness is not contemplated when the data is originally collected. Careful consideration must be given to the potential future relevance of the data collected or generated by the research project. Some datasets will need to be accorded long-term stewardship, whereas other datasets will not. As expressed by the Natural Environment Research Council (NERC) in the UK:

In managing data a balance must be struck between the scale of the commitment needed (across the entire lifetime of the activity) and the anticipated benefits, both scientific and financial.⁷⁴⁵

9.76 Where the data and database are sustained long-term, it is important to consider funding. Provision should be made in the initial budget planning of the project and in grant applications for funds to sustain the ongoing management of the data and the database.

9.77 A DMP should describe:

- whether long-term preservation of the data collected by the research project is necessary;
- if so, how the data will be preserved;
- who will be responsible for the future preservation, maintenance and management of the data and the database; and
- the ongoing, long-term funding of the database (even after the research project that gave rise to the database is finished and no longer funded).

Expenditures

9.78 Management of expenses will be particularly important for long-term maintenance of the database and preservation of the data. Expenditures associated with data management will include the cost of developing a database, costs associated with acquiring existing datasets needed by the project and expenditure on the auditing of compliance with data management.

9.79 A DMP should anticipate these expenses and describe:

- how they will be managed;

⁷⁴⁵ Natural Environment Research Council, *NERC Data Policy Handbook* (2002) [4.2] <http://badc.nerc.ac.uk/data/NERC_Handbookv2.2.pdf> at 21 March 2007.

- any enforceable obligations relating to expenditure on data management; and
- mechanisms to periodically review the cost benefits of continuing to manage and maintain the data.

8. Data Management Toolkit

9.80 Researchers should be provided with practical guidelines for implementing a DMP in the form of a Data Management Toolkit (DMT).⁷⁴⁶

9.81 A DMT needs to be tailored to different levels of research, whether by an individual researcher, a research team within an institution, a collaborative distributed research group or an institution. A DMT should provide practical guidance to enable researchers to actively manage their research data, in compliance with the legal framework applying to their activities.

9.82 A DMT should enable researchers to understand:

- the data access policy and principles of the research institution;
- the research project's DMP;
- the legal and management issues addressed in the DMP and the legal context in which data is collected;
- ownership and licensed use of intellectual property rights;
- legal restraints applying to collection, storage, handling and use of data;
- technical restraints or requirements that may apply, such as compliance with metadata standards; and
- restraints arising from:
 - confidentiality agreements or obligations;
 - third party licensing arrangements;
 - privacy obligations (especially where the data is personal or health information);
 - policy or administrative measures; and

⁷⁴⁶ For example, JISC has developed a toolkit to address generic technological problems applicable to resource discovery, but does not address legal issues relating to data. See Alan Tonge and Peter Morgan, *Project SPECTRa: Submission, preservation and exposure of Chemistry Teaching and Research Data*, JISC Final Report (2007) 22
<http://www.lib.cam.ac.uk/spectra/documents/SPECTRa_Final_Report_v10.doc>.

- any other applicable statutory provisions.

9.83 The need for researchers to be provided with practical guidelines such as the DMT has been recognised in the UK by JISC. The March 2007 final report for the JISC-funded SPECTRa Project⁷⁴⁷ recommended that guidelines based on UK and EU legal practice be developed for intellectual property rights relating to scientific data. In particular, it was recommended that researchers should be provided with guidance on:

- ownership rights of individual researchers, their employers and the bodies funding their research;
- ownership issues arising at different stages in the life cycle of research data;
- intellectual property rights issues arising from collaborative work within a single institution, and between multiple institutions;
- intellectual property rights issues relating to composite research outputs such as databases;
- relationships between intellectual property rights for raw data and intellectual property rights for research papers that publish that data; and
- intellectual property rights relating to the licensing of research data for reuse.⁷⁴⁸

9.84 Additionally, in the recent report, *Sustainable paths for Data-intensive Research Communities at the University of Melbourne* (2006), APSR observes:

[As] research practices are rapidly adopting information and communications technology (ICT), researchers should be made aware of the services and expertise available to them; locally, nationally and globally. An awareness and basic understanding of research data policies, responsibilities, collections, curation, preservation, copyright/IP, metadata and standards must be included in a researcher and postgraduate induction program and reinforced throughout their candidature. An essential part of such a training program would include information about the terminology and underlying principles for managing data throughout its entire life cycle.⁷⁴⁹

⁷⁴⁷ Project SPECTRa: Submission, Preservation and Exposure of Chemistry Teaching and Research Data is a joint project between the libraries and chemistry departments of the University of Cambridge and Imperial College London, with Cambridge University Library as the lead institution, and is funded under the JISC Digital Repositories Programme.

⁷⁴⁸ Alan Tonge and Peter Morgan, *Project SPECTRa: Submission, Preservation and Exposure of Chemistry Teaching and Research Data*, JISC Final Report (2007) <http://www.lib.cam.ac.uk/spectra/documents/SPECTRa_Final_Report_v10.doc>.

⁷⁴⁹ Anna Shadbolt, Dirk van der Kniff, Eve Young and Lyle Winton, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories*, August 21 2006, <http://www.apsr.edu.au/aeres/sustainable_paths.pdf> at 18 June 2007, p40-41

KEY POINTS

In developing effective frameworks for the management of research data, the following steps should be taken:

(1) Formulate a Data Access Policy and Principles:

A Data Access Policy setting out the principles governing data access and reuse should be developed. The policy and principles may be developed at a governmental, institutional or research group level and should address a range of matters such as:

- *the research discipline/s to which the policy will apply;*
- *research funding arrangements;*
- *the kind of data that is collected or generated;*
- *the basis on which researchers will contribute data and incentives to encourage the timely contribution of data;*
- *rights held in relation to the data by the contributing researcher/s, research organisation or institution;*
- *who is permitted to access to the data and under what conditions;*
- *the legal obligations imposed on the institution, researcher/s, data users and other parties;*
- *any constraints applying to the distribution of data held in the database, including time limits for the public release of data under contractual or confidentiality obligations; and*
- *how the database or repository will be managed and how sustainability will be ensured.*

(2) Identify the Specific Kinds Data to be Made Available for Access and Use;

(3) Ascertain the Conditions of Access and Use; and

(4) Adopt Mechanisms for Practical Data Management:

Mechanisms should be developed to provide practical assistance to researchers in managing the data generated through their research projects:

Data Management Plans (DMPs) should be developed. A DMP covers a range of matters, including:

- *how data will be collected or generated in the legal context in which the*

research is conducted;

- *how data is integrated and stored in the database;*
- *data ownership and legal controls;*
- *how data will be managed;*
- *regulatory mechanisms and controls that apply to the data;*
- *quality assurance and control measures, including auditing of compliance with data management and access requirements;*
- *data security measures;*
- *how access will be provided to the data and how the data will be disseminated;*
- *the timeline for distribution of data;*
- *roles and responsibilities of researchers and database managers;*
- *database infrastructure;*
- *sustainability of data;*
- *the ongoing, long-term funding of successful research project databases, even after the project expires and the project is no longer funded;*
- *how researchers will be able to identify material or data received from third parties, and how they will ensure that they are not breaching licence terms in redistributing data; and*
- *where material can be made publicly available, whether open content licences, such as Creative Commons licences, should be used.*

Data Management Toolkits (DMTs) should be developed. A DMT covers a range of matters, including:

- *practical guidance for researchers on how to manage their research data in compliance with regulatory provisions applying to their field of research;*
- *how to develop, implement and comply with a data access policy and principles;*
- *the legal and management issues addressed in the DMP and the legal context in which data is collected;*
- *ownership and licensed use of intellectual property;*

- *legal restraints applying to collection, storage, handling and use of data;*
- *technical restraints or requirements that may apply; and*
- *restraints arising from:*
 - *confidentiality agreements or obligations;*
 - *third party licensing arrangements;*
 - *privacy obligations;*
 - *policy or administrative measures; and*
 - *any other applicable statutory provisions.*

CHAPTER 10 - CONCLUSIONS AND RECOMMENDATIONS

*“We are going to be deluged with data in almost every
field”⁷⁵⁰*

⁷⁵⁰ Tony Hey in Richard Poynder (2006) *A Conversation with Microsoft's Tony Hey*,
<<http://dialspace.dial.pipex.com/town/parade/df04/Hey.pdf>> at 24 May 2007

CONCLUSIONS AND RECOMMENDATIONS

Aims:

1. Consider the guidance required by individual researchers, research institutions and research funders in addressing legal aspects of data generation, management, dissemination and use;
2. Describe steps to be taken to develop effective data access policies and principles, model data management plans and model data management toolkits and template agreements;
3. Propose further work to be undertaken in analysing data access and sharing practices and open access policies, to develop a better understanding of the various strategies and mechanisms available for data access and use.

1. Conclusions

10.01 The collection, management and use of research data occurs in a legal context and raises a host of legal issues. Quite simply, data is surrounded by law. It is now widely acknowledged that if research data is to be effectively made available within the research community, it must be properly managed within an appropriate e-Research or e-Science infrastructure including protocols for dealing with the legal issues arising in relation to research data.⁷⁵¹ Increasingly, it is also appreciated that failure to establish legal protocols for data management is likely to jeopardise the ability to access and use valuable research outputs.

10.02 Failure to ensure that the legal framework effectively implements researchers' intentions regarding access to and use of research data gives rise to numerous risks, including:

- third-party data or data generated under a funding arrangement may not be able to be accessed or used because doing so would breach of copyright, contractual or other legal rights;
- data deposited by researchers into a database or digital repository may not be able to be accessed due to uncertainty about who owns or has rights in relation to the data, or uncertainty about the effect of legislative or other regulatory provisions;
- research may have to be postponed while the legal status of the data is investigated and ascertained;

⁷⁵¹ Christine Hine, *New Infrastructures for Knowledge Production: Understanding E-Science*, from the preface located online at <http://www.igi-pub.com/books/additional.asp?id=5558&title=Preface&col=preface> at 3 April 2007

- commencement of research projects may be delayed pending negotiations for data access with data owners; and
- research objectives may have to be changed or abandoned because satisfactory terms for data access and sharing cannot be negotiated.

10.03 Risks of the kind outlined above - and associated delays and costs – will be exacerbated in situations where a research project has been commenced without ensuring the appropriateness of the legal framework for dealing with the project's data outputs.

2. Key Points of the Report

10.04 This report has examined the legal context in which research data is generated, organised, managed, disseminated and used, identifying the key legal issues that must be addressed in developing legal frameworks for the effective management of research data. It emphasises the importance of developing and implementing appropriate legal frameworks to support open access availability and sharing of research data outputs among members of research communities.

10.05 This report recommends that researchers involved in e-Research should:

- be able to identify the legal rights that exist in relation to data, how and when they arise and how they can be allocated and contractually dealt with among the parties;
- adopt data policies and principles as the basis for legal frameworks for data management, access and use;
- implement legal frameworks for management of research data (including access to and use of the data), based on an understanding of the operation of relevant laws and regulations;
- use practical tools that guide and assist them in the management of research data (such as Data Management Plans and Data Management Toolkits); and
- effectively manage data generation and use, so that legal rights, obligations and requirements do not act as an impediment to open access and use of research data.⁷⁵²

3. Proposed Actions

10.06 This report proposes further work that is required to develop coherent and effective legal frameworks for research data:

- analysis of the data access and sharing practices described in Chapter 4;

⁷⁵² For example either where access is not permitted, or the legal position is unclear or the transaction costs of negotiating agreements for access and sharing are too high in terms of money and time.

- analysis of the open access policies and principles described in Chapter 8; and
- development of data access policies and principles, model data management plans, model data management toolkits and template agreements, as discussed in Chapter 9.

Proposal One – Analysis of Data Access and Sharing Practices

10.07 When structuring a database and determining access arrangements, reference should be made to existing models of data access developed by research organisations in Australia and overseas. An understanding of how issues such as ownership, use and control of data have been dealt with in a range of research databases established in different research fields will be of considerable assistance in the further development of data access strategies and mechanisms.

10.08 Chapter 4 surveyed data access frameworks and infrastructures established by research organisations and described:

- how a sample of existing databases are structured;
- the kind of data included in the databases;
- the arrangements for deposit of data into the databases; and
- the basis on which deposited data can be accessed and used by other researchers.

10.09 It is proposed that the following work is required in relation to data access and sharing practices:

- Further analyse existing databases to ascertain:
 - the various approaches adopted in relation to ownership, access, use and control of data;
 - policy statements, licensing terms, restrictions and conditions; and
 - consistency with legal requirements.
- Consider:
 - the effectiveness of the existing frameworks in facilitating access to and sharing of research data; and
 - the compatibility between the technical and management practices used in Australian databases and those established in overseas jurisdictions.

- Develop standard models and templates for research data frameworks, addressing matters including:
 - intellectual property rights in data;
 - the nature and scope of usage rights;
 - the rights and obligations of third parties;
 - legal disclaimer notices;
 - granting and reservation of rights;
 - jurisdiction and choice of law issues; and
 - the use of technological solutions and existing information exchange frameworks such as Creative Commons and Science Commons.

Proposal Two – Analysis of Open Access Policies and Principles

10.10 It is important for research organisations to take into account and build upon existing open access policies when developing their own policies on data access.

10.11 Chapter 8 considered existing open access policies and principles specifically relevant to research data and information in Australia and internationally.

10.12 It is proposed that the following work is required in relation to data access policies and principles:

- Conduct further comparative examination of policies, principles and guidelines on open access to research data;
- Consider the appropriateness of a range of open access policies and principles across different fields of research; and
- Analyse how various open access policies and principles have influenced or been given effect in data infrastructures, considering variations in the implementation of open access policies and principles in different research fields.

Proposal Three – Development of Data Access Policies and Principles, and Model Data Management Plans, Toolkits and Templates

10.13 Model and template documents should be provided for the guidance of individual researchers, research groups and organisations to practically assist in addressing the legal issues integral to data management in the research context.

10.14 Chapter 9 considered data access policies and principles, Data Management Plans (DMPs) and Data Management Toolkits (DMTs) and emphasised the importance of these mechanisms in the development of research data infrastructures.

10.15 It is proposed that the following work is required in relation to data access policies and principles:

- Develop template statements of open access policies and principles to:
 - clarify the value of open access and the importance of open access to data in the particular research context;
 - provide clear statements of open access policies and principles;
 - provide guidelines on the development of open access policies and principles; and
 - assist institutions and researchers in implementing open access policies and principles in specific fields of research.

10.16 DMPs address how data is to be managed in accordance with relevant legal controls, how data and the database will be preserved and maintained in the long-term and how this will be funded, how data is to be made accessible to the public and who will be primarily responsible for the management of the database and data. DMTs guide researchers in the collection and handling of data in accordance with the DMP, how to deposit data into the database and how to manage data in compliance with legal requirements applying to their field of research.

10.17 It is proposed that the following work is required in relation to DMPs and DMTs:

- Undertake further examination of various research disciplines, the data generated in those disciplines and the arrangements for funding of the research;
- Consider how researchers' data management practices could be assisted by the availability of DMPs and DMTs;
- Make recommendations as to the types and formats of DMPs and DMTs that should be adopted for different research fields;
- Develop model DMPs which cover a range of matters, including:
 - the management of data in accordance with relevant legal requirements;
 - preservation and maintenance of data and the database in the long-term;
 - maintenance of funding;
 - public accessibility of data;
 - identification of third party data;
 - use of open content licences for publicly available material; and

- identification of the persons who are primarily responsible for the management of the database and the data;
- Develop model DMTs which cover a range of matters, including:
 - practical guidance for researchers on how to manage and deposit their research data;
 - compliance and management guidelines tailored to regulatory and legal requirements applying to the particular field of research; and
 - guidelines for data collection and organisation in accordance with the DMP; and
- Draft a template Access Agreement and Repository Deposit Licence, consistent with the DMP and DMT to enable:
 - effective management of research data; and
 - support of the legal framework and compliance with legal and legislative requirements.

Index

Access...

Agreements...7.12-7.15
Definition of...2.27-2.29
Open access...8.01-8.04

Administrative Arrangements...3.53

Confidential Information...

Confidential nature...6.08-6.10
Defences...6.18-6.19
Employment...6.21-6.28
Government information...6.29-6.30
Loss of secrecy...6.31-6.32
Obligation of confidence...6.11-6.13
Remedies...6.20
Secret or confidential information...6.02-6.05
Unauthorised use or disclosure...6.14-6.17

Contract...

Access agreements...7.12-7.15
Click-wrap agreements...7.04
Confidentiality agreements...7.06-7.11
Contractual licences...7.17-7.19

Control...

Definition of...2.18

Copyright...

Assignment...5.21
Digital Rights Management (DRM)...5.36
Electronic Rights Management Information (ERMI)...5.37-5.41
European Database Right...5.89-5.96
Infringement...5.30-5.35
Licensing...5.42-5.47
Material form...5.04
Open content licensing...5.48-5.49
Originality...5.13-5.17
Rights...5.24-5.29

Statutory licences...5.69-5.73
United Kingdom...5.80-5.88
United States...5.75-5.79

Creative Commons...5.50-5.63, 9.28-9.29

Cultural Protocols...3.54

Cyberinfrastructure...1.12-1.14

Data...

Definition of...2.03-2.09

Data Management Plan

(DMP)...9.33-9.79, 10.13-10.17

Data Management Toolkit

(DMT)...9.80-9.84, 10.13-10.17

Data Security...9.63-9.68

Data Sharing...

AERES survey...4.342-4.349
Attitudes and practices...4.334-4.341
Clarity and understanding...4.292-4.297
Consistency of language...4.298-4.299
NCRIS survey...4.350-4.355
QUT e-Research survey...4.356-4.360

Data Sharing Frameworks...

Access and use...9.24-9.27
Australia...8.78-8.118
Bermuda Principles...8.05-8.11
Bethsheda statement...8.14
Berlin Declaration...8.15-8.16
Budapest Initiative...8.12-8.13
Canada...8.50-8.62
Data access policy and principles...9.02-9.17
Database infrastructure...9.73-9.74
EU Directive...8.71-8.77
ICSU...8.36-8.39

Identifying data...9.18-9.23
 Novartis...8.129-8.131
 OECD...8.26-8.35
 Practical management...9.30-9.32
 Wellcome Trust...8.119-8.128
 WSIS declaration...8.17-8.25
 United Kingdom 8.63-8.69
 United States - NIH Data Sharing
 Policy...8.42-8.49
 University and Innovation
 Summit...8.132-8.133

Dataset...

Definition of...2.10

Database...

Definition of...2.11-2.14
 Infrastructure...9.73-9.74

Database Examples...

ADT...4.219-4.236
 AODCJF and AODC...4.211-
 4.218
 ASSDA...4.195-4.204
 BRENDA...4.149-4.158
 Capture Wales...4.237-4.244
 CIRM...4.115-4.125
 EPP-ATLAS...4.273-4.279
 ESG...4.190-4.194
 HILDA...4.280-4.290
 MalariaGen...4.126-4.136
 MCD...4.164-4.171
 MT...4.205-4.210
 Nanowerk...4.159-4.163
 NCBI...4.93-4.94
 NCBI-dbGaP...4.95-4.108
 NCBI-Entrez...4.109-4.114
 NIH...4.08-4.11
 NGHRI-ENCODE...4.82-4.92
 NIH-GAIN...4.22-4.29
 NIH-Genbank...4.12-4.21
 NHGRI-IHGSC...4.41-4.44
 NHGRI-MGC...4.68-4.81
 NGHRI-MGSC...4.45-4.54
 NIH-NHGRI...4.31-4.40
 NHGRI-RGSC...4.55-4.62
 NHGRI-TSC...4.63-4.67
 PARADISEC...4.245-4.255
 RettBase and InterRett...4.179-
 4.189

SDSS...4.256-4.272
 UniProtKB/Swiss-Prot...4.137-
 4.148
 WAGER...4.172-4.178

Diagrams...Pages 178-192

e-Infrastructure...1.12

e-Research...

Definition of...2.15

Government Data...

Electronic Transactions
 legislation...3.49-3.50
 Freedom of Information Act...3.41,
 3.42, 3.47
 Information Standards...3.53
 Public Records legislation...3.41-
 3.48

Information...

Definition of...203-2.09
 Information Standards...3.53

Infrastructure

Data Sharing Infrastructure...9.01-
 9.84
 Database Infrastructure...9/73-9.74

Legal Frameworks...3.05-3.40, 4.07- 4.33,10.04-10.17

Open Access...8.01-8.133

Ownership...

Definition of...2.16-2.22
 Copyright...5.18-5.20
 Computer-Generated
 Works...5.22-5.23

Patents...

Data...4.300-4.313
 Open Licensing...4.314-4.333
 Bermuda Principles...4.314-
 4.319
 CAMBIA...4.326-4.333
 HapMap Project...4.320-4.325
 Parasitic patenting...4.314-4.319,
 4.322

Principles (Statements of)...

- Data access and sharing...8.01-8.04, 9.02-9.17, 10.05-10.06, 10.13-10.17
- Government and Public Sector...8.40-8.118
 - ANZLIC...8.98-8.100
 - APSR...8.101-8.106
 - ARC...8.78-8.82
 - Australian Government...8.107-8.108
 - Canadian Institutes of Health Research...8.50-8.52
 - European Research Council...8.70
 - European Union...8.71-8.77
 - Geoconnections Canada...8.53-8.62
 - NCRIS...8.114-8.118
 - NHMRC...8.78-8.82
 - NIH...8.42-8.49
 - Office of Spatial Data Management (OSDM)...8.83-8.97
 - PMSEIC...8.112-8.113
 - Productivity Commission...8.109-8.111
 - Research Councils UK...8.65
 - Research Information Network (RIN)...8.66-8.69
- International...8.05-8.39
 - Berlin Declaration...8.15-8.16
 - Bermuda Principles...8.05-8.11
 - Bethesda Statement...8.14
 - Budapest Open Access Initiative...8.12-8.13
 - International Council for Science (ICSU)...8.36-8.39
 - OECD...8.26-8.35
 - WSIS...8.17-8.25
- Open access...10.10-10.12, 8.01-8.04
- Private sector...8.119-8.133
 - Novartis...8.129-8.131
 - University and Industry Innovation Summit...8.132-8.133
 - Wellcome Trust...8.119-8.128

Privacy...

- Data linkage...3.26-3.30
- De-identifying personal information...3.20-3.22
- Draft National Health Privacy Code...3.32
- Genetic health information...3.39-3.40
- Guidelines under section 95A...3.17- 3.19
- Health information...3.05-3.40
- Information Privacy Principles...3.07-3.08, 3.10, 3.12, 3.15, 3.34
- National Health Act...3.33
- National Privacy Principles (NPP)...3.09-3.16
- Personally identifying information...3.05-3.40
- Privacy Act 1988...3.06-3.31
- Responses to Privacy Act...3.23-3.25
- State legislation...3.34-3.38

Proposals of this Report...10.06-10.17**Recommendations...10.06-10.17****Science Commons...5.64-5.68****Statutory Frameworks for Data Sharing...3.01-3.04, 3.06-3.25, 3.33-3.34, 3.41-3.52****Statutory Licences (Copyright)...5.69-5.73****Sustainability...4.342-4.349, 8.101-8.106, 9.04, 9.35, 9.75-9.77****Use...**

- Definition of...2.27-2.29

Open Access to Knowledge (OAK) Law Project

In today's ever-changing world, open access to knowledge is increasingly important. The **Open Access to Knowledge (OAK) Law Project** (321H<http://www.oaklaw.qut.edu.au>) hosted at QUT and funded by the Department of Education, Science and Training (DEST) aims to ensure that people can legally and efficiently share knowledge across domains and across the world. This will be of significance to the every day citizen through to top-end researchers.

The QUT team, led by Professor Brian Fitzgerald, acknowledges the increasing importance of access to knowledge as a key driver in enhancing social, cultural and economic development. The project recognises the need for clear protocols for copyright management across the research sector. With the objective of enabling access to knowledge the project is mapping out a program of action designed to develop practical and effective copyright management protocols for implementation in the Australian academic and research sector.

Legal Framework for e-Research Project

The face of global research is changing due to rapid advances in information and communications technology (ICT) and the rise of e-Research. National and international multi-disciplinary collaboration is now possible using a spectrum of advanced ICT capabilities that enhance and allow fast-paced, real-time and large-scale access to knowledge. Advances in ICT present enormous opportunities for Australian researchers.

The **Legal Framework for e-Research Project** (320H<http://www.e-Research.law.qut.edu.au>) led by Professor Brian Fitzgerald at QUT and funded by the Department of Education, Science and Training (DEST) aims to make the legal conditions as dynamic and effective as the advancing technology. By investigating issues such as contractual frameworks, data ownership, access and reuse, IP licensing, privacy and liability the Legal Framework for e-Research project will analyse the role of law in the e-Research environment and propose a more effective legal framework that will better enable the adoption of e-Research methods.