CHAPTER SEVEN

UNDERSTANDING THE LEGAL IMPLICATIONS OF DATA SHARING, ACCESS AND REUSE IN THE AUSTRALIAN RESEARCH LANDSCAPE[1]

Professor Anne Fitzgerald,[2] Kylie Pappalardo,[3] Anthony Austin[4]

INTRODUCTION

In today's world, researchers are increasingly involved in data-intensive research projects that cut across geographic and disciplinary borders.[5] 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

---

[1] This chapter is derived from the publication: Dr Anne Fitzgerald and Kylie Pappalardo (with the assistance of Professor Brian Fitzgerald, Anthony Austin and others), *Building the Infrastructure for Data Access and Reuse in Collaborative Research: An Analysis of the Legal Context* (2007) OAK Law Project <http://www.oaklaw.qut.edu.au/reports>.

[2] Professor, Faculty of Law, Queensland University of Technology, JSD (*Columbia*), LLM (*Columbia*), LLM (*London*), LLB (Hons) (*Tasmania*), Grad Dip Welfare Law (*Tasmania*), BSW (*UQ*), Barrister (*Qld*), Practitioner – High Court of Australia, am.fitzgerald@qut.edu.au

[3] Research Officer, Open Access to Knowledge (OAK) Law Project, Faculty of Law, Queensland University of Technology LLB (*QUT*) (Hons), BCI (Creative Writing) (*QUT*), Grad Dip Legal Prac (*QUT*), k.pappalardo@qut.edu.au

[4] Research Officer, Open Access to Knowledge (OAK) Law Project, Faculty of Law, Queensland University of Technology, LLB (*QUT*), Grad Dip Legal Prac (*QUT*), LLM (*QUT*), Solicitor of the Supreme Court of Queensland and the Federal Court of Australia, anthony.austin@qut.edu.au

[5] International Council for Science (ICSU), *Scientific Data and Information: A report of the CSPR Assessment Panel* (2004) 7; see also Dr Anne Fitzgerald and Kylie Pappalardo, *Building the Infrastructure for Data Access and Reuse in Collaborative Research: An Analysis of the Legal Context* (2007) 6, OAK Law Project <http://www.oaklaw.qut.edu.au/reports> (hereinafter A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007)).

participation and accelerate discoveries.[6] The result of such large-scale collaborations has been the production of ever-increasing amounts of data. In short, we are in the midst of a data deluge.[7]

Accompanying these developments has been a growing recognition that if the benefits of enhanced access to research are to be realised, it will be necessary to develop the systems and services that enable data to be managed and secured.[8] It has also become 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.[9]

This chapter provides an overview of the current research landscape in Australia as it relates to the collection, management and sharing of research data. The chapter then explains the Australian legal regimes relevant to data, including copyright, patent, privacy, confidentiality and contract law. Finally, this chapter proposes the infrastructure elements that are required for the proper management of legal interests, ownership rights and rights to access and use data collected or generated by research projects.

THE AUSTRALIAN DATA LANDSCAPE

The last few years have seen a revolution in the way that research data is produced, stored, analysed and disseminated.[10] Now, vast amounts of data can be generated and accessed through distributed networks online. In response to the enormous growth in data collection and generation in recent years, there has been increased interest from Australian government and research sectors in developing systems to manage data

[6] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 6.

[7] In an interview with Richard Poynder, Tony Hey said, 'We are going to be deluged with data in almost every field': Richard Poynder, Interview with Tony Hey 'A Conversation with Microsoft's Tony Hey' *Open and Shut?* (Blog, 12 December 2006) <http://poynder.blogspot.com/2006/12/conversation-with-microsofts-tony-hey.html> at 5 May 2008.

[8] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 6–7.

[9] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 9.

[10] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 3.

and facilitate access to research outputs.[11] This section provides a brief overview of some of these initiatives.

Government Initiatives

In May 2004, then Prime Minister John Howard announced that the Australian Government would establish quality and accessibility frameworks for publicly funded research as part of the *Backing Australia's Ability – Building Our Future through Science and Innovation* package.[12] The Accessibility Framework for Publicly Funded Research was designed to manage research information, outputs and infrastructure in order to enable them to be more readily discovered, accessed and shared. 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.[13]

A project funded under the *Backing Australia's Ability* package is the National Collaborative Research Infrastructure Strategy (NCRIS). The NCRIS capability known as Platforms for Collaboration supports technological platforms that enhance researchers' ability to generate, collect, share, analyse, store and retrieve information.[14] A central component of Platforms for Collaboration is the Australian e-Research Infrastructure Council (AeRIC), established by the federal Government's Department of Education Science and Training (DEST)[15] upon the

---

[11] See for example, National Collaborative Research Infrastructure Strategy (NCRIS) Committee, Submission to the Review of the National Innovation System (NIS) (April 2008) and AeRIC, *Closing the Gap: Connecting Researchers to the Innovation System Through Sustained Investments in Collaborative Research Infrastructure*, Submission to the Review of the National Innovation System (NIS) (April 2008).

[12] See <http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/accessibility_framework/> and <http://backingaus.innovation.gov.au/> at 24 April 2008.

[13] See <http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/accessibility_framework/> at 24 April 2008. See also A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 3.

[14] <http://www.ncris.dest.gov.au/capabilities/collaborative_investment_plan_platforms.htm> at 24 April 2008; see also, National Collaborative Research Infrastructure Strategy (NCRIS) Committee, Submission to the Review of the National Innovation System (NIS) (April 2008).

[15] Since the change of Federal Government, AeRIC now falls under the auspices of the Department of Innovation, Industry, Science and Research (DIISR): see Dr Rhys Francis on

recommendation of the Australian Government e-Research Coordinating Committee.[16] In the report, *An Australian e-Research Strategy and Implementation Framework*, the e-Research Coordination Committee had recommended that:

> the Government convenes a working group to develop an Australian Research Data Strategy that will support a standardised national approach to the management of data collected, generated and used by the Australian research community.[17]

This recommendation was endorsed in the NCRIS Platforms for Collaboration Final Investment Plan.[18]

AeRIC's responsibilities were established at its inaugural meeting on 23 July 2007, as:

> ensur[ing] that world class services and expertise are identified, developed and delivered nationwide in ways that support effective e-Research within and across all research disciplines … includ[ing] services and expertise relating to: data capture, management, retention, publication, discovery and reuse … [19]

AeRIC undertakes an important coordination role in relation to the NCRIS Platforms for Collaboration infrastructure.[20] It is tasked with ensuring the integration and sustainability of research infrastructure and

---

behalf of Professor Tom Cochrane, AeRIC submission to the National Innovation System (NIS) Review – coversheet, 30 April 2008.

[16] In October 2004, the Australian Government committed to the formation of an overarching e-Research Coordinating Committee, which would provide expert advice to the Government on a strategic framework for the development of Australia's e-Research capacity: see for example, Catherine Harboe-Ree, 'eResearch Coordinating Committee' (CAUL Presentation, September 2005) <http://www.caul.edu.au/caul-doc/caul20052eresearch.ppt> at 3 May 2008.

[17] Final Report of the e-Research Coordinating Committee, *An Australian e-Research Strategy and Implementation Framework,* DEST, (April 2006) 55.

[18] See <www.ncris.dest.gov.au/capabilities/documents/PfC_Investment_Plan_Summary_pdf.htm> at 24 April 2008. The NCRIS Committee accepted the Final Investment Plan on 13 April 2007: <http://www.pfc.org.au/bin/view/Main/PlatformsHistory> at 24 April 2008.

[19] See <http://www.pfc.org.au/bin/view/Main/AeRIC> at 24 April 2008.

[20] AeRIC, *Closing the Gap: Connecting Researchers to the Innovation System Through Sustained Investments in Collaborative Research Infrastructure*, Submission to the Review of the National Innovation System (NIS) (April 2008) 9.

services capitalising on the Government's substantial investments in NCRIS capabilities.[21]

In October 2007, DEST and AeRIC released the report, *Towards the Australian Data Commons,*[22] proposing the establishment of the Australian National Data Service (ANDS). A similar proposal had previously been put forward in the Platforms for Collaboration Final Investment Plan as a means of addressing the recommendations of the Prime Minister's Science Engineering and Innovation Council (PMSEIC) Data for Science Working Group[23] in its December 2006 report:

> **Recommendation 1:** 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.
>
> **Recommendation 6:** That 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.
>
> **Recommendation 9:** That in the context of developing the strategic framework for scientific data management, Australia's intellectual property approaches be checked to ensure they do not impede the sharing of data …[24]

---

[21] AeRIC, *Closing the Gap: Connecting Researchers to the Innovation System Through Sustained Investments in Collaborative Research Infrastructure*, Submission to the Review of the National Innovation System (NIS) (April 2008) 9.

[22] AeRIC, *Towards the Australian Data Commons: A proposal for an Australian National Data Service*, DEST (October 2007).

[23] See <www.ncris.dest.gov.au/capabilities/documents/PfC_Investment_Plan_Summary_pdf.htm> at 24 April 2008. The NCRIS Committee accepted the Final Investment Plan on 13 April 2007: <http://www.pfc.org.au/bin/view/Main/PlatformsHistory> at 24 April 2008.

[24] PMSEIC Working Group on Data for Science, *From Data to Wisdom: Pathways to Successful Data Management for Australian Science* (December 2006) 11–12.

ANDS offers common services in support of research data collections and integration infrastructure to facilitate sharing and reuse of data.[25] The ANDS Utility Program will provide a national registry covering issues such as data access policies, usage rights and licensing requirements associated with data access.[26] It will also provide template data access policies that can be adapted for discipline specific needs.[27]

At an AeRIC meeting on 22 February 2008, it was reported that a contract was signed in November 2007 with Monash University to conduct the ANDS Establishment Project through to the end of June 2008.[28] Under this agreement, Monash University will work with the Australian National University (ANU), the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and other relevant parties to develop the necessary elements to move to full ANDS implementation from July 2008.[29]

Research projects

In addition to the larger scale initiatives described above, there is an abundance of smaller projects that focus on the collecting and compiling of research data in a specific scientific field. One example is the Integrated Marine Observing System (IMOS), which is coordinated by staff at the University of Tasmania supported by CSIRO Marine and Atmospheric Research.[30] IMOS is a nation-wide collaborative program designed to observe the oceans around Australia, including the coastal oceans and the 'bluewater' open oceans.[31] One of the five IMOS research 'nodes' is the Great Barrier Reef Ocean Observing System (GBROOS), which is an observation network covering the eastern Coral

[25] AeRIC, *Towards the Australian Data Commons: A proposal for an Australian National Data Service*, DEST (October 2007) 4.

[26] AeRIC, *Towards the Australian Data Commons: A proposal for an Australian National Data Service*, DEST (October 2007) 36.

[27] AeRIC, *Towards the Australian Data Commons: A proposal for an Australian National Data Service*, DEST (October 2007) 36.

[28] AeRIC Executive Director's Report, *Meeting #5* (22 February 2008) <http://www.pfc.org.au/bin/view/Main/AeRIC-5> at 24 April 2008.

[29] AeRIC Executive Director's Report, *Meeting #5* (22 February 2008) <http://www.pfc.org.au/bin/view/Main/AeRIC-5> at 24 April 2008.

[30] <http://imos.org.au/about.html> at 20 May 2008.

[31] <http://imos.org.au> at 20 May 2008.

Sea and the Great Barrier Reef.[32] Among other things, GBROOS will monitor the effect of rising ocean temperatures on the incidence of coral bleaching over the next decade.[33] GBROOS includes the world's first large scale reef-based Internet Protocol (IP) network.[34] Data generated by the IMOS project will be made available to researchers through the electronic Marine Information Infrastructure (eMII) located at the University of Tasmania.[35] After defining specific data streams, IMOS will eventually develop end-to-end protocols, standards and systems to join the related observing systems into a unified data storage and access framework.[36] Data will be archived within the Australian Ocean Data Network (AODN), which is a distributed data storage and discovery network based at leading Australian marine research facilities.[37] Data storage and retrieval in IMOS is designed to be interoperable with other national and international programs.[38] IMOS is an NCRIS funded project.[39]

Another example is the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC).[40] PARADISEC offers a facility for digital conservation and access for endangered materials from the Pacific region, defined broadly to include Oceania and East and Southeast Asia.[41] PARADISEC is also a national repository for recorded material relating to indigenous cultures of regions in and around Australia. PARADISEC has established a framework for accessing, cataloguing and digitising audio, text and visual material, and preserving digital copies.[42] The project has been funded by the

[32] <http://imos.org.au/newsitem.html?&tx_ttnews[tt_news]=64&tx_ttnews[backPid]=2&cHash=d32f9070cb> at 20 May 2008.

[33] <http://imos.org.au/gbroos.html> at 20 May 2008.

[34] <http://imos.org.au/newsitem.html?&tx_ttnews[tt_news]=64&tx_ttnews[backPid]=2&cHash=d32f9070cb> at 20 May 2008.

[35] <http://imos.org.au/about.html> at 20 May 2008, see also <http://imos.org.au.emii.html> at 20 May 2008.

[36] <http://imos.org.au/emii.html> at 20 May 2008.

[37] <http://imos.org.au/data_access.html> at 20 May 2008.

[38] <http://imos.org.au/emii.html> at 20 May 2008.

[39] <http://imos.org.au/about.html> at 20 May 2008.

[40] <http://paradisec.org.au/home.html> at 25 April 2008.

[41] <http://www.paradisec.org.au/about.html> at 25 April 2008.

[42] <http://www.paradisec.org.au/about.html> at 25 April 2008.

Universities of Sydney, Melbourne and New England, ANU, the Australian Research Council (ARC) and GrangeNet.[43]

Funding policies

Australian funding bodies have taken an interest in the management of and provision of access to research data. 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. Both policies encouraged 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.[44]

The same guidelines are contained in the ARC *Discovery Project Funding Rules for funding commencing in 2009*,[45] and the NHMRC *Project Grants Funding Policy for funding commencing in 2009*.[46]

The introduction of open access requirements for data resulting from funded research projects was supported by the Australian Government Productivity Commission in its 2007 report, *Public Support for Science and Innovation*.[47] The Productivity Commission commended the steps taken

---

[43] <http://www.paradisec.org.au/about.html> at 25 April 2008.

[44] Australian Research Council (ARC), *Discovery Projects Funding Rules for funding commencing in 2008*, [1.4.5.1] and [1.4.5.3] <http://www.arc.gov.au/pdf/DP08_FundingRules.pdf> at 25 April 2008; National Health and Medical Research Council (NHMRC), *Project Grants Funding Policy for grants commencing in 2008*, [16.2]. See also A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 4.

[45] Australian Research Council (ARC), *Discovery Projects Funding Rules for funding commencing in 2009*, [4.4.5.1] and [4.4.5.3] <http://www.arc.gov.au/ncgp/dp/dp_fundingrules.htm> at 25 March 2008.

[46] National Health and Medical Research Council (NHMRC), *Project Grants Funding Policy for funding commencing in 2009*, [16.2] <http://www.nhmrc.gov.au/FUNDING/apply/granttype/projects/index.htm> at 25 March 2008.

[47] Productivity Commission, *Public Support for Research and Innovation*, Research Report (2007) 240, 243 <http://www.pc.gov.au/study/science/docs/finalreport> at 25 April 2008.

by the ARC and NHMRC to promote open access to the results of the projects they fund. However, the Productivity Commission 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.[48]

Surveys of researchers

While the benefits of data sharing have been widely recognised by government agencies and scientific organisations, there is a degree of reluctance among researchers to embrace data sharing practices. Recent surveys of the Australian research community provide indications of current attitudes and practices in relation to data ownership and sharing.

The NCRIS Platforms for Collaboration Data Management Survey conducted in September and October 2006 surveyed key stakeholders in the management of research data throughout Australia.[49] The results of the survey demonstrated that while some researchers are aware of the complexity of the issues involved in data ownership, most have only a rudimentary understanding.[50] Further, the survey made clear 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 data by others, especially if this would diminish the credit attributed to the researcher who generated the data in the first place.[51] The NCRIS survey made apparent the need for researchers to be provided with guidelines and data management infrastructure to assist in developing a better understanding of data ownership and management.[52]

In October 2006, the Australian Partnership for Sustainable Repositories (APSR) project published the results of a survey of researchers based

---

[48] Productivity Commission, *Public Support for Research and Innovation*, Research Report (2007) 240, 243 <http://www.pc.gov.au/study/science/docs/finalreport> at 25 April 2008. See also A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 4.

[49] See <www.pfc.org.au/twiki/pub/Main/DataWorkshop1/NCRISsurveyanalysis.pdf> at 25 April 2008.

[50] See A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 128.

[51] See A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 128.

[52] See A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 128.

across several research institutions and research service providers throughout Australia. The results of this Australian e-Research Sustainability Survey (AERES) were published in a report entitled, *Sustainability Issues for Australian Research Data: the report of the Australian e-Research Sustainability Survey Project.*[53] The AERES study found a distinct lack of formal policies for data management utilised by the surveyed researchers.[54] The report concluded that current data practices generally see data managed sufficiently for research needs but not professionally; discoverable through scientific publication but not otherwise; and having a value placed on it for present needs but not for the future.[55]

In August 2007, the Legal Framework for e-Research Project based at the Queensland University of Technology (QUT) published the report, *Legal and project agreement issues in collaboration and e-Research: Survey Results*.[56] This report documented a survey that was conducted online during May 2007 and was open to all Australian participants involved in collaborative research.[57] The QUT survey found that many researchers consider legal agreements to be an impediment to timely research and will often commence collaborative research projects before finalising agreements dealing with data ownership and other legal interests.[58] One participant, a university researcher in the Arts and Social Sciences, responded, 'Perhaps the biggest problem facing e-Research is the lack of understanding and agreement as to what is required in terms of local and

---

[53] Markus Buchhorn and Paul McNamara, *Sustainability Issues for Australian Research Data: The Report of the Australian e-Research Sustainability Survey Project*, Australian Partnership for Sustainable Repositories (APSR) (2006) <http://www.apsr.edu.au/aeres> at 25 April 2008.

[54] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 126.

[55] Markus Buchhorn and Paul McNamara, *Sustainability Issues for Australian Research Data: The Report of the Australian e-Research Sustainability Survey Project*, Australian Partnership for Sustainable Repositories (APSR) (2006) 44 <http://www.apsr.edu.au/aeres> at 25 April 2008.

[56] Maree Heffernan and Nikki David, *Legal and project agreement issues in collaboration and e-Research: Survey Results*, Legal Framework for e-Research Project, Queensland University of Technology (QUT) (2007) <http://www.e-research.law.qut.edu.au/> at 25 April 2008.

[57] <http://www.e-research.law.qut.edu.au/> at 25 April 2008.

[58] See Maree Heffernan and Nikki David, *Legal and project agreement issues in collaboration and e-Research: Survey Results*, Legal Framework for e-Research Project, Queensland University of Technology (QUT) (2007) 38, 62 <http://www.e-research.law.qut.edu.au/> at 25 April 2008.

national information infrastructure to support e-Research activities'.[59] The QUT survey highlighted the need for simple and easy-to-use resources to assist researchers in managing the legal rights surrounding data and e-Research, particularly where collaborative research projects are concerned.

LEGAL IMPLICATIONS SURROUNDING DATA ACCESS, SHARING AND REUSE

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.[60] For example, arrangements between a researcher and other researchers, research institutions or funding bodies may be governed by contract. Data compilations may attract copyright protection and data may also attract protection under confidentiality or privacy laws. This section provides an overview of the different legal regimes that may apply to and impact upon data collection, access, sharing and reuse.

Copyright

A general principle of copyright law is that copyright protects the material form in which ideas, information or facts are expressed and not the ideas, information or facts themselves. It follows that under this general principle, copyright law will not protect raw data. However, in Australia, copyright law may operate to protect compilations of data, such as datasets or databases, provided that the compilation meets the originality threshold required by law. Under the *Copyright Act 1968* (Cth), a compilation is protected as a literary work.[61]

Compiled data will not always be raw data – a compilation may also include written materials, reports, diagrams, tables and graphs. Where a data item meets the form and originality requirements under the *Copyright Act*, it may be protected by copyright as an independent work.

---

[59] Maree Heffernan and Nikki David, *Legal and project agreement issues in collaboration and e-Research: Survey Results*, Legal Framework for e-Research Project, Queensland University of Technology (QUT) (2007) 62 <http://www.e-research.law.qut.edu.au/> at 25 April 2008.

[60] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 263.

[61] *Copyright Act 1968* (Cth) s 10(1).

An important distinction lies between copyright in discrete data items and copyright in a database as a whole. In the latter, copyright serves to protect 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 parties.[62]

Copyright will only protect a work that possesses the requisite level of originality under law. In *Desktop Marketing v Telstra*,[63] the court considered the issue of whether a compilation is sufficiently original to attract copyright protection. The question for the court was whether Telstra held copyright in their White Pages and Yellow Pages directories, which are essentially a compilation of names, addresses and phone numbers listed alphabetically. In a landmark judgment, the court held that Telstra did own copyright in their compilations, thereby establishing that the originality threshold for copyright protection is low. The court held that copyright can be claimed in a compilation that:

1. has been produced as a result of the exercise of skill, judgment or knowledge in the selection, presentation or arrangement of the materials; or
2. has required the investment of a substantial amount of labour or expense to generate or collect the material included in the compilation (the so-called 'sweat of the brow' approach).[64]

Telstra, in undertaking substantial labour and incurring substantial expense, had met the originality threshold in compiling the Yellow Pages and White Pages directories, notwithstanding that there may have been minimal intellectual input or creativity involved in the selection and arrangement of the material.

Significantly, the court in this decision prescribed a lower threshold for originality under Australian copyright law than that required in the

---

[62] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 137.

[63] *Desktop Marketing Systems Pty Ltd v Telstra Corporation Ltd* [2002] FCAFC 112. See also *Nine Network Australia Pty Ltd v Ice TV Pty Ltd* [2008] FCAFC 71.

[64] *Desktop Marketing Systems Pty Ltd v Telstra Corporation Ltd* [2002] FCAFC 112, [409].

United States, where there must be a degree of creativity applied in the selection, coordination or arrangement of the compilation.[65]

The owner of copyright in a database, dataset or where applicable, a discreet item of data will be able to control how that database, dataset or data is used, copied and shared. It would be wise for a research project that intends to allow its data to be openly shared and reused to formulate plans and policies that properly define, allocate and manage copyright interests in the data and database.

Patents

Patents protect products and processes that are novel, useful and involve an inventive or innovative step.[66] They confer on the patentee the exclusive right to exploit the patented product or process for a period of time (usually 20 years from the time of filing the patent application).[67] 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. This situation has most commonly arisen in the context of patenting genomic data.[68]

Researchers collecting data may be concerned with patents for one of two reasons. Firstly, some researchers may be interested in obtaining a patent over a product or process that incorporates data which they have collected. For these researchers, disclosure of data could prevent a patent being obtained because releasing information into the public domain could preclude the 'novel' or 'inventive' aspect of a product or process that is required by law to secure a patent.[69] In these circumstances, prior to obtaining a patent, data should only be disclosed under confidentiality agreements to ensure that the data is kept out of the public domain.

Secondly, some researchers may want to ensure that their data is kept free of legal restrictions including patents, in order to allow sharing and

---

[65] See *Feist Publications Inc v Rural Telephone Service Co Inc*, 499 US 340, 349 (1991).

[66] *Patents Act 1990* (Cth) s 18.

[67] *Patents Act 1990* (Cth) ss 13 and 67.

[68] For more information, see A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 116–23.

[69] See *Patents Act 1990* (Cth) s 7.

reuse by themselves and others. For these researchers, simply releasing data into the public domain may be enough to create prior art and thus prevent successful patent applications by others.[70] However, even where data is released publicly it may be possible for another party to make improvements to the disclosed data and then make these improvements proprietary. Where 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.[71] Fortunately, there are contractual and licensing options that can be employed to keep data free of restrictive patent claims. One option is to release data via an online database where users accessing the database are required to first enter into a click-wrap agreement that governs use of the data. The agreement can prohibit patent applications based on certain data, or may allow patent applications but provide that the patent must not be restrictive and must allow further use of the patented data.[72] Another option is to actually obtain a patent over a product or process based on or encompassing the research data, but then to licence the use of the protected data under liberal terms.[73]

---

[70] This was the approach underlying the Bermuda Principles, which were developed in 1996 by a consortium of researchers involved in the Human Genome Project. For more information, see A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 118–20.

[71] See A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 119–20; Donna M Gitter, 'Resolving the Open Source Paradox in Biotechnology: A Proposal for a Revised Open Source Policy for Publicly Funded Genomic Databases' (2007) 43(4) *Houston Law Review* 4 <http://ssrn.com/abstract=901994>; 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 Technology Law Journal* 1187, 1207; Claire T Driscoll, 'NIH data and resource sharing, data release and intellectual property policies for genomics community resource projects' (2005) 15(1) *Expert Opinion on Therapeutic Patents* 4.

[72] This was the approach adopted by the International Haplotype Project (commonly known as the HapMap Project), which ran from 2002 to 2005. For more information, see A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 120–21.

[73] This was the approach adopted by the CAMBIA project. See CAMBIA, 'About BiOS (Biological Open Source) Licenses and MTAs' <http://www.bios.net/daisy/bios/licenses/398.html> at 11 April 2008. See also, Richard Jefferson, 'Science as Social Enterprise: The CAMBIA BiOS Initiative' (2006) 1(4) *Innovations: Technology, Governance, Globalization* 13; and A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 121–3.

Privacy

Some research, particularly research in medical fields, will give rise to privacy concerns about the handling and use of personally identifying and health information. In response to privacy concerns, the *Privacy Act 1988* (Cth) requires Commonwealth public sector entities to act in accordance with Information Privacy Principles and private sector entities to act in accordance with National Privacy Principles. The Information Privacy Principles 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.[74] The National Privacy Principles provide that personal information cannot be used except for the lawful purpose for which it was collected.[75] 'Personal information' is defined in the *Privacy Act* as 'information or an opinion … about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion.'[76]

The National Privacy Principles differentiate between 'personal information' and 'sensitive information'. Sensitive information is accorded a higher level of protection and is defined to include health or genetic information about an individual.[77] An organisation must generally not collect sensitive information about an individual unless the individual has consented.[78] Obtaining consent to collect sensitive information for research purposes will usually involve explaining to the participant the purpose, methods, possible risks and potential outcomes of the research, including the likelihood that research results will be published.

There are limited exceptions to the requirements imposed in the Information Privacy Principles and the National Privacy Principles relating to the collection and disclosure of personal and sensitive information. For example, consent to disclose personal information will not be required where the participant was reasonably likely to have been

[74] *Privacy Act 1988* (Cth) s 14: Information Privacy Principle 1.

[75] *Privacy Act 1988* (Cth) Schedule 3.

[76] *Privacy Act 1988* (Cth) s 6.

[77] *Privacy Act 1988* (Cth) s 6; Schedule 3: National Privacy Principle 10.

[78] *Privacy Act 1988* (Cth) Schedule 3: National Privacy Principle 10.

aware or would reasonably expect that the information would be disclosed, or where it is impractical to obtain consent.[79]

The definition of 'personal information' refers to information that can be used to identify an individual. Where information has been de-identified such that it cannot be re-identified, it can usually be used and disclosed in research and data-linkage without fear of infringing the *Privacy Act*. Studies show that individuals generally support the idea of researchers being able to access health information from databases, provided that the information is identified by a unique number rather than a name.[80] The National Health and Medical Research Council (NHMRC) has recommended 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.[81]

All Australian States and Territories except Queensland and South Australia have enacted privacy legislation or introduced privacy bills relating to health information and/or the collection and use of personal information in the State public sector.[82] In South Australia, the Privacy Committee is responsible for administrative protocol PC012 – Information Privacy Principles Instruction, which applies to public

---

[79] See *Privacy Act 1988* (Cth) s 14: Information Privacy Principle 11.1; Schedule 3: National Privacy Principles 2.1 and 10.3; Australian Law Reform Commission, *Review of Privacy* (Issues Paper 31, 2006) [8.124].

[80] Australian Law Reform Commission, *Review of Privacy* (Issues Paper 31, 2006) [8.237].

[81] National Health and Medical Research Council (NHMRC), *National Statement on Ethical Conduct in Research Involving Humans* (1999) 53 <http://www.nhmrc.gov.au/publications/synopses/_files/e35.pdf>. Note: the 2007 revised National Statement on Ethical Conduct in Research Involving Humans was tabled in Parliament on 28 March 2007.

[82] *Privacy and Personal Information Act 1998* (NSW), *Health Records and Information Privacy Act 2002* (NSW), *Information Privacy Act 2000* (Vic), *Health Records Act 2001* (Vic), *Australian Capital Territory Government Service (Consequential Provisions) Act 1994* (Cth), *Health Records (Privacy and Access) Act 1997* (ACT), *Information Act 2002* (NT), *Personal Information Protection Act 2004* (Tas), *Health Complaints Act 1995* (Tas). For Western Australia, see *State Records Act 2000* (WA) s 49 and *Freedom of Information Act 1992* (WA) s 3. In March 2007, the *Information Privacy Bill 2007* (WA) was introduced into the Legislative Assembly. It passed the Legislative Assembly on 27 November 2007 and reached the second reading speech stage in the Legislative Council on 4 December 2007 (see <http://www.parliament.wa.gov.au/web/newwebparl.nsf/iframewebpages/Bills+-+Current> at 20 May 2008).

sector handling of personal information. In Queensland, there are two administrative protocols applying to the State's public sector. Information Standard 42: Information Privacy applies to the collection of personal information in the public sector generally, while Information Standard 42A: Information Privacy for the Queensland Department of Health applies to the collection of health information. The differences in privacy regulation at Commonwealth and State levels has caused some confusion for medical researchers, prompting the NHMRC, the Australian Government Productivity Commission and the Australian Law Reform Commission (ALRC) to recommend a nationally consistent approach to privacy regulation of health information.[83]

Confidential Information

Data that has not been released into the public domain may be protected by the law of confidentiality. A researcher who has expended considerable time and energy in generating or collecting data may have an interest in protecting that the data from others who have not contributed to its production. In such a situation, the action for breach of confidence can be used to control access to the data.

The law of confidentiality is based on the equitable principle that a person who receives information in confidence shall not take unfair advantage of that information.[84] A successful breach of confidence action must establish three elements:

1. the information is confidential in nature;
2. the information was imparted in circumstances importing an obligation of confidence; and
3. an unauthorised use of the information to the detriment of the person claiming the right to maintain confidentiality.[85]

Data will only be protected as confidential if it is not in the public domain. A breach of confidence action can still be established where

[83] Australian Government Productivity Commission, *Public Support for Science and Innovation*, Research Report (2007) 189, 217; Australian Law Reform Commission, *Review of Australian Privacy Law: An Overview of Discussion Paper 72* (2007) 5–6.

[84] *Seager v Copydex Ltd* [1967] 2 All ER 415, 417.

[85] *Coco v A N Clark (Engineers) Ltd* [1969] RPC 41, 47.

more than one person knows about or has access to the data, provided that not so many people know about the data that it can no longer be regarded as secret. Usually, confidentiality will be protected through the use of confidentiality agreements, which provide for the disclosure of information on the condition that the contracting party does not further disclose the information and does not use the information except for the purposes set out in the agreement.

Confidentiality will be lost if enough people know about the data such that it passes into the public domain, or if the data is independently discovered by someone else.[86] Where data loses its quality of secrecy, it is still possible for a researcher to control access to and use of the data through contract.

Contract

In practice, the most important legal mechanism used to allocate rights to data is the contractual agreement. There are three main types of agreement relevant to regulating data access and use – the confidentiality agreement, the copyright licence and the access agreement.

Confidentiality agreements, also called non-disclosure agreements, serve to protect secret information by disclosing the information in a controlled setting so that it remains confidential and is not released into the public domain. Confidentiality agreements will generally: identify the owner of rights in relation to the confidential information; identify the information that is to be treated as confidential; impose obligations on the person to whom the information is disclosed to maintain the secrecy of the information; define the scope of the permitted use of the information; and provide for the consequences of a failure to comply with the confidentiality obligations.[87]

Copyright licences grant permission to a person to deal with a database or a dataset in a way that would otherwise infringe copyright. For example, a copyright owner may permit - through a licence - a researcher to reproduce copyright material and make the material available on a website where it can be be accessed and downloaded by other researchers. A contractual copyright licence may also contain terms that

[86] *Attorney General (UK) v Heinemann Publishers Australia Pty Ltd* (1988) 10 IPR 153.

[87] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 175.

are not strictly related to copyright. For example, the licence may require the researcher to undertake not to hold the copyright owner liable for consequences resulting from any inaccuracies that may be contained in the data supplied.[88] Contractual licences will usually indicate the copyright material to which the licence refers; the permitted acts that the licensee is authorised to do; any restrictions upon the party acting under the licence; the consideration provided for the licence; and whether or not the licence is exclusive (or non-exclusive) and whether it can be revoked or is irrevocable.[89]

Access agreements will operate where a researcher or research organisation has control over the database in which their data is stored. The researcher or research organisation may require persons interested in obtaining access to the data to first enter into an access agreement. Access agreements may: identify the data to be accessed; identify the person/s or class of persons who are permitted to access the data; state that access rights cannot be transferred to third parties; limit the purposes for which the data may be used; contain a disclaimer that the researcher is not responsible for any inaccuracies in the data; and provide for the consequences of a failure to comply with the agreement.[90] For example, an access agreement may provide that the data can be accessed and used for non-commercial purposes only, or may provide that if a user engages in commercial uses of the data, they must account back to the researcher for a proportion of the profits. Access agreements can be used to control access to and use of data that was formerly protected through confidentiality agreements but which has lost its quality of confidence.

DATA SHARING INFRASTRUCTURE

It will not be sufficient for researchers and database managers to simply be aware of the laws that surround the data they collect. If data is to be effectively made available within the research community, it is necessary that it is properly managed. Research projects would be wise to adopt protocols for dealing with the legal issues that may arise in relation to the

[88] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 173.

[89] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 177.

[90] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 176–7.

data they collect. Failure to establish legal protocols for data management may jeopardise the research community's ability to access, share and use valuable research outputs.[91] Data sharing infrastructure, such as data management policies, principles, plans and toolkits, can assist researchers and database managers to effectively manage their legal rights, interests and obligations in relation to the data collected, generated and compiled by the research project.

Data management policies and principles

A forward-thinking research project will have in place a data management policy containing high-level statements about how data generated or complied by the research project is to be made available for access and use. The data management policy may also contain principles expanding on the high-level statements and indicating how they are to be applied.[92]

A data management policy will take into consideration the research discipline of the project; the funding arrangements for the research project; the kind of data generated or collected by the project; how and when data is to be deposited into a database; how, when and on what basis data is to be made available for access by other researchers; and any legal obligations imposed on the research project or individual researchers.[93]

A research project must give careful consideration to formulating a policy which ensures that researchers' objectives, needs and responsibilities in each research situation are properly addressed.[94] For example, where a research project is publicly funded, it may be appropriate for a policy to strongly support immediate open access to research data. However, immediate open access may not be appropriate for data generated by private sector research projects.

The Australian Partnership for Sustainable Repositories (APSR) has highlighted the importance of all data management policies including

---

[91] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 263.

[92] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 240.

[93] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 241.

[94] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 241.

clear definitions of concepts and terms used within the policy.[95] Additionally, research projects should take care to distinguish in their policies data that is to be made accessible from data that is not. This is particularly important where a research project is collecting data subject to privacy limitations or data that is to be commercially exploited.

Data management policies and principles will also explain the conditions under which data is to be made available for access and use. For example, access may be limited to certain categories of researchers or researchers may only be permitted to use the data for specified purposes. In order to properly ascertain and set out the conditions of access and use, each research project should develop a clear and comprehensive listing of all legal restrictions applying to the management, dissemination and reuse of the different kinds of data that may be generated by the project.[96]

Data management plans

Similar to a data management policy, a data management plan (DMP) will address how data is collected, stored, managed and disseminated. It will also be concerned with data ownership and the legal controls surrounding data. However, a DMP will focus on practical measures rather than making broad policy statements. It will also consider expenditures and technical measures to ensure sustainability of data .[97]

A DMP should be in place from the conception and commencement of a research project. A comprehensive DMP will recognise that there are many different parties involved in a research project and will have relevance to all of the different parties. These parties include collectors and compilers of data, data analysts, database managers, parties that have funded the research project and consumers or users of the data and database.[98]

---

[95] Anna Shadbolt et al, *Sustainable Paths for Data-intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable Repositories* (August 2006) 38–9 <http://www.apsr.edu.au/aeres/sustainable_paths.pdf>.

[96] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 244.

[97] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 247–56.

[98] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 247–8.

It is important that a DMP addresses unusual situations that may arise in the collation of data. For example, where data that is generated by the research project is to be integrated with existing data from other sources, the DMP will need to explain how this will be done and how data from each source will be identified once combined. It must also ensure that legal rights and obligations are respected.

Two central issues for each research project to consider in its DMP are:

1. who owns the data generated or collected by the research project; and
2. who is responsible for managing the data?

Data may be owned by more than one person. An owner may be the researcher who has collected or generated the data; the researcher's employer, 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. Each party's ownership rights will need to be defined in the DMP. Additionally, the DMP should set out who is responsible for managing the data. Management responsibilities may include recording, organising and archiving the data and managing access to the data. A comprehensive DMP will address the management roles of each party and will set out the formal levels of responsibility required for database management and maintenance.[99]

As explained above, data collection, access and reuse will be affected by legal controls. It is imperative that a DMP considers the legal and regulatory controls applying to the data that is collected by the research project. Such legal controls may include confidentiality restrictions for secret information, copyright assignments and licences, deposit agreements for inclusion of data in a database and agreements governing access to that database. All contractual obligations should be considered and addressed. In particular, a DMP should describe the conditions under which the research project is funded and any obligations – contractual or otherwise – that the researchers have to the funding body. Finally, a DMP should consider whether legislation applies to the collection or use of data, such as the application of privacy legislation for projects dealing with personal information.[100]

---

[99] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 250, 254.

[100] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 251–2.

Data security and sustainability are two important considerations for any DMP. The level of security that will operate in relation to the data collected will vary depending on the type of data concerned. For example, more stringent security may be applied to data that is confidential or which may form the basis of a patent application. For these types of data, access may be limited to select individuals (access may be password protected) and reuse rights may be minimal. Contractual agreements may regulate what disclosures can and cannot be made in relation to the data. For less sensitive data, the applicable security measures are likely to be less strict. A DMP will need to set out the different security measures relevant to the different levels of data and how these security measures are to be implemented.[101]

Careful consideration must be given to the potential future relevance of any data collected or generated by the research project. Where it is envisaged that data could be useful for future research, sustainability of data will be an important issue to address. A DMP should describe whether long-term preservation of the data is necessary and if so, how long the data will be preserved and who will be responsible for ensuring its preservation. A related issue will be how to ensure the ongoing, long-term funding of the database even after the research project that gave rise to the database is finished.[102]

Data management toolkits

A data management toolkit (DMT) is a document aimed at researchers within a research project, which provides practical guidelines about implementing the DMP. A DMT can assist individual researchers in ascertaining their role and level of responsibility within a research project and with understanding what is to be done with the data collected or generated by the project. A DMT can inform researchers about who will be able access the data collected by the researchers and how they may reuse that data. It can also assist researchers in determining their obligations, both legal and otherwise, in relation to the data that they generate or collect.

---

[101] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 252–4.

[102] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 255.

A DMT can be tailored to different levels of research and researchers. It may be appropriate to have a different DMT applying to researchers than that applying to database managers, or a different DMT applying to a small research team within a single institution than that applying to a larger research team that is part of a collaborative project spread across many institutions.

A DMT provides practical guidance to assist researchers in managing their data in compliance with the project's data management policies and procedures, DMP and the relevant legal framework. Therefore, a DMT should take the form most accessible to a project's researchers, whether this be in the form of a textual document, a series of questions, diagrams or multimedia tools. Yet regardless of form, all DMTs should enable researchers to understand the ownership and management issues surrounding data collection and compilation; the legal and technical restraints applying to collection, storage, handling and use of data; and the access, sharing, use and reuse framework surrounding the project's data.[103]

Licensing models

As far as data and databases attract copyright, licences can be used to allow access to and reuse of the data by other researchers. The emergence of open content licensing models has made it much easier for copyright owners to licence their material to a wider range of people, especially where it is distributed over the internet.[104] Open content licensing involves making copyright material available on liberal terms, to ensure that it is readily accessible and available for reuse.[105] The last few years have seen an increasing appreciation of open content licences to grant access to copyright-protected data collections in open collaborative research projects.[106]

---

[103] See A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 256–7.

[104] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 146.

[105] A Fitzgerald and K Pappalardo, *Building the Infrastructure* (2007) 146.

[106] A Fitzgerald and K Pappalardo, *Building the Infrastructure*, 148. See for example, Editorial, 'Let data speak to data' (2005) 438 *Nature* 531 <http://www.nature.com/nature/journal/v438/n7068/full/438531a.html>; Don Tapscott and Anthony Williams, 'The New Science of Sharing' (March 2007) *BusinessWeek.com* <http://www.businessweek.com/innovate/content/mar2007/id20070302_219704.htm?chan=technology_technology+index+page_more+of+today's+top+stories>;

The leading model of open content licensing is the suite of Creative Commons licences developed by the Creative Commons Project.[107] The Creative Commons (CC) Licences make copyright works freely available for use, on certain conditions as selected by the licensor. Where one or more elements of a database attracts copyright, a CC licence can be used to licence that copyright to users. For example, the CC licensing model is utilised by the Universal Protein Resource (UniProt), a comprehensive resource for protein sequence and annotation data and a collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (EBI).[108] UniProt has chosen to apply the Creative Commons Attribution-No Derivatives Licence to all copyrightable parts of its databases.[109]

Science Commons is a project related to Creative Commons that extends open access principles to scientific data and publications.[110] Formerly, the Science Commons 'Databases and Creative Commons FAQ' stated that a CC licence could be applied to copyrightable elements of a database, but advised database providers to:

- understand and make clear on the database website which elements of the database are licensed under the CC licence, based on the existence of copyright in those elements;
- understand and make clear on the database website which parts of the database are not subject to copyright (ie raw data and information) and which are therefore free to be used and reused independently of the CC licence;
- ensure that they have the necessary authority to apply a CC licence to the database (ie that they are the copyright owner or have permission from the copyright owner);
- where applicable, inform users that the CC licence only applies to the database elements and not the underlying software; and

---

Charlotte Waelde and Mags McGinley, 'Designing a licensing strategy for sharing and re-use of geospatial data in the academic sector' (2007) GRADE <http://edina.ac.uk/projects/grade>.

[107] See <http://www.creativecommons.org> or <http://www.creativecommons.org.au>.

[108] 'About UniProt' <http://beta.uniprot.org/help/about> at 22 April 2008.

[109] *License & disclaimer* <http://beta.uniprot.org/help/license> at 22 April 2008.

[110] See <http://sciencecommons.org>.

- be aware that CC licences do not licence all types of legal rights, but only licence copyright, and so legal restrictions relating to patents, privacy, confidentiality, contract and other relevant legal frameworks will not be affected by the adoption of a CC licence.[111]

Science Commons has since moved away from endorsing the application of CC licences to databases. The recommendation has been withdrawn because of difficulties identified by Science Commons with:

1. identifying the copyrightable and non-copyrightable elements of a database, such that obligations based on copyright (eg the option under some CC licences to require that use of the copyright material is non-commercial) are imposed in situations where copyright does not apply and the obligation is inappropriate; and
2. proper attribution (a requirement under all CC licences), where hundreds or even thousands of scientists have potentially contributed to or deposited data in the database.[112]

Science Commons has instead developed a 'Protocol for Implementing Open Access Data', which sets out the principles for open access data and provides a protocol for implementing those principles. Additionally, Science Commons distributes an Open Access Data Mark and metadata for use on data and databases that conform to the protocol. The protocol is not a copyright licensing model. Instead, the protocol requires a waiver of legal rights and all legal grounds for database protection in order to dedicate the data to the public domain.[113] Science Commons acknowledges that the protocol will not be appropriate for all

---

[111] See <http://sciencecommons.org/resources/faq/databases/> at 22 April 2008.

[112] Science Commons, *Protocol for Implementing Open Access Data* <http://sciencecommons.org/projects/publishing/open-access-data-protocol/> at 22 April 2008. See also Science Commons, *Database Protocol FAQ* <http://sciencecommons.org/projects/publishing/open-access-data-protocol/> at 22 April 2008.

[113] Science Commons, *Protocol for Implementing Open Access Data* <http://sciencecommons.org/projects/publishing/open-access-data-protocol/> at 22 April 2008. See also Science Commons, *Database Protocol FAQ* <http://sciencecommons.org/projects/publishing/open-access-data-protocol/> at 22 April 2008.

types of data, but believes that the protocol offers a system that is both legally accurate and easier for scientists to understand than many copyright licensing models.[114]

The problems identified by Science Commons are indeed apparent in many situations involving licensing of database elements and are worthy of careful consideration. However, the legal position regarding copyright protection of data and databases is much more straightforward in Australia than in either the United States or Europe. It is considerably easier to distinguish between the copyright and non-copyright elements of databases in Australia than in the United States, where the creativity of a compilation must be assessed before copyright applies. Further, there are fewer legal considerations in Australian than in Europe, where a sui generis database right operates to protect databases irrespective of whether the database or its contents attract copyright protection.

The concerns raised by Science Commons highlight the importance of each and every research project adopting a DMP that properly considers and manages issues of data ownership and legal rights including copyright. It is entirely possible to successfully apply a copyright licensing model to a database and its copyrightable contents. However, in order to ensure the successful operation of the licence, the research project's DMP must clearly identify which legal rights apply to which database elements and which database elements are to be licensed to the public and on what terms. The DMP should also state how the data and database are to be attributed and make this information readily apparent on the database website. For example, UniProt provides a webpage that informs users how to cite resources and publications obtained from the UniProt website or databases under a CC licence.[115]

---

[114] Science Commons, *Protocol for Implementing Open Access Data* <http://sciencecommons.org/projects/publishing/open-access-data-protocol/> at 22 April 2008. See also Science Commons, *Database Protocol FAQ* <http://sciencecommons.org/projects/publishing/open-access-data-protocol/> at 22 April 2008.

[115] *Publications – How to cite us* <http://beta.uniprot.org/help/publications> at 22 April 2008.

CONCLUSION

For any research project, several important legal and management decisions will need to be made about the data collected or generated in the course of the research. How will ownership interests in data be determined and allocated? Will data be made accessible to the public, and if so, on what basis? Will sharing and reuse of data be permitted? What legal restraints apply to the data? All these questions must be carefully considered, answered and agreed upon by members of the research project, including researchers, database managers, hosting institutions and funding bodies.

Different bodies of law – copyright, patent, privacy, confidentiality and contract law – will be relevant to the collection, storage and dissemination of data. Proper management of data requires an understanding of how these legal regimes impact on the data's generation, handling and dissemination. By adopting mechanisms such as data management policies, plans and toolkits, researchers and research organisations can effectively manage the data they collect or generate, based on a practical understanding of how the various legal regimes apply to it. Implementation of such measures will ensure that research data can be made available online to other researchers in a manner that is openly accessible, timely and in compliance with legal requirements.