

# FIELD, FILE, DATA, CONFERENCE: TOWARDS NEW MODES OF SCHOLARLY PUBLICATION

**Ross Coleman**

*Sydney eScholarship*

*University of Sydney Library*

New methods of e-research practice are overtaking and destabilising the traditional processes of scholarly publication. These traditional modes of publication do not adequately recognise current forms of research practice and methods of expressing research outcomes. This inability to validate new forms of research expression as publication points to a growing structural deficiency in handling and recognising the new outputs and product of the emerging e-research programs. The consequences of this deficiency will be debilitating for e-research practice.

As a contribution to this debate this paper examines infrastructure models that manage the 'whole life cycle of primary data' from ingest to storage, access and publication (considering the Sydney eScholarship program and the publication of this conference as an example); current definitions of research publication; and the need and suggested requirements for data sets (as outputs from e-research programs) to be recognised as research publications.

## Introduction

The words that set the introduction to this conference also provide the context audience for this paper:

researchers ... using digital methods for the whole life cycle of their primary data, from capture to organisation, submission to a repository or archive, and later access and dissemination in publications, teaching resources and conference presentations.  
(University of Sydney Faculty of Arts, 2006)

New and emerging methods and forms of research practice—as evidenced in this conference—have overtaken and destabilised the traditional modes and processes of scholarly publication. This imbalance in the research cycle is creating tensions in the work practices and priorities of active researchers.

These new research practices have been facilitated by information technologies, web and grid networks, new research tools and repositories that provide new ways of conceptualising, gathering, representing, storing, analysing, accessing and publishing primary data. It is application of these technologies that has both challenged traditional forms (and ownership) of publication and also provided the means to develop new modes of scholarly communication. The development of digital repositories and archives, new means of peer assessment, accessibility through open archives compliant web services and new publication technologies have provided the opportunities to address this imbalance in the research cycle and develop new kinds of ownerships and relationships between research and publication.

This paper explores two particular issues: forms of infrastructure that facilitate 'life cycle' management, and the requirements for data to become a recognised form of research publication.

New kinds of infrastructure services that integrate repository, access and publication processes provide a means of building these new relationships between research and publication, and a very simple example is the organisation and of publication of this very conference, facilitated through the Sydney eScholarship service.

More difficult—and certainly not fully resolved in this paper—is the recognition of forms or sets of data as valid research publication. Quality criteria and even the definition of a research publication need to be re-assessed—though perhaps not as much as may seem.

## Defining a Research Publication

There can be many definitions of research publication, but in Australia there is a formal definition, one that helps determine Commonwealth research funding to universities. This relatively prescriptive definition determines criteria for official recognition and for the publication forms themselves.

The definition is set out in the Australian Department of Education, Science and Training (DEST) draft publication *2007 Higher Education Research Data Collection (HERDC), specifications for the collection of 2006 data* (Australian Department of Education Science and Training [DEST], 2006).

While one can decry any bureaucratic process of reducing the definition of research to administrative criteria, it is the process that determines recognition and reward for one aspect of individual research activity. The reward takes the form of a system of points, depending on the type of publication (one point for a journal article or conference paper, five for a book, and so forth, as outlined in the HERDC specifications). As evidence of staff research activity, these points translate into a component of research quantum funding to universities from DEST.

The DEST/HERDC definition of research is reproduced here in Appendix 1, and the requirements for peer review in Appendix 2. The specifications define a research publication as follows:

For the purposes of these specifications, research publications are [defined as] books, book chapters, journal articles and/or conference publications which meet the definition of research [reproduced in this paper as Appendix 1] and are characterised by:

- substantial scholarly activity, as evidenced by discussion of the relevant literature, an awareness of the history and antecedents of work described, and provided in a format which allows a reader to trace sources of the work, including through citations and footnotes
- originality (that is, not a compilation of existing works)
- veracity/validity through a peer validation process [see Appendix 2] or by satisfying the commercial publisher processes
- increasing the stock of knowledge
- being in a form that enables dissemination of knowledge. (DEST, 2006, p. 9).

The specifications go on to provide greater detail and guidance in regard to those particular recognised forms of publication (books, book chapters, journal articles, and conferences).

Staying with our theme, to be included in the conference category

... the conference publication must meet the definition of research (Appendix 1) as amplified in the general requirements for research publications [above] and must:

- be published in full. The papers may appear in a number of different formats: for example, as a volume of proceedings, a special edition of a journal, a normal issue of a journal, a book or a monograph, CD-ROM or conference or organisational web site;
- be peer reviewed (Appendix 2 to this paper);
- be presented at conferences, workshops or seminars of national or international significance;
- the author must be affiliated with the claiming HEP (Higher Education Provider)...

Keynote addresses and invited papers may be included where all other papers for the conference are peer reviewed. HEPs must retain in verification material evidence of the keynote status of the address (for example, the contents page) and of other contributions to the conference being peer-reviewed (for example, a statement in the introduction to proceedings indicating this).

The types of conference publications that are unlikely to meet the criteria include papers that appear only in a volume handed out to conference participants. (DEST, 2006, p. 28)

These specifications are directed primarily toward institutional research office staff who collect publication data and verify it against the criteria set out in the specifications for reporting purposes.

While these are in some ways bureaucratic criteria—and subject to administrative verification and audit—the definition of research is reasonably sound, as are those relating to the traditional forms of publication. These criteria have been tested over time, and though there has been a bit of experimentation with recognising other forms of publication (such as exhibitions for a brief time), they continue as the traditional approach, partly for historical purposes, partly because of difficulties in providing comparative criteria with new forms of, mostly, non-textual research output.

These specifications do accept electronic publication, as long as it meets research and publication criteria:

Electronic works are eligible to be counted, provided they meet all the relevant criteria in these specifications ... for the publications category against which they are being claimed' (DEST, 2006, p. 32).

Notwithstanding this, however, the definitions are based solely on traditional modes of print publication and do not recognise current forms of research practice and methods of recording and expressing research outcomes. This inability to validate new forms of research expression as publication points to a growing structural deficiency in handling and recognising the new outputs and product of the emerging e-research programs. The consequences of this deficiency will be debilitating for e-research practice.

The use of these criteria and definitions in the proposed Research Quality Framework (RQF) assessment is yet to be clarified, but will most probably be applied in some way in regard to the 'quality' aspects of that assessment. Assessment for impact will be much broader.

In comparison the UK Research Assessment Exercise (RAE) 2008 (UK Research Assessment Exercise 2008, n.d.) takes a much broader approach in terms of recognised research outputs. Staff can nominate four research outputs. These can be: books, chapters in books, articles in journals, conference contributions, or other formats including creative media and multimedia, standards documents, patents, products and processes, items of software, or technical reports, including consultancy reports. The RAE also recognises 'transient outputs' (such as performances) and 'changing research content', which 'do not exist in a stable form and their research content may change over time by virtue of the nature of the medium through which they are disseminated' (includes research data sets and databases). These additional outputs provide a much more appropriate recognition of current research practices and methods.

## Mode of Publication for Sustainable Data from Digital Fieldwork: From Creation to Archive and Back

The publication for this particular conference will meet the HERDC criteria for a research publication. But the publication process is also distinguished by the general integration of organisation, review, storage and publication as one operation. This integration, while perhaps as not as seamless as it could be, is characterised by use of several different systems and processes that move and refine the papers into forms that are reviewed and accepted, archived and then published. These systems include:

- OCS (Open Conference System) (Public Knowledge Project, n.d.) for organisation, registration, and paper review;<sup>1</sup>

- uploading of final accepted papers into an open access institutional repository as a collection (the Sydney eScholarship repository,<sup>2</sup> a DSpace installation) for immediate announcement and discovery through the web. The collection and each paper will have persistent citable URLs;
- consolidation of the papers (with appropriate introduction) into a more formally published form through Sydney University Press for distribution through print-on-demand or in PDF form. This publication will meet DEST requirements. As a digitally archived publication with print-on-demand and sale capacity these conference papers will remain citable and not go out of print.

There is nothing particularly radical or ground-breaking about this process, except that it could only be facilitated in this way through the existence of an infrastructure that provides such integration between research and publication.

That infrastructure is Sydney eScholarship, initiated and managed by the University of Sydney Library. The Sydney eScholarship<sup>3</sup> program is a suite of innovative services that integrates the management of digital content with new forms of access and scholarly publication. The program combines digital library collection creation, digital archiving, repository content management and project consultancy with new and hybrid forms of publication, including Sydney University Press.<sup>4</sup> Sydney eScholarship provides a unique capability and capacity to integrate and add value and impact across the life cycle of digital content.

Sydney eScholarship content is archived and maintained by the University Library as part of the Sydney Digital Library collections, a digital archive and content management service striving to meet the key aspects of the attributes and responsibilities of a Trusted Digital Repository (Research Libraries Group, 2002)<sup>5</sup> (discussed further below). Sydney eScholarship is an active partner in the Australian Partnership for Sustainable Repositories (APSR).<sup>6</sup>

It is this capability (and the skills and technologies surrounding it) that provides the potential to address issues around 'whole of life cycle' management of digital content, and provide a platform for facilitating new modes of publication, such as data sets.

## Data as Research Publication

The need for data collection to be recognised as a valid form of research publication goes beyond recognition of an individual research activity (though this is critical). Such recognition is necessary to encourage and facilitate the development of sustainable and managed data sets as a legitimate e-research activity/output in its own right. It acknowledges the importance of such activity as a valid counted (for DEST purposes) research product. It goes to the heart of the need for discovery and access to primary research data – data that is unique and re-usable.

The need to recognise research data publication as an incentives model has been identified by a number of European scientific researchers (Klump et al., 2006) as essential to promote scientific knowledge, particularly in the terms of the 2003 'Berlin Declaration on Open Access to Knowledge on the Sciences and Humanities' (Max Planck Gesellschaft, 2003). To facilitate recognition of data as publication the authors argue two main criteria are necessary:

- persistence, for reliable citation over time;
- quality standards, including such aspects as credibility, usability and interpretability. (These standards also reference the ISO 9000 standards for quality management systems (QMS))

They also argue the need for a reliable licence model to protect the intellectual property rights of the researchers, citing the Science Commons project<sup>7</sup> in applying the principles of the Creative Commons licence model to the sciences.

## Suggested Requirements for Data as Research Publication

It is possible to begin to consider a set of requirements and criteria against which data sets and collection can be measured and assessed as a valid form of research publication.

The DEST/HERDC definition of research (Appendix 1) specifically excludes 'general purpose or routine data collection' as part of research. Indeed for a data set to be substantial enough to meet any research criteria it would need to provide elements beyond the 'routine'—such as originality, peer validation, and so forth—that are specified as being of a research publication. Elements such as persistence, citability, technical and metadata standards, rights and permissions, and means of discovery or dissemination need also to be present. These are also some of the

requirements necessary for any serious consideration of sustainability of data over the long term.

These requirements need to exist in a sustainable organisational context. This organisational context can be assessed through the attributes of a Trusted Digital Repository (Research Libraries Group, 2002), which include:

- OAIS (Open Archival Information System)—compliance with the system models and frameworks developed to ensure common approaches and principles are used across repository systems;
- Administrative responsibility—ensuring that repositories provide best practice in operational management, including physical environment, as well as the capacity to manage permissions, rights, and so forth;
- Organisational viability—a demonstrated organisational commitment for the long-term management and retention of data;
- Financial sustainability—including the development of sustainable business and financial planning, including budget strategies and risk assessment;
- Technological suitability—plans and strategies for technical preservation and systems renewal over time;
- System security—technical security matters including disaster recovery, back-up, appropriate authentication systems, and so forth;
- Procedural accountability—documentation of policies and procedures, and so forth.

Many of these attributes tend more to ambition than achievement, and may only be attained in part. But the significance of this checklist is as a framework for managers to make some assessment of the context in which they need to provide sustainable support for research data and, in turn, provide that assurance to researchers.

The DEST definition of a research publication can be adapted outside of those traditional—mostly textual—publication forms or categories (be they in print or electronic form). Figure 1 identifies, at first cut, a range of requirements that could meet publication criteria and be applied to the development of data sets as a recognised research activity and as a new form of publication output.



These requirements may raise many practical questions and many researchers could add other discipline specific standards and requirements. However Figure 1 does indicate it is possible to develop an acceptable set of requirements that would provide defensible criteria for recognition as a research publication.

<i>DEST research publication criteria</i>	<i>Data set publication recognition: suggested equivalent requirements</i>
substantial scholarly activity, as evidenced by discussion of the relevant literature, an awareness of the history and antecedents of work described, and provided in a format which allows a reader to trace sources of the work	<ul style="list-style-type: none"> <li>• credibility of the researchers</li> <li>• authority of the platform/organisation (in other words, the publisher)</li> <li>• significance of the subject matter</li> <li>• conceptualisation of the data collection</li> <li>• meeting standard requirements for data and metadata (descriptive, technical, provenance, and so forth)</li> <li>• relationship/linkage to other data sets</li> <li>• persistent citeability</li> </ul>
originality (that is, not a compilation of existing works)	<ul style="list-style-type: none"> <li>• unique data collection</li> <li>• replicated data necessary for testing or verification</li> </ul>
veracity/validity through a peer validation process or by satisfying the commercial publisher processes	<ul style="list-style-type: none"> <li>• use of recognised data and metadata standards</li> <li>• peer review process for data inclusion</li> <li>• credible/authoritative review panel</li> <li>• usability/functionality for research community</li> </ul>
increasing the stock of knowledge	<ul style="list-style-type: none"> <li>• unique primary data</li> </ul>
being in a form that enables dissemination of knowledge	<ul style="list-style-type: none"> <li>• persistence of citation</li> <li>• being an identifiable set of data for citation purposes</li> <li>• IP licence model</li> <li>• OAI compliance for harvesting (OAI-PMH)</li> </ul>

Figure 1: Suggested criteria for recognising data sets as publication.

## Conclusion

The dynamic nature of research methods, tools, outputs and expressions being developed in many areas of e-research, typified by those in ethnography and the humanities, requires not just new infrastructures, but

new ways in which such activity is better and more appropriately recognised as scholarly publication.

The recognition of quality requirements for data sets (as in the table above) as part of the criteria for research publication will ensure researcher compliance with best practice in the collection, description and sustainable management of that data. However data sets will need to be described in forms of standard citation to meet such requirements.

The role of a 'publisher' (and that of ownership) in this process is also an open question. The emergence of library-based e-presses and eScholarship programs such as at Sydney,<sup>8</sup> the development of systems such as OJS and OCS facilitating effective small scale electronic publishing, the growth of global discipline-based research networks and archives<sup>9</sup> all provide flexible scholarly publishing platforms.

Fundamental to these platforms being effective is a functional infrastructure that will provide the kinds of capacities to meet the archival, citation, access and publication requirements of these new kinds of research practice and expression. Infrastructures need to persist as long term trusted and reliable services.

Infrastructure requirements for e-research vary by scale, complexity, user and performance needs, and an integrated service such as Sydney eScholarship occupies one niche in this range. This service works at the small to medium data collections level, and like all new services is still building and evolving the business, administrative, operational and the strategic processes and systems that will allow it to meet its role in this environment, within and beyond the institution.

E-research programs will not only generate new, more, massive and complex types of research data, they will also create new forms of scholarly expression, beyond the textual. One form of expression can be data sets themselves. If these expressions—properly meeting research publication criteria—are not recognised as forms of publication, as incentives, then the full potential of e-research will not be realised.

## Endnotes

<sup>1</sup> The Open Conference System (OCS) like the Open Journal System (OJS) has been developed by the Public Knowledge Project (PKP) based at the University of British Columbia. PKP has developed free, open source software for the management,

publishing, and indexing of journals and conferences. Release of version 2 of OCS is due in Jan 2007.

<sup>2</sup> Sydney e-Scholarship Repository website: <http://ses.library.usyd.edu.au> (University of Sydney Library, 2006b).

<sup>3</sup> Sydney e-Scholarship website: <http://escholarship.usyd.edu.au> (University of Sydney Library, 2006a).

<sup>4</sup> Sydney University Press website: <http://www.sup.usyd.edu.au>. (University of Sydney Library (2006c). Sydney University Press is recognised by DEST as a commercial publisher.

<sup>5</sup> Developed through a collaboration led by the Research Libraries Group (RLG). 'A trusted digital repository is one whose mission is to provide reliable long-term access to managed digital resources to its designated community, now and in the future' (Research Libraries Group, 2002, p i).

<sup>6</sup> Australian Partnership for Sustainable Repositories website: <http://www.apsr.edu.au> (Australian Partnership for Sustainable Repositories, n.d.).

<sup>7</sup> *Science Commons*, n.d.

<sup>8</sup> Many university libraries have initiated electronic publishing programs, some as named e-presses (for example, Monash, ANU, UTS in Australia), some as scholarly publishing programs (for example, Michigan, Cornell), others as integrated eScholarship programs (for example, California, Sydney).

<sup>9</sup> Groups such as SSRN, the Social Science Research Network 'devoted to the rapid worldwide dissemination of social science research' can potentially function as publishers (website: <http://www.ssrn.com>).

## References

Australian Department of Education, Science and Training [DEST]. (2006). *2007 Higher Education Research Data Collection: specifications for the collection of 2006 data*. Retrieved October 10, 2006, from [http://www.dest.gov.au/sectors/research\\_sector/online\\_forms\\_services/higher\\_education\\_research\\_data\\_collection.htm](http://www.dest.gov.au/sectors/research_sector/online_forms_services/higher_education_research_data_collection.htm)

Australian Partnership for Sustainable Repositories [APSR]. (n.d.) *Australian Partnership for Sustainable Repositories*. Retrieved October 10, 2006, from <http://www.apsr.edu.au>

Klump J., Bertelmann R., Brase J., Dipenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., Schindler, U., Sens, I. & Wächter, J. (2006). Data publication in the Open Access Initiative. *Data Science Journal*, 5, 79-83.

Max Planck Gesellschaft. (2003). *Open Access Conference - Berlin Declaration*. Retrieved October 10, 2006, from <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

Public Knowledge Project, University of British Columbia. (n.d.). *Open Conference System*. Retrieved October 10, 2006, from <http://pkp.sfu.ca/ocs>

- Research Libraries Group. (2002). *Trusted Digital Repository Attributes and Responsibilities*. Mountain View, CA: Research Libraries Group. Retrieved October 10, 2006, from <http://www.rlg.org/legacy/longterm/repositories.pdf>
- Science Commons*. (n.d.). Retrieved October 10, 2006, from <http://sciencecommons.org>
- United Kingdom Research Assessment Exercise 2008. (n.d.). *Research—information about the collection and storage of research outputs for the 2008 RAE*. Retrieved October 10, 2006, from <http://www.rae.ac.uk/aboutus/policies/default.htm>
- University of Sydney Faculty of Arts. (2006). *Sustainable Data from Digital Fieldwork Conference Website*. Retrieved October 10, 2006, from <http://conferences.arts.usyd.edu.au/index.php?cf=11>
- University of Sydney Library. (2006a). *Sydney e-Scholarship Program*. Retrieved October 10, 2006, from <http://escholarship.usyd.edu.au>
- University of Sydney Library. (2006b). *Sydney E-Scholarship Repository*. Retrieved October 10, 2006, from <http://ses.library.usyd.edu.au>
- University of Sydney Library. (2006c). *Sydney University Press*. Retrieved October 10, 2006, from <http://www.sup.usyd.edu.au/>

## Appendix 1

(Excerpted from DEST, 2006, p. 8)

### Research

For the purposes of these specifications, research comprises:

- creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications<sup>1</sup>
- any activity classified as research which is characterised by originality; it should have investigation as a primary objective and should have the potential to produce results that are sufficiently general for humanity's stock of knowledge (theoretical and/or practical) to be recognisably increased. Most higher education research work would qualify as research
- pure basic research, strategic basic research, applied research and experimental development.

Activities that support research and meet this definition of research include:

- provision of professional, technical, administrative or clerical support and/or assistance to staff directly engaged in research
- management of staff who are either directly engaged in research or are providing professional, technical or clerical support or assistance to those staff
- activities of students undertaking postgraduate research courses
- development of postgraduate research courses
- supervision of students undertaking postgraduate research courses.

Activities that do not support research must be excluded, such as:

- preparation for teaching
- scientific and technical information services
- general purpose or routine data collection
- standardisation and routine testing
- feasibility studies (except into research and experimental development projects)
- specialised routine medical care
- commercial, legal and administrative aspects of patenting, copyright or licensing activities
- routine computer programming, systems work or software maintenance (research and experimental development into applications software, new programming languages and new operating systems would normally meet the definition of research).

## Appendix 2

(Excerpted from DEST, 2006, p. 30)

### Peer review

For the purposes of the HERDC, an acceptable peer review process is one that involves an assessment or review of the research publication in its entirety before publication by independent, qualified experts. Independent in this context means independent of the author.

Peer review is relevant for journal articles and conference publications being counted in the Research Publications Return – Return 2.

For journal articles, any of the following are acceptable as evidence of peer review:

- the journal is listed in one of the Institute for Scientific Information indexes ([www.isinet.com/journals](http://www.isinet.com/journals))
- the journal is classified as 'refereed' in Ulrich's International Periodicals Directory (Volume 5—Refereed Serials) or via Ulrich's web site [www.ulrichsweb.com](http://www.ulrichsweb.com)
- the journal is included in the department's Register of Refereed Journals (*no longer maintained*)
- there is a statement in the journal which shows that contributions are peer reviewed
- there is a statement or acknowledgement from the journal editor which shows that contributions are peer reviewed
- a copy of a reviewer's assessment relating to the article.

*Note: a statement from an author that a publication was peer reviewed is not acceptable. The existence of a national or international advisory board is also not sufficient evidence that all relevant publications were assessed by members of it.*

For books and book chapters, the concept of a commercial publisher is used as a surrogate test of quality in place of a peer review requirement.