

Growing Institutional Support for Data Citation: Results of a Partnership Between Griffith University and the Australian National Data Service

Author

Simons, N, Visser, K, Searle, S

Published

2013

Journal Title

D-Lib Magazine

DOI

[10.1045/november2013-simons](https://doi.org/10.1045/november2013-simons)

Rights statement

© The Author(s) 2013. This is the author-manuscript version of this paper. It is posted here with permission of the copyright owners for your personal use only. No further distribution permitted. For information about this journal please refer to the journal's website or contact the authors.

Downloaded from

<http://hdl.handle.net/10072/54418>

Griffith Research Online

<https://research-repository.griffith.edu.au>

D-Lib Magazine

November/December 2013

Volume 19, Number 11/12

Growing Institutional Support for Data Citation

Results of a Partnership Between Griffith University and the Australian National Data Service

Natasha Simons

Griffith University, Brisbane, Australia

n.simons@griffith.edu.au

Karen Visser

Australian National Data Service, Canberra, Australia

karen.visser@ands.org.au

Sam Searle

Griffith University, Brisbane, Australia

samantha.searle@griffith.edu.au

doi:10.1045/november2013-simons

Abstract

Data is increasingly recognised as a valuable product of research and a number of international initiatives are underway to ensure it is better managed, connected, published, discovered, cited and reused. Within this context, data citation is an emergent practice rather than a norm of scholarly attribution. In 2012, a data citation project at Griffith University funded by the Australian National Data Service (ANDS) commenced that aimed to: enhance existing infrastructure for data citation at the University; test methodologies for tracking impact; and provide targeted outreach to researchers about the benefits of data citation. The project extended previous collaboration between Griffith and ANDS that built infrastructure at the University to assign DOI names (Digital Object Identifiers) to research data produced by Griffith's researchers. This article reports on the findings of the project and provides a case study of what can be achieved at the institutional level to support data citation.

Introduction

There is a growing recognition that research data is a first class output of research and, as such, needs to take its place amongst more traditional measures of scholarly endeavours, such as journal articles. The amount of money poured into research on a global scale is significant and major funders are increasingly recognising data as a valuable product of research. As the volume and complexity of data continues to grow, various governments and research councils have responded to the "data deluge" by outlining policy and principles that support better management of research data and by providing funding to assist research institutions in achieving this goal. The result is a range of international initiatives that have a shared goal of ensuring that research data can be well managed, connected, published, discovered, cited and reused.

Within the context of improved management of data produced through research, data citation is a relatively new concept. Data citation refers to the practice of citing research datasets and collections in the same way that other types of information, such as articles and books, are cited. Citation is the norm for scholarly acknowledgement of publications and as a first class output of research, ideally data would be treated in the same way. In addition to providing a reward structure for sharing data, data citation allows for the identification, retrieval, replication and verification of data underlying published studies [1]. The international not-for-profit organisation [DataCite](#) suggests,

"Data citation can help by: enabling easy reuse and verification of data; allowing the impact of data to be tracked; creating a scholarly structure that recognises and rewards data producers."

Over the past few years, various governments and funding agencies have included or expanded references to data management, access and more recently, citation, in their policies and programs. For example, the Research Councils UK state in their "Common Principles of Data Policy" that "all users of research data should acknowledge the sources of their data" [2] and the National Science Foundation (US), now allows for citable data to be listed as a product of research, like a journal article [3].

The [Dryad Data Repository](#), in partnership with various journals, provides a framework for data deposit and data citation in conjunction with article publication. However, this type of initiative is more the exception than the rule. At the level of cross-disciplinary research institutions, such relationships with journals are rare and developing policy and improving technical infrastructure to enable routine data deposit and long-term data management is the much-needed prerequisite to data citation. On a global scale, a cultural shift is required within the scholarly community in order for data citation to become the norm. As the practice of data citation evolves, partnerships and shared learning among institutions and between countries are critical. An outstanding example of collaboration and shared learning is the partnership between the [Australian National Data Service](#) (ANDS) and Griffith University in developing infrastructure, assessing impact measurement tools, and engaging with researchers to support data citation practices.

Australian initiatives in data management and citation

Enabling support for researchers and institutions in building a culture of data citation is an objective of ANDS. An initiative funded by the Australian Government, ANDS is building the [Australian Research Data Commons](#): a cohesive collection of research resources from all research institutions, to make better use of Australia's research data outputs. [Research Data Australia](#), ANDS' flagship service, provides a comprehensive window into the Australian Research Data Commons with a rapidly growing collection of almost 90,000 Australian records of Australian research data collections. This discovery service is designed to provide rich connections between data, projects, publications, researchers and institutions, and promote visibility of Australian research data collections in search engines. Within the overall goals of the organisation, "An important aim of ANDS is to enable more researchers to re-use research data more often. To achieve this aim ANDS is engaged in activities that will make it easier to share data, to recognise the importance of making data available and to make data citation a standard procedure." [4]

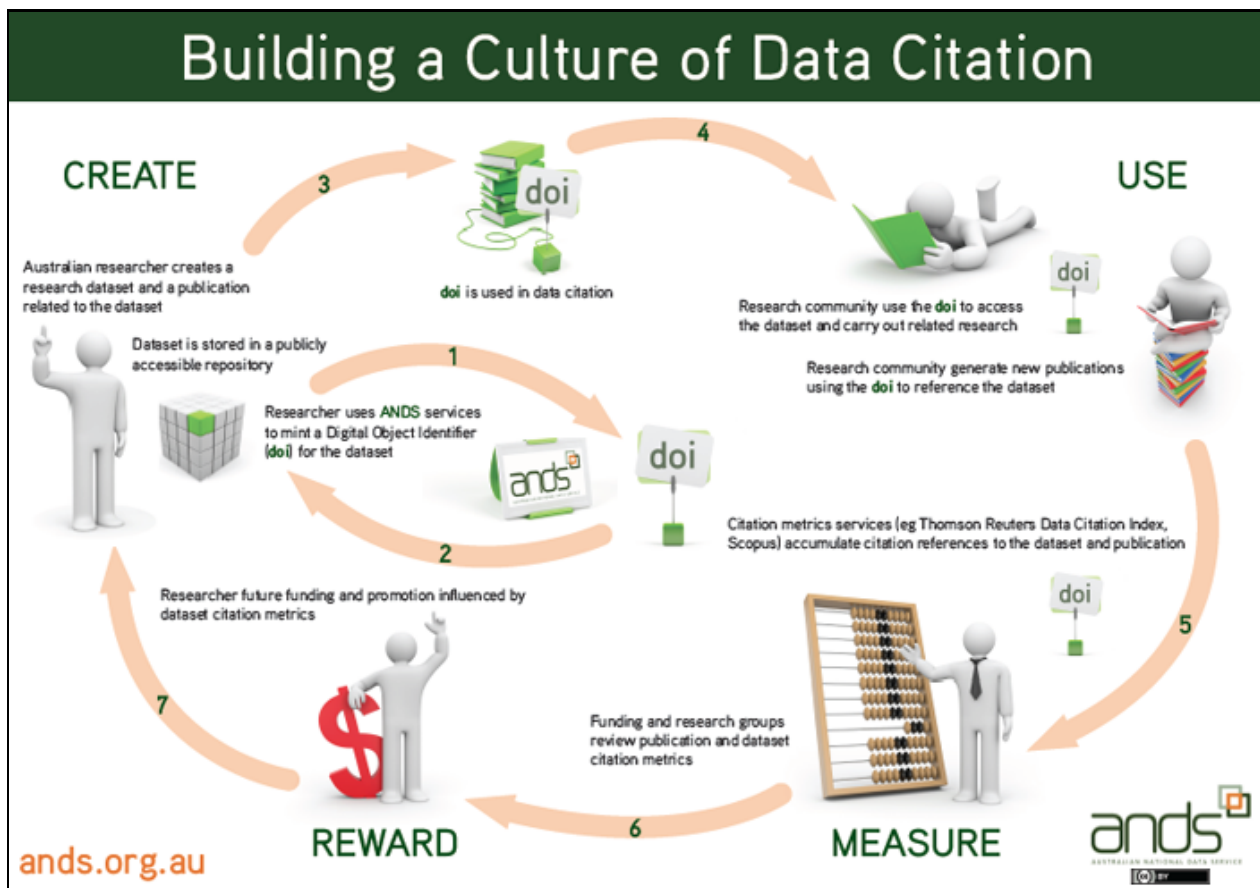


Figure 1: Data Citation Poster by the Australian National Data Service
[\[Larger version of Figure 1.\]](#)

To assist institutions in managing persistent access to data, and facilitating data citation practices, ANDS provides a "Cite My Data" service. This machine-to-machine service allows ANDS-partner institutions to mint DOI® names ([Digital Object Identifiers](#)) free of charge for their datasets using DataCite (of which ANDS is a partner) as the DOI Registration Agency. This process has been facilitated by the provision of extensive support materials on the ANDS website, including step-by-step guides and responses to FAQs from Australian research institutions. ANDS has also facilitated the growth of an Australian Data Citation Community of Practice by sponsoring a series of workshops, meetings and webinars that has drawn on the experience of the international and local DOI and data citation community. Further, ANDS has provided funding to a small number of institutions to assist them in developing infrastructure and guidance for researchers that supports building a culture of data citation.

ANDS and Griffith University: a case study in collaboration

In 2012, Griffith University's Division of Information Services began a new project funded by the Australian National Data Service that aimed to: enhance existing infrastructure for data citation at the University; test methodologies for tracking impact; and provide targeted outreach to researchers about the benefits of data citation. The project, known as the data citation project, extended previous collaboration between Griffith and ANDS that built infrastructure at the University to assign DOIs to research datasets and data collections produced by Griffith's researchers. Concluding in May 2013, the data citation project was the first of its kind at an Australian university, though similar initiatives have been made at other Australian research institutions such as the CSIRO and the Australian Antarctic Data Centre. The project generated new experiences and findings regarding data citation infrastructure, impact and practice.

Developing DOI infrastructure

DOIs are globally unique, resolvable persistent identifiers that contribute an important component to the research data infrastructure. Persistent identifiers are critical in managing access to online resources so that links are not broken and resources are not lost. There are a large number of persistent identification schemes available for use and it is useful to select which ones to use against criteria that include uniqueness, trustworthiness, reliability, scalability, flexibility, and transparency to users of the scheme.

Within the global context of improving management of, and access to, research data, there is a growing international effort to improve citation of research data using the DOI system. DataCite is the international not-for-profit organisation playing a leading role in this effort; and as a partner of DataCite, ANDS are also key contributors. DataCite promote the use of DOIs in data citation as a way of: helping researchers track reuse of their data; helping data centres in establishing a mechanism that supports discovery and reuse; and supporting publishers with an elegant link between an article and its underlying data [5].

The benefits of assigning DOIs to datasets and data collections also extend beyond their value in the context of data citation. Assigning DOIs to research data collections enhances the concept of data as being a valued research output, to be managed persistently for the long-term. DOIs require a commitment to maintaining links to the data and therefore signal an institution's willingness to manage the data for the foreseeable future. DOIs are also routinely assigned to publications, in fact the system began in the publishing industry. When applied to data, they indicate that data is to be treated with the same respect as publications – to be well managed, persistently available and cited over the scholarly lifecycle of the research. Finally, DOIs are key to the collection of citation metrics and altmetrics. While citation metrics track formal citations, altmetrics tools such as [ImpactStory](#) use DOIs to track mentions in social media and non-traditional scholarly communications across the web. Without a DOI, this tracking is made more difficult.

Developing resources, policy and procedures for DOIs

Since September 2011, ANDS and Griffith University have worked closely together to ensure that emergent practices associated with DOIs attached to Australian research datasets are aligned with world practices and are, in fact, setting standards for this global movement. Griffith University was the first to test the Cite My Data service, providing feedback and advice to ANDS and contributing their experiences and ideas to the emerging data citation community of practice.

The Griffith experience revealed that minting DOIs was technically straightforward, however it raised a number of questions such as: what material should have a DOI, how to manage versioning, level of granularity, landing page resolution, metadata requirements and data citation format [6]. Griffith sought answers to these questions from ANDS, as well as the national and international communities of practice. This resulted in the development of the *"Digital Object Identifiers (DOIs): Introduction and Management Guide"*, a document that provides a framework for minting and maintaining DOIs at the institution. The target audience for the document is internal; it begins with an overview of the DOI system, data citation and the ANDS service. The DOI management guide section outlines the business rules for minting DOIs and makes clear the agreed approach to granularity, versioning, citation format and so on. The document includes a section on the DataCite metadata schema, as a minimal amount of metadata is required to mint a DOI, and concludes with a technical summary that points to the scripts developed to mint DOIs for Griffith data collection and to information on the ANDS service.

Drawing heavily on the Griffith University experience, ANDS developed an extensive matrix of publicly available, support materials for both data citation and DOIs. These resources include general technical and informational materials as well as in-depth materials on specific topics. The materials are presented in a variety of formats including "how to" checklists, Guides, current examples of national and international practice, academic articles, webinars, recordings and linked webpages. Several webinars featured international perspectives and speakers, such as Dr Heather Piwowar (ImpactStory), Dr Louise Corti (UK Data Archive) and Ryan Scherle (Dryad). As data citation is of interest to both researchers and institutions, and sometimes for quite different reasons, there are resources specifically for researchers [7] and institutions [8].

Once a DOI has been assigned, it is used as the web link in a citation element that is included in the metadata record describing the data collection. For example:

Cara Beal, Rodney Anthony Stewart ([2013]): South East Queensland Domestic Water Usage Collection (1st Spring 2012 Read). Smart Water Research Centre. <http://doi.org/10.4225>

[/01/513E57E0F1577.](#)

This record is made available for discovery in the [Griffith Research Hub](#). The Hub has addressed the need for a comprehensive view of the institution's research output and contains profile pages for researchers and their associated publications, projects, collections, groups and so on. The citation element is formed according to the basic DataCite guidelines for a data citation [9] and it is included in the data collection metadata record that Griffith then provides to the [ANDS Research Data Australia](#) discovery portal.

Griffith has a roadmap for future DOI-related activities including: further work to embed DOI minting into automated workflows for data deposit in the repository; assigning DOIs to grey literature such as theses; a watching brief on a number of international developments including the [ODIN](#) project and altmetrics tools; and a future review of guidelines and rules.

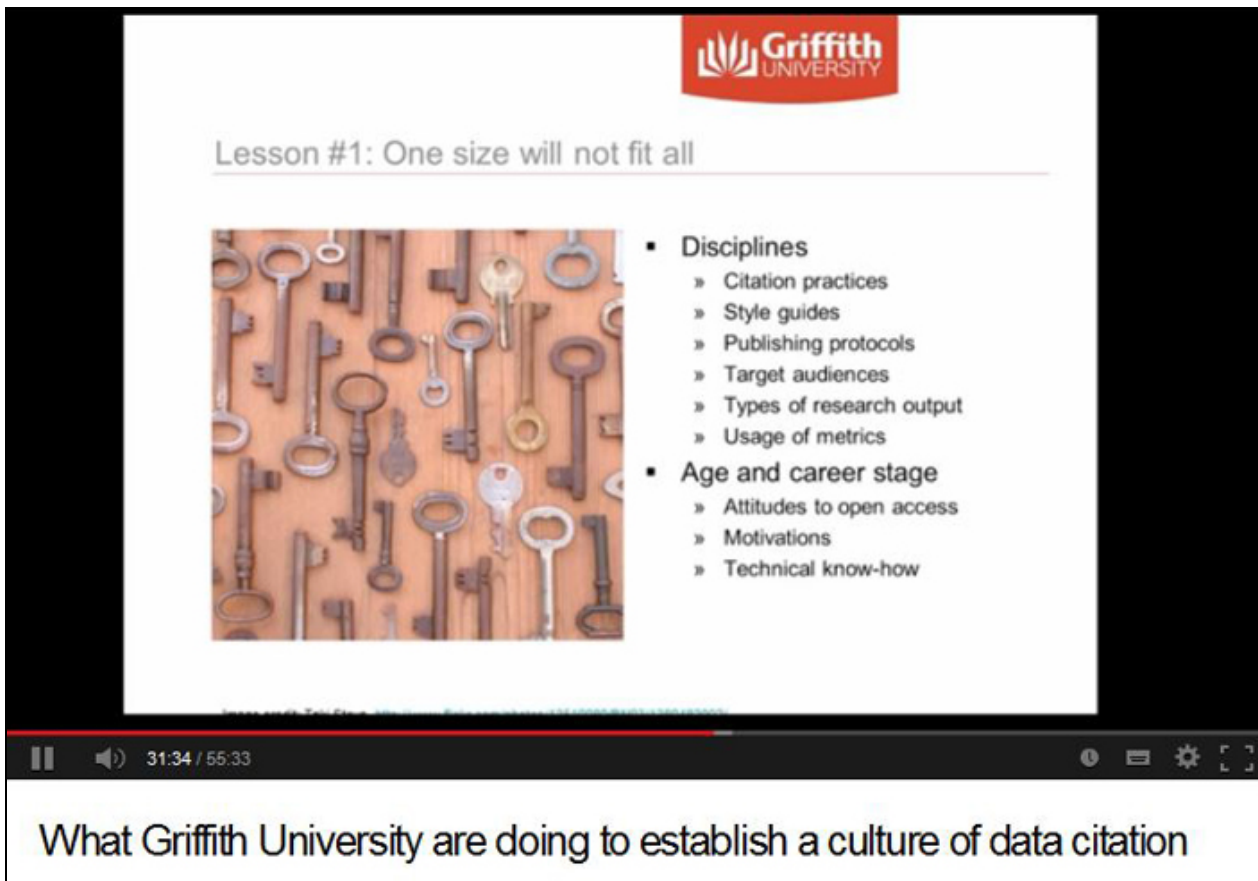
Assessing impact tools

As part of the Data Citation Project, Griffith arranged a one-month trial of the [Thomson Reuters Data Citation Index](#) during April 2013. The DCI forms part of the Web of Knowledge platform and was launched in November 2012. It provides subscription-only access to metrics associated with research data from global repositories covering multiple disciplines. The altmetrics tool ImpactStory was also trialed as part of the project and in connection with ORCID identifiers [10]. Results were shared with the Data Citation Community of Practice via a webinar in June 2013 [11] and via the project blog [12]. A key conclusion of these trials is that bibliometric and altmetrics tools may yield few, if any, results for datasets at this point in time. The reasons for this are multi-faceted and include: the lack of mandates for deposit of data; the early stage of data citation as a practice among researchers; delay between the availability of a dataset, its use and subsequent citation; and the need to expand the bibliometric product to include datasets from the ANDS Research Data Australia service. However, it is clear that data citation is a new scholarly practice and metrics for data citation will change considerably in the near future.

Engaging with researchers and librarians


As part of the ANDS-funded Data Citation project, Griffith provided targeted outreach to researchers about the benefits of data citation. The outreach strategy included conversing with subject librarians about citation practices in different disciplines, introducing data citation as part of a standard consultation with a specific research group, and engaging with researchers at the point of data deposit into the institution's data repository.

While a formal qualitative study was not part of the project, anecdotal evidence gathered from outreach activities suggests that a number of factors contribute to researchers' openness to discussing how sharing their data might contribute to the impact of their own research. This could include the types of publishing outlets in their discipline, their target audiences, and the processes by which their work is currently assessed. Age and career stage may also be a factor, with early career researchers seemingly more receptive. Subject librarians may also benefit from data citation awareness sessions, as this is still a new and evolving area.



Griffith UNIVERSITY

Lesson #1: One size will not fit all



- Disciplines
 - » Citation practices
 - » Style guides
 - » Publishing protocols
 - » Target audiences
 - » Types of research output
 - » Usage of metrics
- Age and career stage
 - » Attitudes to open access
 - » Motivations
 - » Technical know-how

31:34 / 55:33

What Griffith University are doing to establish a culture of data citation

Figure 2: Screenshot from YouTube Video of Griffith University Webinar on Data Citation

The project found that adequate guidance for constructing a citation for data is an area that requires further attention in formal guides and manuals. Citation guides do not cover the processes of citing data well, if at all. Mooney and Newton (2012) conducted a content analysis of journal articles, author instructions, style manuals and data publishers finding that "roughly half of journals point citation toward a style manual that addresses data citation, but the majority of journal articles failed to include an adequate citation to data used in secondary analysis studies" [13]. Additionally, in many universities training for citation practices is targeted at new undergraduates, a group that are fairly unlikely to be generating or re-using datasets until later in their academic careers. The shortcomings of journal policies and citation style guides (which may also feed into the templates commonly used in reference management software) combined with a lack of training opportunities are likely to inhibit the growth of data citation as a scholarly norm. Action in addressing these issues requires international attention in conjunction with small-scale institutional initiatives.

In terms of workflows for data deposit and citation, the project looked at the way that Dryad promotes data citation through their notifications to researchers following data deposit. As part of their community outreach, Dryad provides the text of the author notifications in their submission workflow. The Dryad model allows for data deposit and the linking of a publication to the underlying data to occur as part of the journal article publication. At Griffith, however, there is currently no mechanism for alerting the eResearch Services team as to when a researcher is about to publish an article. While the Dryad model is therefore not applicable at the institutional level, Griffith aims to adopt author notifications that are similar to those used by Dryad in the future, as part of the automated self-deposit process for research data that is currently under development.

The big picture

The formal evidence base for the benefits of data citation is still minimal. Griffith University is a cross-disciplinary institution and generalising the evidence base for data citation to researchers from disciplines not included in the studies risks easy dismissal of the conclusions. Benefits for researchers are not the same as benefits for institutions

or for funders, and this needs to be kept in mind when communicating about citation benefits with people who may feel increasingly pressured by the multiple efforts already in place to measure the value of their research. Griffith's experience suggests there is a need for more comprehensive and compelling evidence of the benefits of data citation for researchers from all disciplines.

Developing a culture of routine data citation is intricately linked to routine data deposit and data management practices. Griffith's new *Best practice guidelines for researchers: managing research data and primary materials* [14] incorporates data citation as part of a holistic view of data management, and over time, information and training materials that reflect this concept and practice are expected to improve. As a result of collaboration with ANDS, Griffith has the infrastructure to mint DOIs for datasets, and include a data citation element in records made available through the Griffith Research Hub and Research Data Australia.

Future plans

Given the success of their collaboration to date, Griffith University and ANDS have a range of initiatives planned for 2013 and beyond. These include:

- Increase the uptake of the Cite My Data service to all Australian research institutions with the capacity to mint and maintain DOIs.
- Strengthen and expand the Data Citation Community of Practice through the data citation webinar series and at various workshops, conferences and events.
- Continue to improve the Cite My Data service and respond to community feedback.
- Continue international contributions and ideas exchange with DataCite, particularly regarding the ODIN project.
- Work together to enable a single feed from Research Data Australia to the Thomson Reuters Data Citation Index and other developing products.

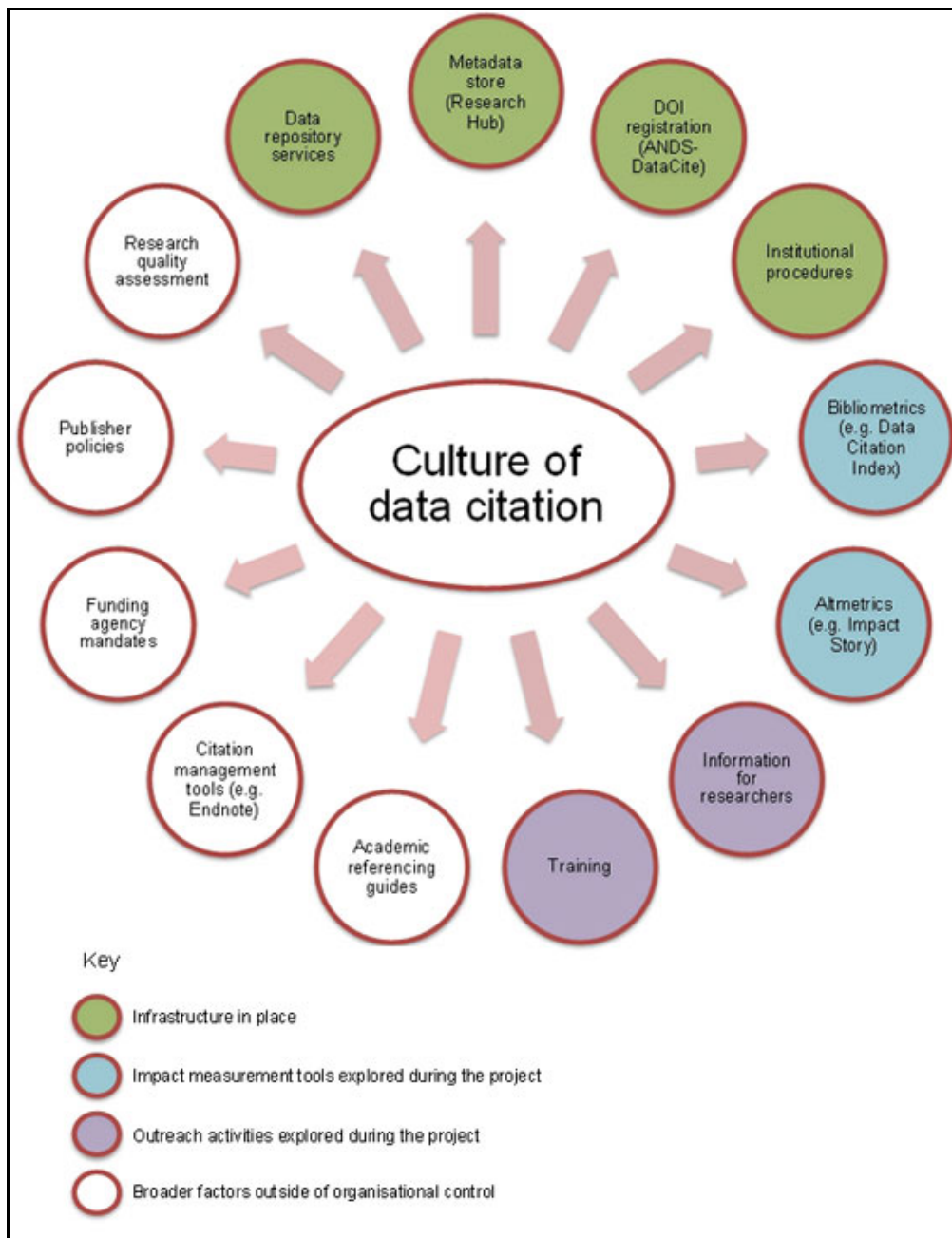


Figure 3: Overview of Data Citation Activities at Griffith University

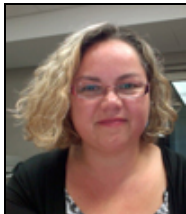
Conclusion

Data citation is an emergent practice, yet it has the potential to become a norm of scholarly attribution, in line with improved data management, access, discovery and reuse. At the institutional level, support for data citation includes developing infrastructure and providing guidance for researchers, as part of a broader strategy of improved data management and deposit. The partnership between ANDS and Griffith University provides a case study of what can be achieved through collaboration and mutual support. However, there are a number of factors beyond the capability of a single institution to address that will determine how well established a culture of data citation can become, such as funding agency mandates for data deposit and citation, the policies of scholarly publishers and the provision of adequate style guides. Collective action and strong leadership is required to address these issues over a period of time.

References

- [1] Mooney, H, Newton, MP. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication* 1(1):eP1035. <http://doi.org/10.7710/2162-3309.1035>
- [2] Research Councils UK. [RCUK Common Principles of Data Policy](#).
- [3] National Science Foundation. (2013). [Proposal and Award Policies and Procedures Guide](#).
- [4] [ANDS Data Citation](#).
- [5] [DataCite: Creating a global citation framework for data](#).
- [6] Simons, N. (2012). Implementing DOIs for Research Data. *D-Lib Magazine*. Volume 18 Issue 5/6. <http://doi.org/10.1045/may2012-simons>
- [7] [ANDS Data Citation for Researchers](#).
- [8] [ANDS Data Citation for Institutions](#).
- [9] DataCite. [Why cite data?](#)
- [10] [ORCID Connecting research and researchers](#).
- [11] [What Griffith University are doing to establish a culture of data citation](#). YouTube.
- [12] [Data Citation and Impact at Griffith University Blog](#).
- [13] Mooney, H, Newton, MP. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication* 1(1):eP1035, p. 1. <http://doi.org/10.7710/2162-3309.1035>
- [14] Griffith University. (2013). [Best practice guidelines for researchers: Managing research data and primary materials](#).
-

About the Authors



Natasha Simons is Senior Project Manager in the Division of Information Services at Griffith University, Brisbane, Australia. She has managed the Griffith Research Hub project, in addition to other projects funded by the Australian National Data Service. Previously, Natasha worked at the National Library of Australia in a variety of roles, including Manager of Australian Research Online. Natasha is a member of the Council of Australian University Librarians Research Advisory Committee.



Karen Visser is Program Leader for Skills and Policy at the Australian National Data service (ANDS), where she works to develop awareness of the benefits to researchers and institutions of making data citation a routine scholarly practice.



Sam Searle has been the eResearch Senior Specialist (Information Management) at Griffith University since October 2012. She contributes information management expertise to software development projects, and coordinates a range of activities related to research data management. She has previously worked at Monash University Library, the Office of the Information Commissioner (Qld), Victoria University of Wellington, and the National Library of New Zealand, and in other universities in Australia, New Zealand and Scotland in a range of research, archives and publishing roles.

Copyright © 2013 Natasha Simons, Karen Visser and Sam Searle

PRINTER-FRIENDLY FORMAT

[Return to Article](#)
