

Just accept it!
**Increasing researcher input into the business of
research outputs**

Lisa Ogle
Research Publications Librarian
University of Newcastle Library
Lisa.Ogle@newcastle.edu.au

Kai Jin Chen
Research Publications Librarian
University of Newcastle Library
Kai.Chen@newcastle.edu.au

Abstract:

This paper discusses the experience of the University of Newcastle in implementing the NURO research publications management system, using Symplectic Elements software. The new system provides an opportunity to value-add library services for researchers, while meeting the demands from university management for richer publications reporting tools. A particular challenge is re-engaging researchers in the research publications management process, after many years of the Library doing much of the work on their behalf.

Introduction

The Higher Education Research Data Collection (HERDC) is an annual publications data gathering and reporting exercise to the Australian government, used to assess the research and research training outputs of higher education providers (HEPs), and is one of the measures used to determine the allocation of research block grant funding worth more than \$1.6 billion to the sector each year. Typically administered by centralised research departments within universities, the responsibility for the publications collection at some universities has been assumed by library staff, who possess the expertise to support and add value to this process. At the University of Newcastle, the Library has managed the publications collection portion of the HERDC for the past five years. The publications collection process in the Library has been characterised by an emphasis on data quality, consistency in interpretation and application of the HERDC specifications in publications assessment and a drive to identify as many of the University's researchers' outputs to maximise the funding returns for the University. However, this activity centred solely on meeting the extensive requirements of the government reporting cycle and was not sufficiently geared to meeting researcher needs. It became apparent over the last two years that insufficient data about publications was being collected. This limited the University's ability to feed into and influence its position in university ranking schemes. Further, the systems being used to collect and manage these data were not equipped to support any change to existing processes.

Academic libraries are also seeking opportunities to deliver enhanced research support services to researchers around demonstrating research impact, using traditional bibliometric measures and exploring new measures such as article level metrics and providing support and advice to researchers about leveraging social media tools and networks as part of promoting their research. At the University of Newcastle, these services are currently provided by Faculty Liaison Librarians, who usually have to pull this data together manually from disparate systems to provide to researchers.

Evolution of a model for managing the HERDC: maximising the return for the University

Until 2008, responsibility for the HERDC publications collection process operated via a distributed model, with Research Services having responsibility for the overall collection, and administrative staff located in the Schools collecting and entering publications data into the Callista Research database. Callista is a manual publications entry system, which the University used to record publications data. The system had not been upgraded or supported for many years prior to the change to Symplectic Elements.

After a University-wide restructure in 2005, the position of Research Publications Officer was lost from the Research Office and it became evident that assistance was required to ensure the University could meet its annual reporting deadline. The University Library was approached to provide assistance, and in 2006, Library staff first became involved with verification of research publications, auditing a small percentage of the total collection prior to the Australian government reporting deadline. Searches of key scholarly databases identified an additional 700 HERDC

eligible publications. Around 20% of the publications audited were inaccurate: incorrectly assigned category, entry of duplicate publications, or incorrect bibliographic data.

In 2007, administration of the HERDC was a joint initiative between Research Services and the Library, with administrative staff in the Schools continuing to collect and enter data into the publications database and Library staff auditing these publications, sourcing additional verification information and value adding through conducting additional database searches to identify publications not reported.

In 2008, the University of Newcastle Library assumed responsibility for the entire publications collection portion of the HERDC for the university. A fulltime Research Publications Librarian position was created to collect, verify and enter the entire reported research outputs for the University. This included publications and other research outputs not HERDC-eligible, such as creative works and reports. A further estimated 500 hours of effort was also contributed by the Faculty Librarian, Health who initially established the Library's HERDC processes and continued to manage the Library's role in the research publications collection. Research Services continued to manage the overall HERDC submission.

A report on the 2008 HERDC (Booth) identified the following benefits from Library staff involvement:

“

- Library staff possessed expert understanding of scholarly publishing context and ability to interpret bibliographic data;
- Library staff had the ability to continuously monitor online databases for publications and associated verification evidence;
- Efficiency gains via centrally managing and storage of publications and evidence digitally, as well as the ability to respond quickly in the event of a publications audit;
- Increased quality control of data entered;
- Verification of evidence to occur just once at the point of entry into the database;
- Consistent interpretation of government specifications and guidelines;
- Continuous entry of publications into Callista ensured that publication lists were more up-to-date;
- School staff could be released from Callista data entry responsibilities.”

Library staff also drove initiatives to improve the ability to identify Newcastle's research publications, through targeting inconsistencies with author affiliations on publications and identifying additional database search strategies to capture the research output of the University's large cohort of medical conjoint staff, who do not always list the University on their publications. As a result, in 2008, the number of publications identified by the Library rose to 1200, capturing approximately 45% of the total HERDC-eligible publications. Most of the remaining publications were self reported to the Library by researchers or their support staff. Additionally, the institutional repository and the travel grants database were utilised as sources of

publications data. A Publications Review Panel was established in 2008 to review the HERDC eligible publications and assist in resolving disputed classification of publications. The Review Panel meets on an annual basis and is chaired by the Deputy Vice-Chancellor, Research, with membership consisting of the Assistant Deans, Research from each Faculty. The number of disputed publications is on average 30 per year.

What about services for researchers?

Researchers are concerned...that universities have created a range of different systems and points of entry which they find complex and overwhelming as they try to deal with the various aspects of their research work.(Research Information Network, 2010).

The gains achieved during these years, while of benefit to the University from a regulatory and funding perspective, were not necessarily valued by researchers. Researchers were now dependent on the resourcing of a single staff member to enter their research outputs, which had to be individually collected, verified and data entered before researchers' online publications profiles were updated. At any time throughout the publications collection process, there was always a backlog of about six months, which meant that Library staff were not able to begin entering the current year publications until HERDC reporting on the outputs of the previous year had concluded.

While in principle the University was now operating on a continuous publications collection cycle, the entire process was geared around the HERDC deadline of June 30 each year. This meant that researchers often experienced delays of more than 6 months to see their publications and outputs appear on their online profiles.

Over time the Library became more proficient at finding University of Newcastle affiliated publications; however, processes to verify, collect evidence and data enter publications did not change. Publications verified and entered increased 10% per year on average in the period 2008-2012. Resourcing remained static with 1.0 FTE staff on paper (excluding the 0.2 FTE support of the Faculty Librarian, Health), supplemented with casual support when available. There was a significant amount of double handling of publications and duplicated effort, as researchers and their support staff reported publications to the Library that had already been sourced, but not verified or entered into the publications system. The research publications team worked weekends, public holidays and during the University's Christmas closedown, in order to address the constant backlog and demands from researchers to update their profiles. This work was hidden from researchers, who were by now very removed from the collection process, as well as frustrated by the time it was taking to update their profiles.

This level of activity appears to match the effort expended across the sector, with the PhillipsKPA Report (2012) showing that,

“...the median estimated staff effort on the HERDC research publications collection in 2011 was 547 days, making it the most labour intensive of all the DIISRTE [Department of Industry, Innovation, Science, Research and Tertiary Education] reporting requirements. The HERDC

research income collection and the...staff hours and indirect costs surveys were also relatively labour intensive involving an estimated median staff effort of 95 days, 85 days and 15 days respectively."

Researchers also potentially had to report their outputs into either/all of the following:

- Newcastle-affiliated publications and required verification evidence to the Research Publications Librarian (publisher versions full-text)
- Self-entry of non-Newcastle affiliated publications into the Research Portfolio Manager (RPM) research profiles system. This system was developed in-house for Research Quality Framework (RQF), and subsequently Excellence in Research for Australia (ERA), reporting. The Callista system could only accommodate entry of University of Newcastle affiliated publications. These publications flow from Callista through to RPM, but if researchers wished to have their earlier unaffiliated publications counted for ERA or appearing on their profile, they must enter them into RPM themselves.
- Notification of publications (if desired) into the University's open access digital repository and submission of post print version of the full-text to the Library

All University of Newcastle researchers are asked to provide feedback on the publications collection process at the end of each collection year. From this feedback, the following priorities were identified:

- Up to date online research profiles and the ability to generate publication lists for grant and promotion applications
- Ease of access to bibliometric data to support grant and promotion applications including h-index, citation counts, top cited publications
- Less 'administrivia' and time spent chasing up internal reporting requirements = more time for research!
- Support/systems that help the researcher to demonstrate impact without creating additional work i.e. data visualisation, collection of social metrics, ability to push their data into social media.

In early 2012, University management realised that the existing model was not sustainable, and agreed to fund a new research publications management system. This decision coincided with the arrival in November 2011 of a new Vice-Chancellor, who was requesting increasingly complex data about research quality and impact, including research outputs, which could not be sourced from the institution's current systems. The University also released the 'NeW Directions Strategic Plan 2013-2015' around this time. One of the key strategies was to 'Increase the number of high quality research outputs by at least 30 per cent, and ensure strategic and timely submission of data supports the highest possible disciplinary rankings for the University in world university ranking schemes' (The University of Newcastle, 2012). Internally, the new management team also set a key performance indicator for researchers around increased collaboration with international authors on research publications. The current systems could not effectively track these data.

The 2012 *Review of Reporting Requirements for Universities: Final Report* by PhillipsKPA outlines a set of principles for the collection and dissemination of data by higher education institutions:

“

- a) **Fit for purpose** - data is used to suit the purposes for which it is designed to be used;
- b) **Privacy** - data collected will comply with laws relating to privacy;
- c) **Consistency** - data is consistently collected and applied across uses and time;
- d) **Auditability** - data can be readily scrutinised and verified back to source;
- e) **Transparency** - data has clear meaning;
- f) **Timeliness** - data is sufficiently timely to enable key stakeholders to analyse performance, risk and outcomes and take responsive action where appropriate;
- g) **Validity and reliability** - data should be robust and measure what it is intended to measure; and
- h) **Efficiency and cost-effectiveness**– data collection should be justified by the value it yields for the various purposes for which it is collected and should not be excessively burdensome on either data suppliers or data users.”

The report also suggests the following additional principles:

“

- **necessity** - data collection elements should be essential for a clearly identified purpose;
- **no duplication** - data collections should make use of existing data (including reasonable proxy data) if at all possible;
- **mutuality** - data collections and dissemination arrangements should recognise the mutual responsibilities of the parties involved;
- **accessibility** - data collections and dissemination arrangements should be developed in ways that are useful and accessible to the institutions providing the data, as well as government;
- **stability** - as far as possible data collections should be stable over time, while being subject to periodic review for currency, consistency with these principles. Wherever practical, reviews should be scheduled in advance and substantial notice given before changes are implemented.”

With reference to these data collection principles, Library and Research Services identified the following as the key areas for improvement, to be addressed by the new system:

- Better alignment of HERDC processes with ERA and NOVA Digital Repository processes, including a single data entry system for all publications.
- Capturing of ‘value-added’ metadata e.g. bibliometric and citation analysis, collaborations with other universities, in order to measure the University's research impact.

- Reduce the amount of manual data entry and duplicated effort around publications, which were now numbering around 3500 outputs per year, 2000 of which were HERDC category publications.
- Improve the user experience for research staff especially by reducing the time lag between reporting of publications to the Library and appearance on their research profiles, in an era of rapid online publication.
- Better capture of non-affiliated publications, which were difficult to capture using the existing database searching techniques. The University of Newcastle has a large cohort of medical conjoint staff who do not always list the University on their publications.

Identifying a suitable system

The University undertook a scan of the products available in the market, and identified Symplectic Elements as the system that best suited its needs. Symplectic is a UK-based company, started in 2005 by researchers from Imperial College London, and was just moving into the Australian market. The company initially developed a publications management product for the College's Department of Physics. The product is known as Elements, and it is now used in 32 institutions in the UK, NZ, USA and Australia. In early 2012, the system was used by one university in Australia, University of Melbourne, with another, University of New South Wales, in its implementation phase.

Symplectic Elements integrates with a suite of databases, including Web of Science, Scopus and PubMed, to automatically search for a researchers' publications on a regular basis. Elements relies on researchers configuring their individual search settings in the system to maximise the number of their publications found, while excluding publications by other authors with similar names. Once records are pulled into the system, researchers receive an automated email asking them to accept or 'claim' the publication as their own. They can also reject publications that are not theirs. Rejecting, rather than deleting, means these publications will not be presented to the researchers again during subsequent database searches. CrossRef data is also available in the system via a 'supplementary search' matching CrossRef records to manual or harvested records.

Publications not found via automated database searches can be manually added to the system. Elements provides a solution to the tedium of data entry, by offering users an 'Assisted entry' form for books, chapters and articles, using data from Google Books and CrossRef. Users can enter minimal information about a publication such as title, author, ISBN or DOI and the system will populate the manual entry form with most of the required data.

Integration with institutional repositories was important, as the Library wanted to automate the process of sending metadata and publication objects to the University's repository, NOVA. Symplectic offers a repository workflow within the system and users can upload full text directly into an institution's repository.

Symplectic had a good track record for developing modules within Elements for government research reporting exercises in the United Kingdom and New Zealand. At the time the University agreed to purchase the system, there was no HERDC

module, and the project commenced on the understanding that the vendor would work with Newcastle and other Australian universities interested in the product to develop a suitable module.

The NURO project

A project team was formed in August 2012 to implement the Symplectic Elements system at the University of Newcastle. The team comprised a Project Manager from IT Services (seconded from the Research Office's data management unit), a Business Analyst from IT Services, a Data analyst from Research Services, and a Subject Specialist from the Library, with a staffing level of approximately 3.0 FTE. Staff members from IT Services were also used throughout the project to perform specific technical roles.

The internal name for the system is NURO (Newcastle University Research Outputs). The team motto became 'Just Accept It!', a play on the Nike slogan, reflecting the action that researchers would need to take to get a publication onto their online profile and included for the HERDC, as well as the challenges faced in getting researchers to accept a new way of managing their publications.

Highlights and Challenges

The highlights...

Implementation of NURO has allowed the Library to review workflows around the HERD collection, and coincided with a review of the publication types collected by the University. The majority of harvested publications link via DOI to the full-text held in online library databases. This means it is possible to verify affiliation, year of publication and publication type through a relatively seamless link. A decision was made to no longer collect electronic copies of these easily accessible and verifiable publications. Inclusion of the ERA 2010 journal rankings for individual journal articles also means that peer review status can be verified immediately onscreen, for many publications.

Creation of the HERDC module means the often-onerous HERDC verification process has been decoupled from the process of updating researchers' online profiles. When a researcher claims or adds a publication, it flows through to their profile, as well as moving into the HERDC module, where it can be nominated by the researcher, or by Library staff, for inclusion in the publications collection. Library staff then verify the publication against the HERDC specification and/or the University's internal guidelines.

The University of Newcastle commenced collecting publications data in 1998, so for well-established researchers their online profiles are often incomplete. In the initial search, NURO harvests a significant backlog of publications for many senior researchers who have been publishing since the 1970s and 80s. Researchers can export an updated CV directly from the system in a number of formats (Word, PDF, CSV (Excel), RIS (Endnote) and BibTex), which is useful for grant, funding, and promotion applications.

NURO provides individual researchers and their managers with an array of bibliometric data (H-indices, WoSWeb of Science and Scopus citations and Altmetrics) for individual publications. Altmetrics is a metrics system measuring the impact of publications in media, social media and research networks.

Reporting from NURO is live and transparent. University management can access reports directly from the system on the number of publications for their school/faculty or the entire organisation, including metrics data. It is possible to compare publications data across the organisation and to drill down to the individual researcher level. Heads of Schools can also view progress on the HERDC and prompt their researchers to nominate their publications, or provide required evidence, where necessary.

For individual publications, there is an onscreen heat map, based on author address information on the publication, showing where all of the authors of an article are based. These data can be used to provide information at an author, school, faculty or institutional level on the number of international collaborations.

The challenges...

Analysis of the data sources indicated that about 50% of Newcastle's entire annual research output could be sourced via the databases, mainly journal articles and conference papers. This meant that around half of the outputs would still need to be manually sourced and entered. There is very limited coverage of books and chapters, as well as other outputs that are collected, such as reports and creative works. Humanities and social science researchers do not experience the same ease of use as their STEM colleagues, with many of their publications still requiring manual data entry.

To address these issues, Library staff decided to maintain the current practice of conducting University of Newcastle affiliation searches in those databases not searched by NURO, such as Informit, SAGE, Taylor & Francis and Springer. Affiliated publications found using this method are downloaded to Endnote libraries and then uploaded to NURO via a locally created 'University of Newcastle' institutional user. Researchers are then manually linked to the publication in NURO. They are asked to claim the publication in the same way as a harvested publication. This process is relatively seamless for the researcher.

Newcastle has traditionally maintained a 'just in case' philosophy and practice around the collection and retention of HERDC verification evidence, but the new system provided an opportunity to review these practices. A decision was made to no longer store PDF copies of accessible publications with Digital Object Identifiers or Uniform Resource Identifiers. The project team negotiated with the vendor for a file attachments function within the HERDC module, but unfortunately this was not available in time for Newcastle's launch. This functionality is planned for early 2014, but in the interim, the Library will continue to maintain existing methods of collecting verification evidence for the more obscure publications.

The HERDC module, while now out of beta mode, is still quite basic and a number of improvements are planned for 2014. As more Australian universities take up the product, the module will evolve and improve.

The fuzzy logic used to match existing data migrated publications is quite stringent, resulting in a number of mismatches between existing migrated data and the data pulled in from databases. From 2006, the Library expended significant effort cleaning up existing bibliographic data held in Callista, as well as improving data entry processes. This was necessitated by the Research Quality Framework and Excellence in Research for Australia exercises, alongside the implementation of the NOVA digital repository. Often mismatches occur because the harvested citation data from Web of Science or Scopus was of poorer quality than the University's internal manually entered data, which was not anticipated.

It took longer to develop the HERDC module than anticipated, and integration with Valet repository workflow software required greater consideration, collaboration and resourcing. It was over a year from the start of the project to researchers interacting with the system, with the initial April launch date pushed back to July and then to October.

The out-of-the-box repository workflows within Symplectic Elements did not suit the needs of the University of Newcastle. Elements is end-user focussed, with researchers able to upload texts from within NURO directly into the digital repository. However, this level of interaction bypasses some of the existing open access processes that aim to improve data quality and ensure all copyright requirements are met.

A decision was made to send only the citation metadata to the repository, with NURO connecting to the existing Valet repository staging system, where the current workflows could be executed. This occurs at the point a publication is 'Accepted' for the HERDC or 'Accepted (Internal)' for the broader publications collection. The Research Publications Librarian decides which record to accept, choosing the most complete and correct record. While an improvement on the current process of uploading all citation data into Valet in bulk at the conclusion of the HERDC year, it is disappointing that full integration could not be achieved. This is a priority project for 2014, with funding approved.

What do the researchers think?

A focused communication plan was developed, starting from the top down, with both the Vice-Chancellor and the Deputy Vice-Chancellor (Research) becoming enthusiastic champions of the system early in the process. In the 6-month period leading up to the launch, the project team conducted many demonstration and information sessions for researchers and management, as well as comprehensive testing with researchers, and the feedback was positive. As a result, after years of waiting weeks, and often months, for their publications to appear on their profile, researchers are excited that they have control over what appears on their profile, as well as when it appears. The features most favourably commented on are the ability to export a CV; the inclusion of article level citations, journal metrics and Altmetrics; and the inclusion of information about international collaborations for their publications.

As at late October 2013, NURO had been available to University researchers for four weeks. There was concern that after five years of the Library managing the entire publications reporting process on behalf of researchers, it would be difficult to get

researchers to engage with a system where they needed to do some of the work themselves. However, take-up of the system has been very good, with over 800 staff logging in and interacting with the system in the first fourteen days, and approximately 2,000 current-year publications claimed or added to the system during that period. To give some perspective to these figures, around 3,500 publications are collected per year, and in the 2012 collection year, there were 1,800 unique researchers who published, including research students, casual researchers and conjoint staff.

Feedback from researchers since the system has gone live includes: “NURO is fantastic”; “NURO is fabulous”; “This is a great system, if only all of our systems were this good”; “a breakthrough system”; and “It offers a great many benefits to researchers”. While the principal reason for adopting the system has been to improve collection of current and future publications, the system has prompted researchers to include their complete publishing catalogue on their online profile. The challenge for the Library will be to keep researchers engaged with the HERDC process, given that they can now update their own research profile, without the need to provide evidence first.

As part of its communications campaign to researchers, the Library has focused on three key messages:

1. Optimise your search settings
2. Accept (claim) or add your publications
3. Don't forget the HERDC!

Conclusion

The University of Newcastle has implemented a new research publications management system that significantly improves the timeframe to update online researcher profiles. Researchers now have access to real time bibliometric data within a single system and can generate up to date publications lists and CVs. At the same time, workflows around the HERDC and the broader research publications collection process have been improved.

Although there are still some issues to resolve, it is anticipated that once the file attachments functionality is delivered, the workflow will improve for both the researcher and Library staff. The University looks forward to engaging with other universities that are implementing Symplectic with the purpose of utilising the HERDC module, to identify and make further improvements. Early feedback from researchers indicates that they find the system useful and easy to use and they are excited they will have an up-to-date and enhanced online profile.

The biggest unknown at present is the degree to which the researchers will respond to requests for verification information from Library staff, to provide information to demonstrate that a publication meets the criteria for a HERDC category. However, given the increased publication targets in the University's latest strategic plan and the ability for senior staff to interrogate NURO for data on publications at any time, it is envisaged motivation to engage will come from outside the Library.

References

Booth, D. 2008. *The University of Newcastle HERDC 2007 Research Publications (2008 Collection Year). Internal Report.*

Research Information Network. 2010. Research support services in UK universities: A Research Information Network report. Last accessed 20 January 2014:
http://www.rin.ac.uk/system/files/attachments/Research_Support_Services_in_UK_Universities_report_for_screen.pdf

PhillipsKPA 2012. *Review of Reporting Requirements for Universities: Final Report*, Richmond, VIC, PhillipsKPA.

The University of Newcastle 2012. *NeW Directions Strategic Plan 2013-15*, Callaghan, NSW, University of Newcastle.