A ‘FACEBOOK’ FOR RESEARCH. A PERSPECTIVE ON THE CHANGING ROLE FOR INFORMATION ABOUT RESEARCH IN THE PROVISION OF RESEARCH SERVICES

Simon Porter, eScholarship Research Centre, Level 2 Thomas Cherry Building, The University of Melbourne Victoria 3010, Australia, Phone (03) 8344 3628, Fax: (03) 9347 4803, Email simon.porter@unimelb.edu.au

Abstract

Anyone who has either worked in research administration or has been on the receiving end of it would probably agree that that the dynamic that between researchers and research administrators is similar to that between market researchers and the public. Market researchers have an inexhaustible need to find out what the public is thinking, and are constantly surveying, running focus groups, and otherwise poking and prodding the public to tell them what they need to know. Likewise, research administrators are constantly requesting information about research from researchers in endless cycles. (Often the atmosphere in which information is collected from researchers is as if a research administrator had called them up during dinner to ask them for their publications!) The perspective that research administrators and marketers share is that they are both on the outside looking in.

For marketers though, with the rise of sites such as Facebook, the perspective is changing. The genius of Facebook is that the creators have managed to build an environment where the public wants to be and share information about themselves. Within this environment, the exchange of social information can be directly mined for marketing information without any specific questions being asked of anyone. Putting aside some of the my uneasiness about this setup, for a research administrator, the model of being able to get what you need without having to directly ask researchers is a tempting one to explore. So what would be the essence of a Facebook equivalent for research? How can we collect information on research as it happens?

INTRODUCTION

Anyone who has either worked in research administration or has been on the receiving end of it would probably agree that that the dynamic that between researchers and research administrators is similar to that between market researchers and the public. Market researchers have an inexhaustible need to find
out what the public is thinking, and are constantly surveying, running focus groups, and otherwise poking and prodding the public to tell them what they need to know. Likewise, research administrators are constantly requesting information about research from researchers in endless cycles. (Often the atmosphere in which information is collected from researchers is as if a research administrator had called them during dinner to ask them for their publications!) The perspective that research administrators and marketers share is that they are both on the outside looking in.

For marketers though, with the rise of sites such as Facebook, the perspective is changing. The genius of Facebook is that creators have managed to build an environment where the public wants to be and share information about themselves. Within this environment, the exchange of social information can be directly mined for marketing information without any specific questions being asked of anyone. Putting aside some of the my uneasiness about this setup, for a research administrator, the model of being able to get what you need without having to directly ask researchers is a tempting one to explore. So what would be the essence of a Facebook equivalent for research? How can we collect information on research as it happens?

The view from the outside looking in

Firstly, let's look at why we should worry about the problems of research administrators. In addition to internal reporting requirements, a university's ability to publicly communicate its research identity: who it is and what it does; the outcomes of its research; and what problems it wants to address are crucially important for engaging with the public, the media and industry. Information on our researchers, such as their projects, publications, and grants, are key information assets that a university can use to expose its research activity to the broadest possible audience.

Yet at most universities, information about research activity is fragmented. It is collected in a community of silos focussed on recording information for narrow sets of purposes. Examples include:

- Administering the funding and contractual obligations of research grants and contracts
- Ethics administration
- Recording publications output for government research assessment exercises
- Separately recording publications in open access repositories, and
- Replicating all of this information on staff, departmental, and faculty profile pages.

One strategy for reducing the fragmentation of research information is to target existing key research data gathering exercises, and deliberately expand their scope to cover the public communication of research identity. By defining policy around the set of information that can be communicated publicly about researchers directly from
administrative systems, public profile pages can be dynamically created to list such information as contact details, qualifications, grants, and publications. In addition to this core set of information, researchers can add information to their profiles such as a short biography, a photo, areas of research expertise, and international linkages.

![Diagram of public and private layers of information]

This approach has been successfully demonstrated by both the University of Melbourne's Find an Expert (www.findanexpert.unimelb.edu.au), and The University of Queensland's UQ reSEARCHers (www.uq.edu.au/uqresearchers). In the US, an impressive example of this approach is Cornell University's VIVO initiative (vivo.cornell.edu). Another example at the national level is Ireland's expertiseireland (expertiseireland.com.) The potential of these approaches is that they not only maximise the value of existing research information collection exercises, but that they also increase the quality of the data that is collected, (because the data is now public.) Additionally, the creation of centrally produced profile pages has the potential to focus a university's web presence, by replacing locally created web pages. At the University of Melbourne, several faculties, including the Faculty of Architecture and the Faculty of Engineering now point to 'Find an expert' pages rather than their own. Provided that all of the information sources represented have clear processes to keep information up to date, and that an effective search interface can be provided, these initiatives have the potential to become the 'front door' to the university for the public, the media, philanthropists, potential research students, and potential research collaborators.
Publicly communicating research identity creates additional reporting stresses

Increasing the scope of research reporting exercises by adding a public dimension adds additional stresses to both the amount of information collected, as well as the way it is collected. For instance, the legal name that a researcher is identified by in the HR system, may not be the name that they go by as a researcher, and additional processes need to be created to make sure profile pages make use of preferred names. A less tractable problem is the additional stress placed on publications reporting. At the University of Melbourne, publications histories going back to 2001 are now listed publicly. A senior researcher arriving at the university now requires 7 years of publication history to be entered into the system to be represented fairly against his or her peers, rather than just the set required for current reporting exercises. Each additional publication still needs to be carefully entered, identifying all internal authors, and adding research classifications. We can maximise the value of the information we collect, but the cost of collecting this information is still high.

Increased pressure on the quantity of information that is recorded, in particular publications, highlights the need to be able to accurately source information where it already exists elsewhere. In the case of publications, publication information clearly exists in publication databases such as Thompson Reuters ISI Web of Knowledge, PubMed, Elsevier, or even Google Scholar. What is lacking is the ability to consistently, and accurately download lists of publications based on a researcher’s public identifier in a similar way to the way authors are uniquely identified against books in library catalogues via a name authority file. One vendor’s approach to this problem is Thompson’s Researcher ID product that allows researchers to create a public identifier for themselves, and then enter in their publications history. A national response to this problem might be to leverage the public identifier infrastructure that exists within the National Library of Australia’s People Australia project (http://www.nla.gov.au/initiatives/peopleaustralia/). If every university in Australia could provide public profiles that could be harvested by People Australia, as researchers move from one Australian university to another, their public research history could be downloaded from the national record.

Limits to a university’s knowledge about its research

One of the significant limitations of relying on aggregating already existing sets of information about research to communicate research identity is that there still remains a large gap in our knowledge about research that is conducted at our universities. This gap is a single comprehensive source of information about the ongoing research projects and programs at a university. Typically information about research groups and projects is distributed across faculty and departmental websites, where it is treated as a content management exercise, and like all content management exercises, there is a constant (losing) battle to keep information up to date.

Whilst websites might be the only public source of information about research projects, within universities, research project information is repeatedly declared on ethics and grant applications, annual reports, faculty and university reporting exercises. All of these processes record project information for a particular purpose, (typically only recording only the Principal Investigators,) and all take a slightly
different slice of information that cannot be easily reused. Should a project span multiple universities, requests for information quickly multiply.

One reason that universities engage in multiple partial collection exercises for research project information is that by themselves, they lack the information infrastructure to be able to accurately define research projects, and programs. As a significant proportion of research projects cross organizational boundaries, the information infrastructure to adequately describe all of the people involved in a given research project is lacking. For internal participants, universities can use unique identifiers based on HR records. For external participants, universities can only use free text fields. Even to call these participants external is misleading, research projects just don’t neatly fit within research institutions. What we end up being able to record within our university walls is just an incomplete projection of the research activity that is going on. To put it another way, the ‘source of truth’ about research projects is distributed. Without proper records of research projects there is very little an individual university can do to offer services targeted at research projects, and subsequently little incentive to coordinate multiple processes around a single research project definition.

Figure 2. The truth about research is distributed

An eResearch response to the problem of Research Projects and Programs
In a parallel world to research reporting, the problem of recording research projects and programs around which services can be offered is the bread and butter of eResearch concepts such as virtual organizations. Over the past decade, universities have begun increasingly to share information infrastructure to enable researchers to collaborate in virtual teams across organizational boundaries (National Science Foundation 2007). Within Australia, these efforts are now coordinated by the Australian Government’s Platforms for Collaboration, which aims to:

- support technological platforms that enhance researchers’ ability to generate, collect, share, analyse, store and retrieve information. These platforms will allow researchers to work seamlessly from desk to desk within and between organisations. Seamless access to collaborative research opportunities, knowledge and information will enable researchers to perform their research more creatively, efficiently and collaboratively across long distances, regardless of location and time, and disseminate their research outcomes with greater effect. (NCRIS 2008)

Platforms for Collaboration is organized into five separate programs. Three operational programs, National Computational Infrastructure (NCI), Interoperation and Collaboration Infrastructure (ICI), and the Australian National Data Service (ANDS) focus on the provision of peak computing facilities, research collaboration tools, and access to research data respectively. Two other programs, the Australian eResearch Infrastructure Council (AeRIC), and National eResearch Architecture Taskforce (NeAT) focus on funding, and strategic direction of the program. (NCRIS 2008)

In addition, Platforms for Collaboration has provided a foundation investment in the Australian Access Federation Project which aims to “develop and deploy an infrastructure to facilitate trusted electronic communications and collaboration within and between higher education and research institutions both locally and internationally as well as with other organizations” (AAF 2008.) The vision of the Australian Access Federation (increasingly known as the authorisation fabric,) is that researchers will be able to authenticate against their own institution, and then to get access to resources and collaborations that are hosted elsewhere.

Based on this authorisation fabric, providers of eResearch Infrastructure can develop eResearch Services that allow access to resources to be negotiated at an individual level, but also at a project level. A significant development in this area is the establishment of a national ‘data fabric’ provided by the Australian Research Collaboration Service (ARCS) from within the ICI program, that allows “researchers, research groups, research organisations and research communities to store their data” (ARCS, 2008.) Other Initiatives built on top of the authorization fabric, focus on the management of virtual organizations and the coordinated provisioning of services such as wikis, and repositories. One example of this approach is IAMSuite developed by the Macquarie E-Learning Centre of Excellence. (MAMS 2008)

To date, the focus on eResearch initiatives such as virtual organizations within Platforms for Collaboration has been on security and access, and the collaborative functionality that researchers require to produce first class research. Often referred to as 'middleware,' the idea is that virtual organizations provide a platform upon which seamless access to resources, collaborations and tools can be provided. It is not generally intended that the definition of the virtual organization itself, especially if
representing a research project or program, could have a wider audience than just the group of researchers that it defines.

The purpose of this paper is to argue that, like the surfacing of statements of researcher identity based on administrative data via appropriately defined policy, research projects and programs expressed as virtual organizations could become a key information asset in communicating information about research to the public. Further, this development is not just required to improve communication channels between researchers and universities about their activities, but is an essential requirement of at least one part of the Platforms for Collaboration vision. Ensuring that researchers have the best possible access to the available research data; a goal progressed by the Australian National Data Service.

The Information Architecture of ANDS

The Australian National Data Service was established in early 2008. As part of the vision for the Australian National Data Service presented in Oct 2007, the following goals were articulated:

- A national data management environment exists in which Australia’s research data reside in a cohesive network of research repositories within an Australian ‘data commons’.
- Australian researchers and research data managers are ‘best of breed’ in creating, managing, and sharing research data under well formed and maintained data management policies.
- Significantly more Australian research data is routinely deposited into stable, accessible and sustainable data management and preservation environments.
- Significantly more people have relevant expertise in data management across research communities and research managing institutions.
- Researchers can find and access any relevant data in the Australian ‘data commons’.
- Australian researchers are able to discover exchange, reuse and combine data from other researchers and other domains within their own research in new ways.
- Australia is able to share data easily and seamlessly to support international and nationally distributed multidisciplinary research teams.

(ANDS Technical Working Group, 2007, p. 6)

One of the programs outlined to support these goals is the ANDS collections / service registry which is designed to enable discovery of research data hosted in research collections across Australia. Based upon the work of the Online Research Collections Australia (ORCA) registry project (APSR, 2007), the vision for the ANDS collections registry is a machine readable, as well as human searchable registry of descriptions.
of Australian research data collections as well as descriptions of the data access methods (services) available for each collection (ANDS Technical Working Group, 2007, p. 35.)

Through the existing registry schema for ORCA, it is possible to get a glimpse of the information architecture required to support ANDS. The current ORCA Registry project is based on ISO2146 - Registry Services for Libraries and Related Organizations (Burton, 2008), and leaves the way open for the possibility that as part of the common practices of associating metadata with research outputs, it should be possible to:

- uniquely identify parties (including people and organizations) across research collections, and
- uniquely identify research activities (including research projects and programs)

**ANDS Collections Registry (ANDS Technical Working Group, 2007, p. 35)**

The ability to discover research data based on unique identifiers for projects and researchers not only significantly enhances the possibilities for exposing research data (such as being able to systematically include this information in research profiles), but it also provides an infrastructure to reward practices such as allowing the open reuse of research data by making it much easier to aggregate usage statistics back up to projects and people.

The first challenge for ANDS is where these public identifiers for researchers and research projects will come from. In the future, it is reasonable to assume based on established 'platforms for collaboration,' that there are fairly well defined workflows that allow data created within a virtual organization to be deposited in data centres.
and repositories as a matter of course. It is also reasonable to assume that the research data that is deposited will be tagged with metadata that uniquely identifies the projects and the people associated with it. If this data is to be harvested by an ANDS collection registry, then these identifiers will need to be public.

As indicated in the previous section, the first challenge for this vision is that the current identifiers used within the AAF to enable security and access to research collaborations have not been designed to be publicly communicated. To identify the policy around what the public identifier for researchers and projects should be, what additional information should be communicated, and how this might be incorporated into the workflow of research, will require national discussion. Fortunately, for the reasons mentioned above, universities should be receptive to having this discussion as they have a vested interest in being able to discover and communicate information about their researchers, and in particular research projects and programs. In fact the policy discussion that ANDS would appear to require could be seen as an extension of the policy decisions around the communication of research identity at the university level. It may even be beneficial for the public identifiers that identify researchers against research projects to be the same as the public identifiers that uniquely identify authors against publications. As changes will also need to occur in the collaborative services that provided to researchers, this discussion will also need to be progressed within 'Platforms for Collaboration.' ANDS would seem uniquely placed to progress this conversation in both of these areas.

Figure 4. An expanded policy discussion
The second challenge for the ‘public identifier problem’ is that there are many different eResearch collaboration environments, and no excepted standards for the public communication of researchers and research project information. Whilst candidate information exchange formats have been developed such as the Common European Research Information Format (Jörg, 2008,) this work has come largely from the research reporting world, and has not yet been embraced by eResearch communities.

The third challenge is that only a comparatively small number research projects currently use eResearch collaboration environments, therefore the current benefit of harvesting research projects definitions from existing virtual organizations would be small. However, by addressing the second challenge and providing a standard way to publicly communicate information about researchers and research projects, providers of eResearch collaborative infrastructure have the opportunity to make the use of eResearch infrastructure mainstream. Like the relationship between social networking and marketers on Facebook, acknowledging the development of a symbiotic relationship between research administration and eResearch could provide the tipping point for a mainstream adoption of eResearch practices.

A model for the mainstream roll-out of eResearch Infrastructure in a University

To see why the metaphor of a 'Facebook for Research,' could apply to the mainstream rollout of eResearch Infrastructure across a university; imagine a university strategy that created a research project definition for every current research program and project in the university (perhaps using IAMSuite). Although only a small minority of projects would be considered 'eResearch' projects, the public component of these definitions – a project description, and list of people involved could be used as the 'source of truth' for program and project information in research portals, and as a basis for government reporting. In a scenario where many universities adopt the same approach, a single project definition could be shared between all organizations involved.

Having a single consolidated source of project definitions would then allow services to be provisioned to research projects in a similar processes to the way teaching and learning spaces are provisioned in Learning Management Systems from subjects and student enrolments defined in Student Administration Systems. Some of these services could be provided locally, such as wikis and collaboration suites such as Sakai. Alternatively these projects could make use of the strength of virtual organizations that defined on a common authorization fabric, and access services that are available in the 'cloud.'

For providers of external research services, the advantage of having an existing project definition to plug into, significantly lowers the barrier for adoption, as well as providing an identifiable target market. An additional benefit for eResearch service providers is that because they are often government funded, and need to report on who is using their services, using a list of actively maintained research project definitions is a significant positive. Should universities restructure some of their reporting requirements so that publications are associated with the projects that created them, external services providers could also benefit from this information.
An existing registry of research projects should also provide institutions with a systematic way to conduct capacity planning for eResearch services. For one thing, universities would well placed to ask that all projects submit a data management plans. Part of this plan could be an estimate of how much data was likely to be created, and where the information is currently stored. This information could be used to budget for the right amount of required data storage.

Finally, as research projects and programs defined within virtual organizations become the accepted way of communicating information about research activity, existing administrative process could be optimised by referring to a single project definition across ethics, grants and contract administration, and publications reporting. An additional benefit of basing administrative processes around virtual organizations is that the same research project definition could be used to support administrative processes that cross organizational boundaries, such as applying for a grant, or submitting a publication to a journal.

This at least, is one perspective on what the essence of a ‘Facebook’ equivalent for research should look like. Or to put it another way, by extending Platforms for Collaboration to include research administrators, and the information needs of universities, Australia can significantly increase its capability to both conduct first class research and effectively communicate that fact to the world.

References


The ANDS Technical Working Group (2007) *Towards the Australian Data Commons, A proposal for an Australian National Data Service* retrieved Oct 2008 from 

MAMS (2008) *IAMSuite* retrieved Oct 2008 from 
http://www.mams.org.au/confluence/display/IAM/IAMSuite
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
PORTER, SIMON

Title:
A 'Facebook' for research. A perspective on the changing role for information about research in the provision of research services

Date:
2008

Citation:

Publication Status:
Unpublished

Persistent Link:
http://hdl.handle.net/11343/35010

File Description:
A 'Facebook' for research. A perspective on the changing role for information about research in the provision of research services