

2013

Research data management in a collaborative network

Darren Gibson

Edith Cowan University, d.gibson@ecu.edu.au

Julia Gross

Edith Cowan University, j.gross@ecu.edu.au

This article was originally published as: Gibson, D., & Gross, J. (2013) Research data management in a collaborative network. *Research Global*, issue 33, pp. 12-15. publishers website available [here](#).

This Journal Article is posted at Research Online.

<https://ro.ecu.edu.au/ecuworks2011/586>

Research data management in a collaborative network

Darren Gibson and Julia Gross explore the research data challenges that may arise within a collaborative research network, and offer possible solutions.

Edith Cowan University (ECU) is one of 12 Australian universities striving to create world-class research capacity and capability through investment in the Collaborative Research Network (CRN) project. The CRN project at ECU aims to accelerate the growth of research activity in four key areas: health, education, ICT, and the environment. These areas align both with the Australian Government's national research priorities and ECU's strategic research plan.

Managing research data is essential, and this is particularly the case within a CRN if maximum benefit is to be achieved from the collaborations. The CRN project will create a significant amount of new data and it is imperative that this data is well managed to ensure that it is secure, discoverable, accessible, useable and re-usable. ECU, however, does not currently have a set model to follow for effective data curation.

In light of this, a qualitative study was undertaken to outline the research data challenges within a CRN and discuss possible solutions. The study will inform policy development and enable robust guidelines, procedures and technical solutions to be developed and deployed resulting in positive benefits for the research community. It may also act to increase research outcomes and productivity.

Method

The study aimed to explore academics' experiences and challenges within a research data management context. An initial discussion was undertaken with a variety of stakeholders – including internal groups (the Information Technology Service Centre, the university's library, and the Office of Research and Innovation) and external groups (the Australian National Data Service (ANDS) and iVEC –

Data must be well managed to ensure that it is secure, discoverable, accessible, useable and re-usable.

Western Australia's supercomputing facility) – to discuss the best course of action to follow. The resulting plan is illustrated in Figure 1.

It was decided that a pilot study would be undertaken to investigate the current research data management procedures, opportunities, and challenges within the ECU's CRN. A suitable research instrument was developed and a series of semi-structured research data management interviews were conducted. Specifically, the interviews were designed to explore the following:

- Gaps in the existing model of data management

- The challenges of data management from a researcher's perspective
- How to develop a model for research data management to ensure data is collected and consolidated consistently
- How to develop an effective model for data sharing across multiple partners (outside ECU)
- The possibilities for partnership – for example, working with iVEC for data sharing and storage

Current practice

Researchers within the CRN at ECU collect and store an assortment of data in a multitude of different formats, depending on the nature of the research and chosen study method (qualitative or quantitative, for example). The size and scale of the total data collected and held varies considerably across research disciplines and even within disciplines and projects. Some of the datasets may also contain confidential information that may have ethical and/or commercial sensitivities.

A range of data formats were discussed during the interviews, including video, audio, images, and statistical data from questionnaires. There are currently a range of software packages being used for manipulating data, as

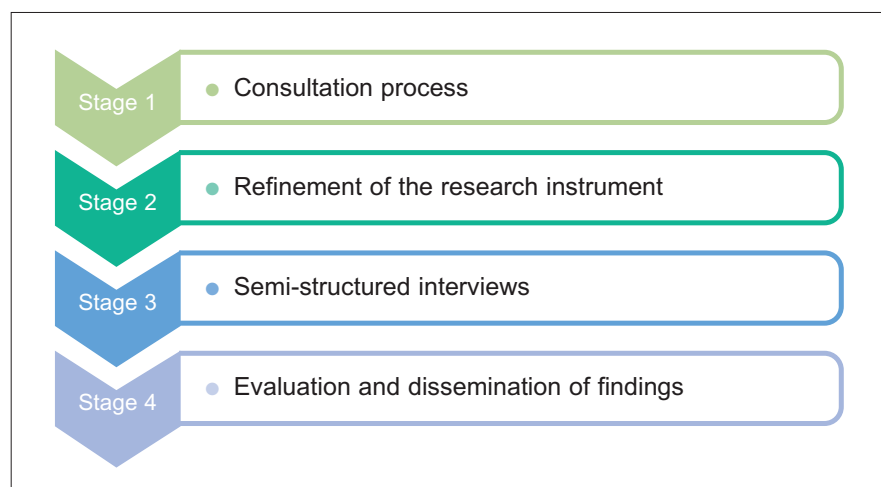


Figure 1: The multi-stage process of the review of research data management at ECU

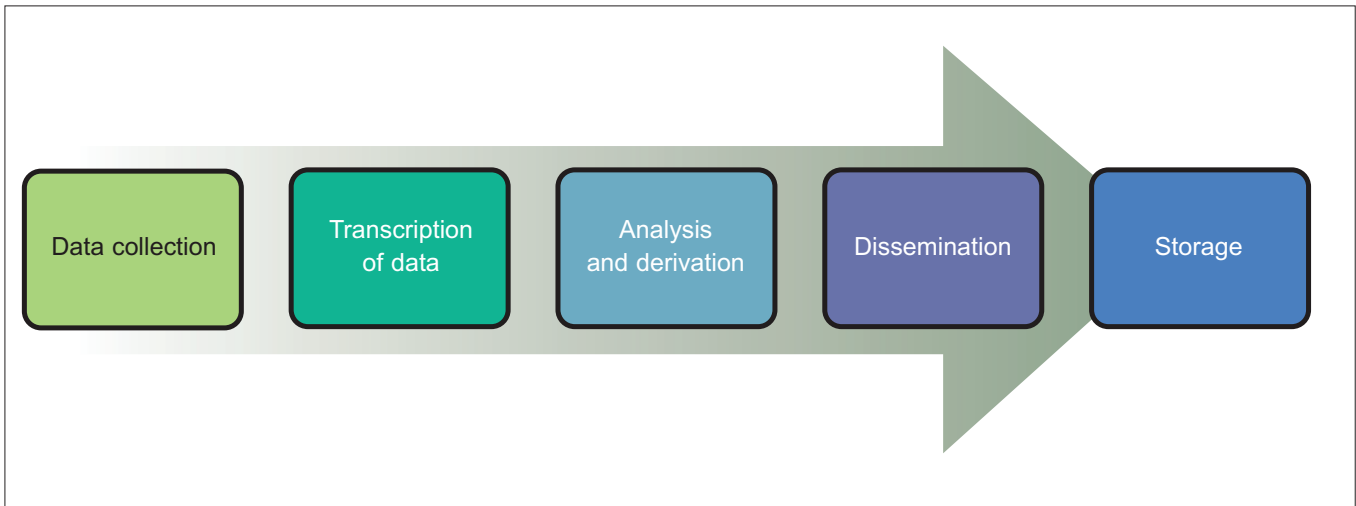


Figure 2: Flow of data within a research study

well as for editing, analysis and transcription. These include Microsoft Office, statistics packages such as SPSS, digital video editing software, qualitative software packages, websites/portals for sharing, questionnaire sites, and databases.

While the majority of digital data will be retained indefinitely, paper copies are normally destroyed after the designated retention period detailed within the current ECU records policy. One issue highlighted in the interviews was the lack of a standard method for managing research data: some research teams are proactive in data management while others lack motivation and/or knowledge.

The majority of research teams interviewed use the internal ECU servers to store their data. Within the data-intensive groups, however,

alternatives have also been sought – external hard drives, laptops, and commercial data storage options, for example. Many groups also store back-up copies of their data, uncompressed video files, or images on external hard drives – mainly because of the space limitations within the current ECU infrastructure. Flash (USB) portable drives are also routinely used but only for transient purposes (such as transferring data from an external piece of equipment onto the server). However, the long-term storage of data using a flash drive is not common. Accessibility issues make sharing data stored on the university’s IT infrastructure problematic for collaboration. The general flow of data within a research study is outlined in Figure 2.

Challenges

Not unexpectedly, the research data challenges highlighted within the interviews varied, depending on the nature of the project. Three main themes emerged (shown in Figure 3) along with some minor themes; these challenges will now be discussed in more detail.

Data sharing: The main issue for the research community is sharing research data during collaborative projects with inter-state and international partners. Collaboration is a critical factor in growing research at ECU, and the issue around sharing data must be resolved to allow the university’s research profile to develop.

There are different layers of complexity in sharing research data with collaborators. Among the relevant factors are the following:

- Data may contain confidential or sensitive information
- The size of data to be shared
- The number and location of partners in the collaboration – the more partners collaborating in a project, the greater the risk of the data set being contaminated
- Sharing data using ‘open access’ sites

Issues experienced by researchers include:

- File size restrictions and quantities of data
- Sharing and accessing internal file sites with partners or externally
- Collecting data off-site and transferring it back to ECU
- The range of supported software packages

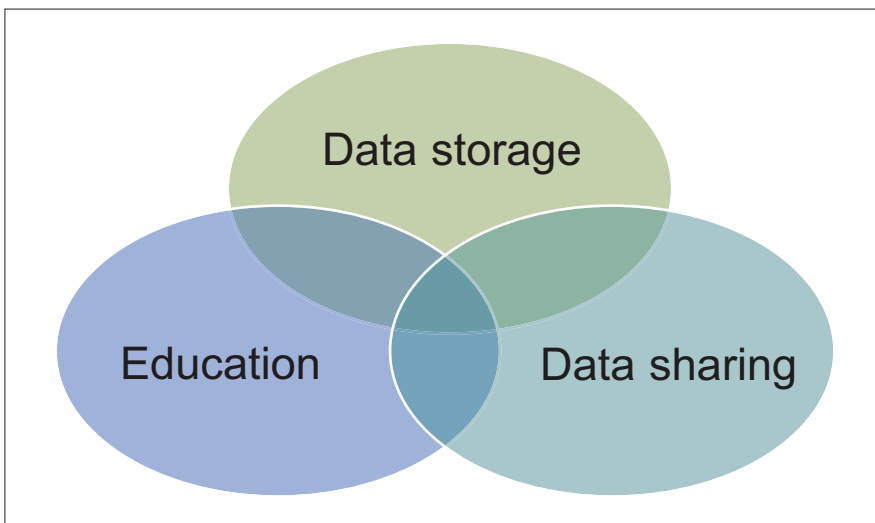


Figure 3: The main challenges facing ECU researchers

Data storage: Secure storage and back-up processes for research data are essential to guarantee that high quality research outputs are successfully archived. Whilst some researchers are using the central data storage infrastructure provided by ECU, many are not. Those who do not cite a range of reasons including: storage size limits and restrictions, the lack of support available, and problems with significant latency in access.

These problems are rapidly expanding due to the increase in research-active staff and numbers of higher degree students. The ECU records department could assist with the storage of archival data; however, there must be agreement around retention periods, the destruction of research data, and the availability of data for re-use. One researcher described how interviews collected during her PhD (over ten years ago) have historical value and will now be made available to the State Library of Western Australia, providing an excellent snapshot of life at that time.

The majority of research teams will almost always retain their data indefinitely. The cost implication of this to ECU is unknown, but is likely to be significant. To develop and support research, ECU may be required to provide increased data storage space for the research teams or to outsource the data storage.

The security and storage of datasets is also an issue because of the nature of research and the potential for data errors resulting from multiple researchers accessing the data. Therefore, restricting access to specific project folders or creating a second working copy of the data is possible. However, a second copy of the data could prove troublesome if multiple collaborators are analysing the data within the same time period. This could result in different copies of the data being used and datasets being modified. Version control and a data log should be created to prevent multiple copies of the data being circulated.

Education: A significant issue within research data management is education. The knowledge base for researchers in research data management is mainly obtained via training and experience gained throughout their careers. Therefore, a PhD supervisor's knowledge –

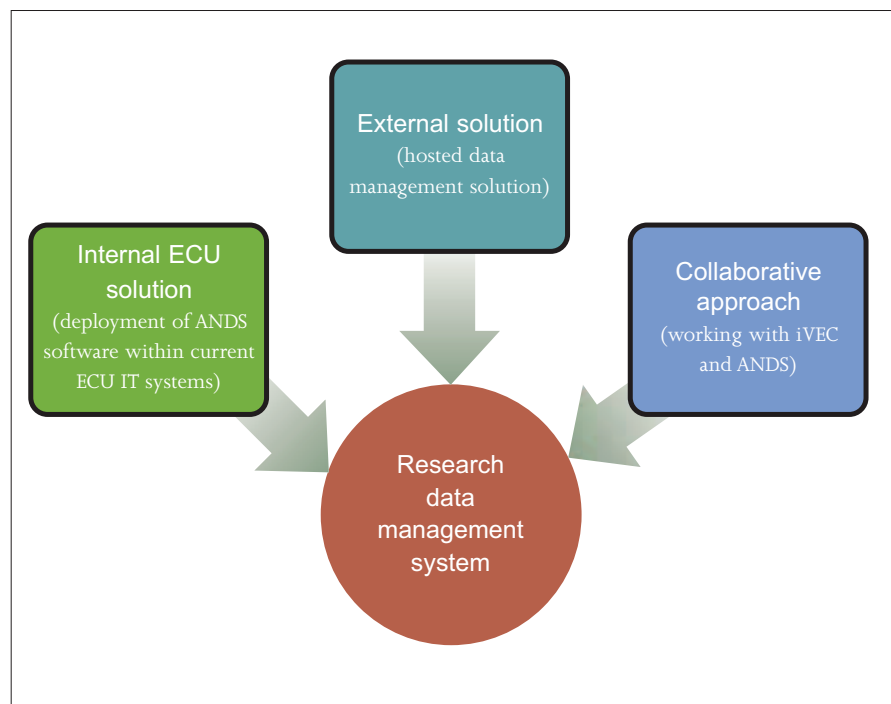


Figure 4: Possible technical solutions to research data management

good or bad – is generally passed down to his or her students. The study highlighted the lack of training offered to the research community within research data management. Even if training was offered, however, there is no guarantee of significant uptake by the community. Within a collaboration, two parties may have differing views on best practice for research data management; in this case, such issues must be resolved at the start of a research study.

Generally, research teams were found to be either unaware of the current institutional support available or lacking the time to learn about the advantages of improved custodianship of their research data. One specific example, which was highlighted within multiple groups, was the lack of a standard data management plan. The majority of groups had developed something themselves or intuitively organised their data in a useable format. However, a draft data management plan should be disseminated to research teams to ensure standards throughout the university and across the CRN partners.

The collection, organisation, and effective use of metadata is a challenge for many, and research teams indicated that they would benefit from assistance in generating effective/meaningful metadata. During an audit process or a request for data, ensuring that the research team has a simple, logical metadata system

would inevitably save time and resources. One team described how their partners have a system that automatically produces metadata from deposited data files. Having such a system at ECU could save time and resources and could lead to increased research productivity.

Other aspects for consideration while working in a CRN are the governance arrangements covering research data (such as how and where data is stored and the duration of retention, for example). This information is usually detailed within research ethics applications. However, some researchers only submit an ethics declaration stating that their research does not involve animals or humans (as opposed to addressing issues surrounding data, etc.) This leads to inconsistency in the governance arrangements around the subsequent management of data generated.

Technology has enhanced opportunities for collaboration but, in order to respond to evolving file formats, a central storage of data is imperative. For example, if a researcher stores their own data, technological advancements may result in a data format becoming obsolete. This will present difficulties for the researcher when they need to extract the data; and may be especially problematic in longitudinal studies.

Possible solutions

How can ECU best facilitate collaboration and data sharing with our partners? As the CRN matures and new collaborations are established, more data will be generated and these issues will become increasingly prolific and problematic to overcome.

To support the ECU policies and guidelines being developed and implemented, a software solution may be required, such as a content or research data management system. For example, a research data management system that assists the research community with data storage and sharing could be developed. A number of options are currently available and these are illustrated in Figure 4.

All these models have benefits and risks. An effective research data management system is pivotal for the sustainability of the CRN within ECU and beyond. It is hoped that the research data management themes highlighted in the study can be developed and integrated into other research streams within ECU's future research activities.

Storage and access: Researchers always need more data storage space and this should be easily accessible to allow further analysis and collaboration. Currently, accessibility of files off-campus is an issue due to the substantial latency inhibiting the easy viewing of large data files. As detailed earlier, some researchers store research data in other locations to ensure improved access to the data files and to allow collaborators easy access to information. One solution would be to use the current national infrastructure within Australia, accessible via high-speed networks (iVEC, for example). There is, however, a caveat when using external providers in that there may be security challenges around data storage (especially for clinical or social data). Consequently, the ideal solution may be the creation of a large ECU research data store with a simple, secure interface, allowing easy access to data files off-site and providing a route for external collaborators to login.

For researchers using video-based research data, accessible storage could be outsourced to a commercial organisation. It is understood that commercial organisations provide web-

based storage using a 'shopping cart' interface that can convert video data into any required digital format; data access can be restricted and simultaneously accessed anywhere in the world. Such a facility would allow collaborators to access their data, but with access control remaining with the chief investigator. A facility that could stream data directly from video cameras into a large data store, provided locally by a university or commercially (e.g. in a cloud), could also be useful. However, there would have to be a cost-benefit analysis of outsourcing carried out before proceeding.

Data sharing: In general, the research community has difficulties sharing information, especially research data. Currently, practices vary and include travelling to specific conferences to meet collaborators and exchanging data, using commercial sharing sites (such as Dropbox, Wikis, and Google Docs) and posting flash drives or CDs in the mail. Many ECU research teams would be interested in sharing their metadata after publication, but not necessarily their research data. The ECU library manages the repository Research Online (RO) which is a great asset and research output store. RO could be pivotal in providing an accessible central point for storing a range of data, including images, datasets, conceptual diagrams, and so on.

Another possible access solution to facilitate collaboration could be to develop a common research portal that would allow all collaborators secure access (via Australian Access Federation login credentials, for example). Underlying data could be securely stored within the institutional data store and the researchers, and their external partners, could gain access.

Multiple e-research tools and virtual laboratories are currently being created within the National eResearch Collaboration Tools and Resources (NeCTAR) funding scheme, and will become available to the research community over the next 12-18 months – mostly via the NeCTAR Research Cloud. This new infrastructure will increase collaboration across all research sectors by providing various mechanisms to access, manipulate, process, and analyse data.

Conclusion

To overcome all the issues around storage and sharing will require significant investment, education and, ultimately, behavioural change. However, there are many issues which can be more easily resolved – including the introduction of guidelines and procedures – and doing so will help the research community to advance their research and data management governance arrangements. These issues will only be resolved through effective and ongoing dialogue between researchers and service providers – both internal and external – to ensure needs are understood, documented, and acted upon in a timely fashion. Ultimately, ECU's capacity to improve its research output, particularly surrounding international recognition, depends on its effective resolution of data management issues. RG

Darren Gibson is Collaborative Research Networks Program Manager at Edith Cowan University, Australia.

Julia Gross is Senior Librarian (Research Services) at Edith Cowan University, Australia.