ICT Requirements for Research: Survey, Analysis and Outcomes

Andrew Buttsworth¹, Florian Goessmann²
¹Curtin University, Perth, Australia, a.buttsworth@curtin.edu.au
²Curtin University, Perth, Australia, florian.goessmann@curtin.edu.au

INTRODUCTION
A comprehensive understanding of researcher requirements is vital for any institution that conducts and supports research. As part of our ongoing efforts to improve our understanding of how researchers use Information and Communication Technologies in the research workflow, Curtin University’s eResearch Support team recently surveyed Curtin’s research community. The survey was run in conjunction with Curtin’s Office of Research and Development to maximise its reach. Over 25% of research active researchers completed the survey. The results showed clear trends in terms of the requirements researchers place on ICT services, tools and infrastructure to support their research workflow.

SURVEY STRUCTURE
The survey was targeted at the entire research community at Curtin University, and the goal was to get participation from researchers representing all fields of research. To help achieve this objective we wanted to make sure that the survey was both quick and easy to complete, as well as understandable to non-technical researchers. To that end the survey was formulated around three key concepts:

1. focused on capability, not technology.
2. focused on gathering existing requirements, not future desires.
3. based on a simple generic research workflow, that researchers could easily relate to.

These concepts enabled us to produce a fairly short and straightforward survey, that captured the sort of capabilities that our research community requires access to right now. The diagram below highlights the various sections of the survey and illustrates how researchers only had to answer questions that were relevant to their research workflow. There were only 23 questions in total, and survey abandon rates were quite low at less than 7%. The average time taken to complete the survey was only about 8 minutes.

Figure 1: flow diagram of the survey layout

KEY FINDINGS
Gathering Research Data
- The research of most respondents is based on data collected or generated manually; as opposed to being sourced from sensors, instruments, or computer simulations.
- The vast majority of digital research data is made up of common documents and multi-media files (such as Word documents, PDFs, pictures, movies and audio recordings).
- Over half of the typical research projects required less than 25GB of storage per project.
- External hard drives and USB thumbdrives are the most commonly used medium for research data storage, followed in 2nd place by desktop computers.
- Cloud storage providers (such as DropBox) are used by over a third of the respondents.
Working with Research Data

- According to our respondents, the two major requirements for a research data storage system are:
  a) the ability to access research data at anytime, from anywhere and through any device, and
  b) the ability to collaborate and share data with colleagues.
- Collaboration is by no means limited to within the institution. More than three quarters of respondents collaborate across the Australian university sector, and more than half need to collaborate internationally and/or with industry partners.

Data Computation

- Two thirds of respondents do not think of themselves as undertaking compute intensive research; and of those that do, the vast majority uses their desktop computer.
- Less than a fifth uses external compute infrastructure (such as IVEC or NCI).
- Only a very small number use cloud computing resources.
- Over two thirds are happy with the compute resources currently available to them; those that are dissatisfied are only using their desktop computers for computation.

Research Data Retention

- More than two thirds of respondents want to be able to cite their data in publications.
- Less than one quarter see the need for their data to be published in Research Data Australia.
- A data retention period of up to 10 years meets the needs for over two thirds of the respondents. The vast majority of respondents want their research data to be accessible by others, however in most instances they want to manage this on a case by case basis.

FURTHER DISCUSSIONS

Over the next few months we will be following up with over 40% the of respondents that indicated that they would like to be involved in further discussions. We will be completing a number of in depth use cases, discussing with them requirements that weren’t covered in the survey and gathering any requirements they have for future research projects. All of this information will be compiled and added to the existing survey data, in time for it to be discussed in further detail during the presentation. It will be used to help support the development of user focused IT services for Curtin’s research community that are sustainable, scalable and supportable.

CONCLUSIONS

When looking at a broad spectrum of research activities (as opposed to considering traditional ICT focused research areas), data storage and collaboration are, by a large margin, the highest priority for the bulk of researchers. It is in these areas that significant benefits can be achieved from the uptake of eResearch tools and services by non-traditional users. This is especially true if we develop tools and services that are driven by the actual needs of the target user community, rather than by what we think they should be like according to our eResearch theories. Our survey has also shown that a significant number of researchers engage the services of commercial cloud service providers, because those services offer the kind of capabilities and convenience researchers require, and have come to expect. However, by doing this, researchers expose themselves and their institutions to potentially serious risks. To counteract this development, we, have to provide the research community with data storage and collaboration tools and services that make compliance both convenient and compliant with institutional rules, regulations and policies.

While research data services contain the largest untapped potential (and risk), there are still a considerable amount of benefits to be gained in the computation space. A lot of these benefits can be achieved by assisting those researchers currently limited to their desktop computers, in migrating to compute services (either HPC or cloud) that better meet their needs. In a lot of these cases the services are already freely available to them, so the benefits could be achieved through better information, processes, support and training.

ABOUT THE AUTHORS

Andrew Buttsworth is the Team Leader for Curtin IT Service’s eResearch Support group. He graduated from Curtin University with a BSc in Applied Science and a GradDip in Computing in 1996, before starting work in the Australian higher education IT sector. After working at the University of Western Australia for 4 years, Andrew started work at Curtin University in 2000 for the Faculty of Science & Engineering. He has moved through various positions and roles at Curtin University, and his current focus is on ensuring researchers gain assistance in using ICT to enhance their research outcomes.

Florian Goessmann is the Assoc. Director eResearch Support at Curtin IT Services. Since completing his MSc (Physics), he has held several technical and managerial positions in the Australian eResearch sector. He has comprehensive experience in the planning, development and support of large scale, collaborative data storage systems. His main goal is to provide researchers with sustainable, scalable and supportable IT services that actually meet their requirements.