

User Engagement in Research Data Curation

Luis Martinez-Uribe¹ and Stuart Macdonald²

¹ University of Oxford, e-Research Centre, 7 Keble Road, Oxford OX1 3QG, UK

luis.martinez-uribe@oerc.ox.ac.uk

² University of Edinburgh, EDINA, 160 Causewayside, Edinburgh, EH9 1PR, Scotland

stuart.macdonald@ed.ac.uk

Abstract. In recent years information systems such as digital repositories, built to support research practice, have struggled to encourage participation partly due to inadequate analysis of the requirements of the user communities. This paper argues that engagement of users in research data curation through an understanding of their processes, constraints and culture is a key component in the development of the data repositories that will ultimately serve them. In order to maximize the effectiveness of such technologies curation activities need to start early in the research lifecycle and therefore strong links with researchers are necessary. Moreover, this paper promotes the adoption of a pragmatic approach with the result that the use of open data as a mechanism to engage researchers may not be appropriate for all disciplinary research environments.

Keywords: digital curation, research data management, open data, digital repository services, user engagement.

1 Introduction

Research methods and practice, including scholarly communication, are experiencing a radical transformation in the digital age. New tools and infrastructures make possible the generation of digital research data outputs as well as new ways to use, share and reuse them. There is a growing acceptance of the importance of curating research data in order to preserve them and make them re-usable for future generations with libraries, computing services and other service units within academic institutions working together to develop digital repositories to curate this type of research output.

We believe that engagement with researchers, the user communities in this case, is crucial in order to develop systems that will meet their needs. Whilst some argue that open data is the way forward, it is not clear that it will help engage researchers with digital curation activities. Thus this paper will attempt to answer the following research questions:

Is open data the correct concept to engage the research community?

What other methods can be used to facilitate engagement in data curation?

2 Open Access Repositories and Researchers' Requirements – A Balancing Act

Open Access (OA) enthusiasts have written about the inevitability of 24 hours a day and 7 days a week access to all research papers and their citations “*for free, for all*

and forever.” [1] Primarily led by technological developments, the increase in the overall volume of research, the increasing uncertainty about content preservation and by the strong dissatisfaction of academic libraries subjected to constant increase of journal subscription prices, digital repositories were built and employed within research institutions far and wide [2]. Being content provider as well as user and re-user of these information systems, researchers can be regarded as the key user community. Nonetheless, it has been argued no formal detailed requirements analysis has taken place in order to identify and address researchers’ needs and concerns related to such scholarly communication systems [3]. As a result the user community has been overlooked in the developmental phases of technology design and implementation of the information system ultimately meant to serve them. Arguably the repository infrastructure developed was not, in most cases, built to address researchers’ needs but those of libraries’ and librarians. This has led to a struggle to find ways to populate repositories with researchers’ output. In recent times there have been an increase in the number of institutional and research funders OA mandates, the knock-on effect from which will see the requirement of significant investment in awareness raising activities in order to highlight the benefits to researchers of using and depositing research materials in such repository systems. Such a process may have been expedited had the library and research worlds been more closely involved in a more agile digital repository design and development with an iterative requirements phase.

3 Research Data Repositories - Learning From Experience

When it comes to the research data setting we have to approach the problem from a new perspective. We have to evolve and learn from previous experiences in order to develop repository services capable of dealing with the management and curation of research data by addressing researchers’ needs.

Although open data is becoming a widely used term, there is not a consistent formalisation of the concept. Murray-Rust [4] suggests that the concept of open source software can be extended to that of open data in that data should be freely available for re-use and modification without restriction. The virtues of ‘Open Data’ have been praised and evangelized by many since OECD’s declaration back in 2003 [5] however many research communities are currently not in a position to make their data available on those terms.

The JISC-funded DISC-UK DataShare project explored a number of technical, legal and cultural issues surrounding research data in repository environments. It built on the existing collaboration of data librarians and data managers from the Universities of Edinburgh, Oxford, Southampton and LSE and investigated mechanisms for ingesting and sharing research data in existing institutional repository systems for those researchers willing to openly share them. Project partners identified a number of barriers pertaining to the researcher and the research setting that would impact on data sharing [6], including:

- Reluctance to forfeit valuable research time to prepare datasets for deposit, e.g. anonymisation, codebook creation, formatting
- Concerns over making data available to others before it has been fully exploited