**Invisible Repositories, Re-Use and Reproducible Research**

What are the practical implications Merton’s four institutional norms of good scientific research to a digital repository: We analyze past and contemporary practice in the light of the emerging function of repositories in research data management. We conclude that repositories fulfilling this function will heavily re-use existing data in order to support the reproducibility of research results – and will shift focus from ‘shop windows’ to exposing data sources to machines. Thereby, they might become invisible to end-users.

Merton defined 4 norms of good scientific practice [1]:

1. ‘Communalism’ refers to the claim that research results are property of the community: Repositories fulfill this function when they are open.
2. ‘Universalism’ means that everybody should be able to contribute, for example independent of cultural or national origin: repositories can implement this function if they embed individual repositories in a global network.
3. ‘Disinterestedness’ requires the greater scientific good to be valued higher than personal interests: repositories can support this function by providing a neutral and unbiased platform for knowledge resources.
4. ‘Skepticism’ implies critical scrutiny of research results: repositories can enable scrutiny through exposing research results to the scientific community.

The ways in which open repositories support Merton’s norms are so obvious that its explanation appears almost trivial. But the practical implications may be substantial and manifold.

The recent discussion about using repositories in research data management might result in a renaissance of the idea of original idea of the “Institutional Repository” [2]. And it is noteworthy in this context that Merton refers to his norms as institutional properties. However, the idea of the Institutional Repository (IR) has not been without criticism, especially with respect to the observation that many of them are empty [3]. It is therefore the right time to ask critical questions about lessons learnt and what can be done better in the future.

Considering Merton’s norms, it could be assumed that IRs have failed to put the value for research in the center of its raison d'être and introduced repositories as an alien to research practice.

We will analyze selected examples of practice (see also Table 1). For the sake of simplicity, we will focus on the function of a repository as managing text-based publications for scholarly journal articles, i.e. bibliographic data. Genuine research data management – as well as digitizations and monograph publications – will not be considered in depth. However, as a minimal common denominator between bibliographic data management and research data management, we consider the guidance on open access by RCUK that papers *“… must include … a statement on how the underlying research materials – such as data, samples or models – can be accessed …”* [4].

Table 1. Different practices in Institutional Repositories.

|  |  |  |
| --- | --- | --- |
| **Conventional Practice** | **Emerging Practice** | **Research Data Practice** |
| IRs relying on manual data input | IRs relying on pre-existing data: CrossRef, Web of Knowledge, PubMed, ArXiv, Inspires, monograph catalogues etc | IRs relying on minimal metadata, automatic data creation and the re-use of data from DataCite, Dryad, NCBI, World Data Centers |
| IRs duplicating data available elsewhere | IRs support disambiguation of author names, department names, Grant IDs as well different sources and versions as well as different citation formats | IRs focus on use cases for data not held elsewhere (long-tail) and enrich bibliographic data with data-links |
| IRs providing a ‘shop window’ | IRs support discovery in search engines, embedding in personal pages, departmental pages and offer data via APIs (e.g. for research funders’ systems) | IRs provide private access for researchers and collaborative groups in the data creation phase and support data mining via APIs |

Considering the practices shown in Table 1, the following principles for IRs are proposed:

1. Re-use: IRs should seek the re-use of existing data wherever possible to prevent double effort.
2. Unique Value: IRs should focus on functions that are not already provided elsewhere.
3. Embedding: IRs should be embedded in the researcher’s everyday life and display information in tools and services that the researcher uses.

By providing an institutional version of record for publications and linking this record to (records) of research data, IRs could achieve a vital role in the supporting the reproducibility of research in an ever more complex and volatile system of research communication. However, a deep understanding of research practice, reliability of operations and attention to detail are indispensable requirements.

When following these principles, it may well be that repositories will adopt a completely different shape from what we know today – they might become invisible.

Demonstrations on ORA, and DaMaRO, especially DataBank and DataFinder will be provided in the presentation as practical examples.

Merton, Robert K. *The sociology of science: Theoretical and empirical investigations*. University of Chicago press, 1979.

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *portal: Libraries and the Academy*, *3*(2), 327-336.

Swan, Alma, and Leslie Carr. "Institutions, their repositories and the Web."*Serials review* 34.1 (2008): 31-35.

Research Councils UK. Website accessed 4 March 2013. http://www.rcuk.ac.uk/documents/documents/RCUK%20\_Policy\_on\_Access\_to\_Research\_Outputs.pdf (2013)