Comparing Repository Types: Challenges and Barriers for Subject-Based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication

Chris Armbruster, Max Planck Society, Germany
Laurent Romary, INRIA and Humboldt Universität zu Berlin, Germany

ABSTRACT

After two decades of repository development, some conclusions may be drawn as to which type of repository and what kind of service best supports digital scholarly communication. In this regard, four types of publication repository may be distinguished, namely the subject-based repository, research repository, national repository system, and institutional repository. Two important shifts in the role of repositories may be noted and in regard to content, a well-defined and high quality corpus is essential. This implies that repository services are likely to be most successful when constructed with the user and reader in mind. With regard to service, high value to specific scholarly communities is essential. This implies that repositories are likely to be most useful to scholars when they offer dedicated services supporting the production of new knowledge. Along these lines, challenges and barriers to repository development may be identified in three key dimensions, i.e., identification and deposit of content, access and use of services, and preservation of content and sustainability of service. An indicative comparison of challenges and barriers in some major world regions is offered.

Keywords: Deposit Mandate, Institutional Repository, National Repository System, Repository Infrastructure, Research Repository, Scholarly Communication, Subject-Based Repository

REPOSITORIES AND SERVICES MENTIONED (SELECTION)

ARROW, arXiv, BASE, CiteSeer X, DRIVER, HAL, HighWire Press, Japan DRF, MIT OCW, NARCIS, OAIster, PMC, PMC Canada, Re-...
an identification of challenges and barriers. In service of further analysis, the main repository solutions are here reconstituted as ideal types. Ideal types are abstract types, derived partly from the history of repositories, partly through logical reasoning. The relevant literature on scholarly communication, open access and repositories is appreciated (cf., Bailey, 2008, 2009, 2010), though the following is not a literature review but an argument that moves back and forth between abstract ideal types and specific cases. The idea is not to classify each and every repository as belonging unambiguously to a particular type. Rather, the purpose of creating ideal types is to compare and contrast the types so as to generate insight into repository development generally as well as for each individual instance. This implies that the new knowledge thus constituted may enhance the agency of stakeholders and managers in improving and adapting their repository solution.

The four proposed ideal types may be described as follows:

- **Subject-based repositories** (commercial and non-commercial, single and federated) usually have been set up by community members and are adopted by the wider community. Spontaneous self-archiving is prevalent as the repository is of intrinsic value to scholars. Much of the intrinsic value for authors comes from the opportunity to communicate ideas and results early in the form of working papers and preprints, from which a variety of benefits may result, such as being able to claim priority, testing the value of an idea or result, improving a publication prior to submission, gaining recognition, achieving international attention and so on. As such, subject-based repositories are thematically well defined, and alert services and usage statistics are meaningful for community users;

- **Research repositories** are usually sponsored by research funding or performing organisations to capture results. This typically requires a deposit mandate. Publications are results, including books, but data may also be considered a result worth capturing, leading to a collection with a variety of items. Because these items constitute a record of science, standards for deposit and preservation must be stringent. The sponsor of the repository is likely to tie reporting functions to the deposit mandate, this being, for example, the reporting of grantees to the funder, or the presentation of research results in an annual report. Research repositories are likely to contain high-quality output. This is because its content is peer-reviewed multiple times (e.g., grant application, journal submission, research evaluation) and the production of the results is well funded. Users who are collaborators, competitors or instigating a new research project are most likely to find the collections of relevance;

- **National repository systems** require coordination - more for a federated system, less for a unified system. National systems are designed to capture scholarly output more generally and not just with a view to preserving a record of scholarship, but also to support, for example, teaching and learning in higher education. Indeed, only a national purpose will justify the national investment. Such systems are likely to display scholarly outputs in the national language, highlight the publications of prominent scholars and develop a system for recording dissertations. One could conceive of such a national system as part of a national research library that serves scholarly communication in the national language and supports public policy, e.g. in generating open educational resources for higher education and enhancing public access to knowledge;

- **Institutional repositories** contain the various outputs of the institution. While research results are important among these outputs, so are works of qualification, and teaching and learning materials. If the repository captures the whole output,