Outcomes of the USRN Discovery Pilot Project

Summary: U.S. Repository Network Discovery Pilot

January 2025

Overview

An interoperable network of repositories is an essential component of our national research infrastructure, offering rapid and open access to research, and plays a crucial role in collective efforts to transform global research communications, leading to a more open, inclusive, and equitable system. The recent OSTP memorandum on "Ensuring Free, Immediate and Equitable Access to Federally Funded Research" offers a timely opportunity to raise awareness of the benefits of a distributed network of repositories based at universities and research centers in supporting compliance with emerging policies. The United States Repository Network (USRN) was launched in 2022 to strengthen repositories in the U.S. and create a more cohesive, connected network as part of the national research infrastructure. The initiative is a collaboration between SPARC and COAR.

In 2023, the USRN launched a pilot project to assess the current practices of U.S. repositories in terms of discoverability and other good practices. The pilot involved 23 repositories of varying sizes and types, ranging from prestigious universities to smaller institutions to subject-specific repositories. The aim was to have a representative enough sample that the findings could be usefully extrapolated to the larger U.S. network. The pilot was undertaken by CORE, a U.K. not-for-profit aggregator, and Paul Walk of Antleaf, an expert consultant in repository technology.

The Process

The pilot project involved a number of activities that would help (1) determine current practices at repositories and (2) develop support and resources to help participating repositories improve their practices. This included an initial technical evaluation as well as a survey of current practices, availability of resources, and appetite for technical development.

The initial assessment found that a significant number of repositories were not applying good practices in a number of areas, and this was often due to the limited staff and technical resources at the institution. In addition, there was a relatively low awareness of various tools and functionalities that are available to have a fully operational repository.

Based on this assessment, participating repositories were offered support to help address their current shortcomings. The support was provided in several ways: through one-on-one interactions, webinars, a knowledge base of information resources, and through automated checking via the CORE dashboard.

Outcomes

The pilot project resulted in significant positive outcomes for the U.S. repository community. In particular, an increase in the discoverability of content in participating repositories. At the beginning of the project, about half of the repositories did not have their OAI-PMH interface properly configured and, therefore, could not be indexed by external discovery systems. After just over a year of the pilot, all but one repositories are now OAI-PMH compliant. This has resulted in a 50% increase in indexed content, with 728,770 new records now publicly accessible. In addition, important support resources were developed through the project that will be made available to the entire USRN community:

- An automated tool by CORE to specifically assess repositories' compliance with the USRN Desirable Characteristics for Publication Repositories;
- CORE Data Provider's Guide, which provides comprehensive instructions for repository indexing, is applicable beyond the project;
- USRN Desirable Characteristics Toolkit, a wiki of valuable resources providing guidance on metadata, discoverability, and persistent identifiers.

Conclusions

There are over 1,000 repositories in the United States that play a critical role in collecting, preserving, and providing access to a range of valuable research outputs. However, not all of these repositories use optimal practices that are needed to ensure their outputs are visible both in the US and internationally. In addition, the new functionalities on the horizon that will allow repositories to participate in innovations and leading-edge research are often not being adopted due to a lack of awareness or because of low resourcing at the repository. The outcomes of this pilot show that with a relatively small effort, repositories can significantly improve their practices.

As a next step, we will expand the availability of the resources developed in the project to the broader USRN network while engaging in key strategic discussions about why good practices and interoperability are important.

Project team
- Jennifer Beamer, Visiting Program Officer, USRN/SPARC
- Matteo Cancellieri, CORE Lead developer
- Heather Joseph, Executive Director, SPARC
- Petr Knoth, Founder and Head of CORE; Professor of Data Science, The Open University
- David Pride, CORE, Research Fellow, Open University
- Kathleen Shearer, Executive Director, COAR
- Halyna Torchylo, CORE, Liaison and support
- Michael Upshall, CORE Communication and Engagement Manager
- Paul Walk, Antleaf Ltd, Technical Consultant