Persistent Identifiers to Connect Research in Shared Research Repository and Beyond

Ilkay Holt, Sara Gould, Jenny Basford, Frances Madden, Rachael Kotarski, British Library
Ellen Ramsey, University of Virginia

This work presents implementation of identifiers in British Library's Shared Research Repository[1]. The Shared Research Repository Beta was launched after announcing a pilot project in 2019. It is created in collaboration with five project partners who are all UK cultural and heritage organisations: the British Museum, Tate, National Museums Scotland (NMS), Museum of London Archaeology (MOLA), and Royal Botanic Gardens, Kew. The shared repository allows users to search across the combined content, meaning that common research topics and collaborative activities can be discovered and explored through a single search.

The repository runs on Samvera Hyku application and provides partners with hosted, multi-tenant repository infrastructure sharing the same features at the technical level but also independence in terms of local curation and individual branding. Ubiquity Press has been the service provider and developed the current infrastructure.

British Library is passionate about identifiers in terms of data discovery and re-use of data. It is a founding member of DataCite, which assigns persistent digital object identifiers (DOI names) and the Library acts as a UK consortium lead for the DataCite. The British Library is also a project partner of the project FREYA[2], funded by the European Commission and ending in November 2020. The project aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally.

Identifiers help to connect research entities together across different repositories and silos of information. As persistent and unambiguous links, they are beneficial for integration and data exchange across systems. In repositories in particular, they enable citation of repository content, accurate metrics, and automated population of metadata.

Adoption of identifiers at the Shared Research Repository is a reflection of the commitment to improve access to, reuse of and tracking impact of research. Identifiers are in place for four entities within the Shared Research Repository including author, research, organisations, and funder IDs:

Author IDs
British Library uses both ISNI and ORCID to identify authors in a record. In the deposit form "Creator ISNI" and "Contributor ORCID" are entered where they are available. This functionality supports disambiguation of authors and creators, but will also allow them to link to repository records in their author profiles in the future.

Research IDs
The Shared Research Repository offers an extensive metadata including DOIs. The repository allows users to mint DOI where they can assign a new DOI to a work. This DOI

---

[1] [iro.bl.uk]
[2] [www.project-freya.eu]

can be saved as a draft or made private, but can also be registered and/or findable in DataCite's DOI registry. The repository can also fetch metadata by using existing DOIs to search for a publication from DataCite and Crossref. This populates the metadata fields automatically on the deposit form.

Organization and Funder IDs

ROR, Research Organization Registry[3], is a recent initiative launched in January 2019 and assigns identifiers for organizations. ROR allows crosswalk with other identifiers for the organization e.g. GRID, ISNI, Crossref Funder Registry, Wikidata. The implementation of ROR IDs has recently been done at the Shared Research Repository. The funder example given on the poster has both ISNI and ROR IDs based on a lookup of the Crossref Funder Registry. ISNI and ROR IDs can be added for organisational creators and contributors and the repository also captures Wikidata and GRID for organisational authors at the backend, not visible on the frontend. These functionalities bring full integration of identifiers exposed with FREYA Project and allow long term interoperability with identifier registration agencies e.g. DataCite and Crossref.

Such advances at the Samvera's Hyku instance at the British Library's Shared Research Repository will feed into the objective of making institutional repositories a more attractive dissemination option for research organisations. This objective is set for the Advancing Hyku project[4] which British Library is a partner of, funded by Arcadia, a charitable fund of philanthropists Lisbet Rausing and Peter Baldwin, from October 2019 to August 2021. The project aims to support the growth of open access through institutional repositories, by introducing significant structural improvements and new features to the Samvera Community's Hyku platform. The project partners are University of Virginia Library, Ubiquity Press and the British Library. The advances will drive green open access and will increase value to researchers with a combined approach of auto-population and expanding the array of integrated, open source services.

Activities with in the project include:
- Content auto population by identifying relevant content and importing metadata from sources (Unpaywall, OA Button, and OA publishers such as PLOS, Hindawi, etc.)
- metrics for institutions and authors demonstrating research impact, readership, interaction, possibly ingesting into CRIS systems and for reporting
- author profile sync by making self deposit easier, identifying authors' contents and importing metadata for them from sources (ORCID, Sherpa/Romeo, etc.).
- Investigating pathways to long-term preservation and community codebase contributions.

To achieve these, the project team closely works with the Samvera Community.

The current implementation of Hyku for the British Library's Shared Research Repository will form the baseline for the Samvera Hyku application towards the target state architecture for the Advancing Hyku project. British Library reviewed the architecture and design of the Hyku implementation at its current state as one of the deliverables of the project. This review[5]

---

[3] [ror.org]
[4] [advancinghyku.io]
[5] [doi.org/10.18130/v3-k4an-w022]

outlines the scope of the requirements and the main principles and guidelines that drive the Advancing Hyku architecture and design. The main aspects considered for review of the product are the merging of features from Hyku community into current British Library's version and vice versa, multi-tenancy for a provider-consumer model, interoperability with other systems e.g. DataCite, Crossref, OAI-PMH API communication, portability where components should be deployed for Cloud as well as on-premise infrastructure, stability (e,g. high availability using fault tolerant mechanisms), scalability (e.g. performance should not degrade with increasing volume of data), and security including role-based secure models for tenant admins and data component design to be compliant with regulations.

Advancing Hyku project builds on these requirements and principles. The project team is currently working on merging the existing features and contributing back to the Hyku/Hyrax Core codebase. DOI minting and auto population are features soon to be in the codebase[6]. In terms of identifiers, this will serve the Samvera Community in the short term. The Advancing Hyku project team continues to generate discussions and collaborate with other community projects. Some of those activities are to align with Hyku Community Roadmap, to collaborate with Hyrax Analytics to divide and conquer metrics and analytics, and to partner with Hyku for Consortia to enhance and prioritise codes developed for multi-tenant repositories. Advancing Hyku project is coming to the end of first year and a brief quarterly project updates can be found on the website https://advancinghyku.io .

---

[6] [https://github.com/samvera/hyrax/pull/4458]