

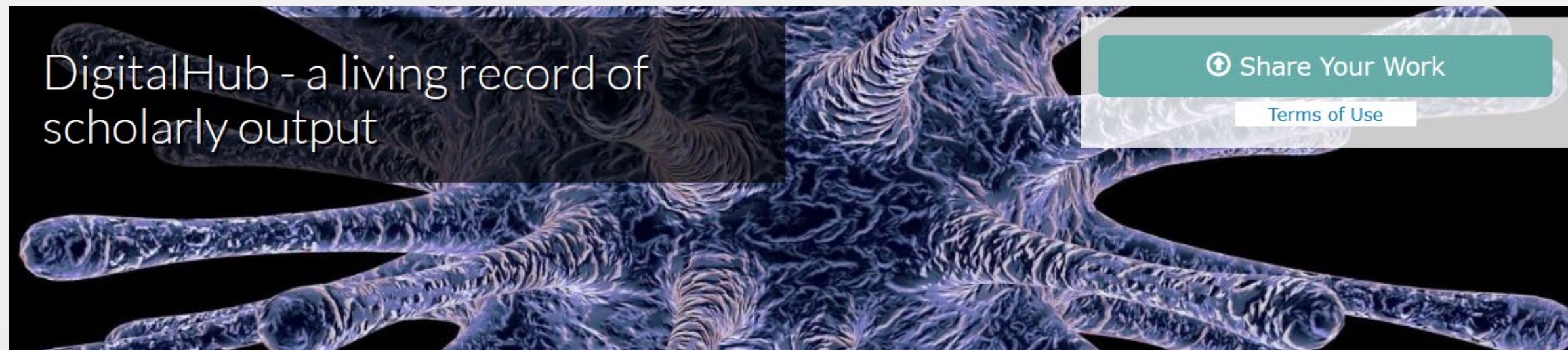
# Beyond Accessibility: Taking the Institutional Repository to the Next Generation

Sara Gonzales, Guillaume Viger, Matthew Carson, Kristi Holmes  
Northwestern University, Chicago, IL



## Introduction

In October 2015 the institutional repository DigitalHub ([digitalhub.northwestern.edu](http://digitalhub.northwestern.edu)) was launched at Northwestern University's Feinberg School of Medicine. The repository is the product of the School, its library, Galter Health Sciences Library & Learning Center (Galter Library), and the Northwestern University Clinical and Translational Sciences Institute's (NUCATS) commitment to reproducible science, discovery of the products of research, and preservation and accessibility of scholarly outputs.



### DigitalHub technical features:

- Fedora-back end (repository software) + Hydra (now Samvera)
- Apache Solr search indexing
- Sufia front-end and user interface

### DigitalHub FAIR<sup>1</sup> data and interoperability features:

- *Findability*: assignment of a unique identifier (DOI) and rich metadata descriptions
- *Accessibility*: record retrievability through free and open Internet protocols that allow authentication and authorization where necessary
- *Interoperability*: Leverages commonly used metadata standards for medical institutional repositories [e.g., Medical Subject Headings (MeSH)]
- *Re-usability*: deposits assigned licenses for clear re-use parameters and provenance

Since 2015, over 4,800 research objects have been uploaded to DigitalHub. Half a dozen popular collections have been created, including Dialogues in Oncofertility and the Science in Society Scientific Images Contest.

## Development Methodology

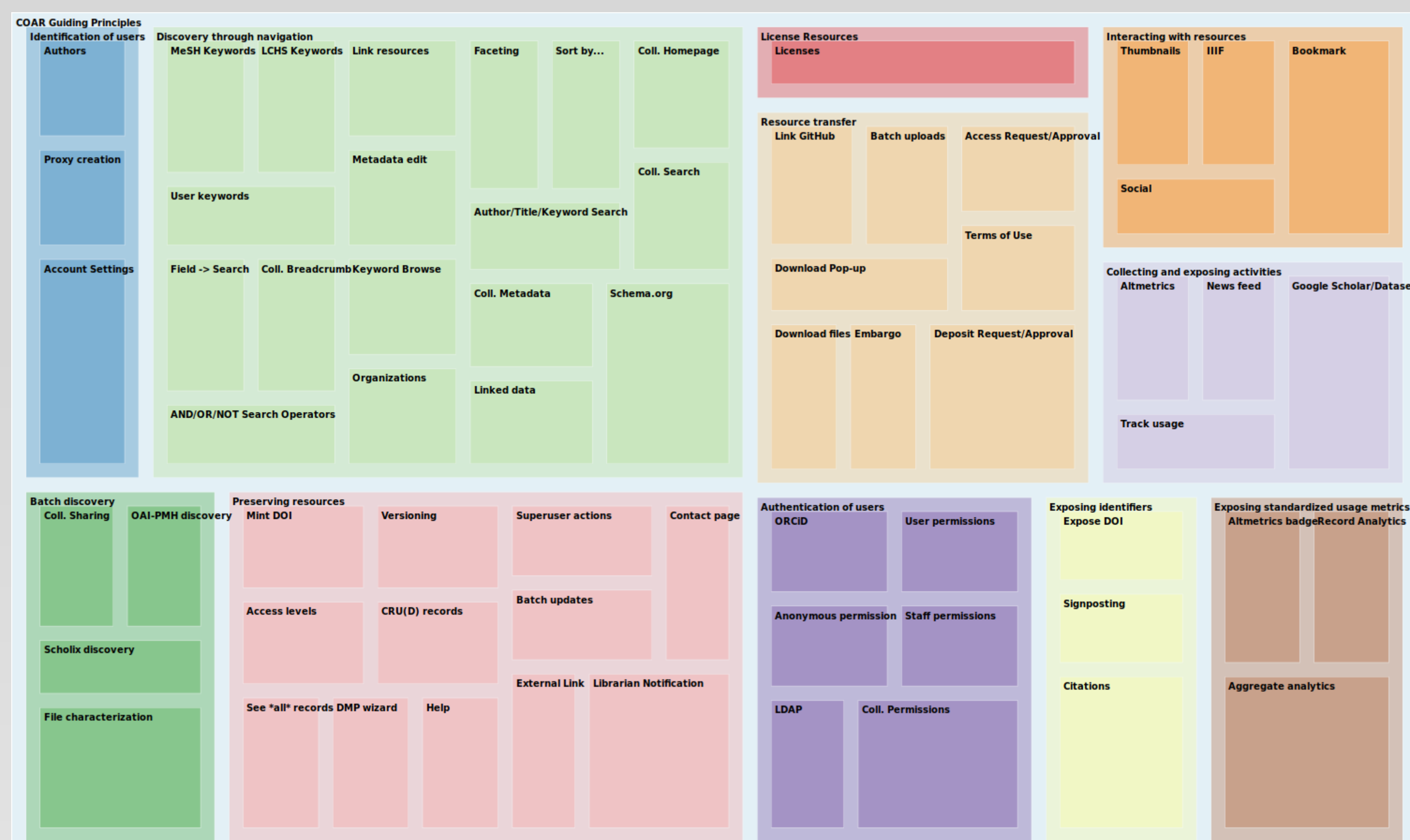
Development on menRva has proceeded locally at Galter Library since mid-2018. GitHub project milestones include:

- 1) DigitalHub parity
- 2) Next Generation Repository: includes functionalities such as leveraging APIs for linked data sources and social activity feeds
- 3) Wishlist, a category consisting of additional next generation features and librarian-required features to enable easier updates and enhancement of records.

## Methodology: Current Analysis

As requirements were gathered and informed from the literature, the project team sought to visualize the ways in which menRva's requirements improved on the repository features of DigitalHub. In addition we wanted to visualize how menRva's customizations differed from Invenio's standard features. While Invenio has many next generation repository features, menRva will build upon and customize them for the Northwestern and CD2H audiences. To achieve this we exported our issues from GitHub and organized them into the functional categories for institutional repositories that emerged through requirements gathering and a literature review. We cleaned the resulting data in OpenRefine and added data from an analysis of Zenodo. This led to the visualization shown in Figure 1, which demonstrates how menRva's customizations meet many of the requirements for next generation repositories, while going beyond in some areas of resource interaction. Meanwhile Figure 2 demonstrates how the requirements listed in the Next Generation Repository milestone map to COAR's Behaviors and Technical Recommendations for Next Generation Repositories.

Figure 2: menRva's next-generation requirements mapped against the COAR's Behaviors and Technical Recommendations for Next Generation Repositories



## COAR, CD2H, and changes in the IR landscape

With recent updates in web and linked data technologies, the institutional repository landscape has continued to evolve. Galter Library has committed to the cutting-edge technology by joining COAR, the Confederation of Open Access Repositories and its Next Generation Repositories working group. Outputs of this group, such as the Guiding Principles of Next Generation Repositories and recommendations on the behaviors and technical aspects of such repositories<sup>2, 3</sup> have had a significant impact on the direction of Galter's repository development.

Galter Library was empowered to switch to the Invenio repository framework through its 2017 award of the National Center for Data to Health (CD2H) grant from the National Institutes of Health.<sup>4</sup> Through the grant Galter Library has partnered with CERN on Invenio development, and to survey both the repository landscape and a group of local users at Northwestern, the Institute for Innovations in Developmental Sciences (DevSci), to develop requirements for the local instance of Invenio, dubbed menRva. In time menRva will be released for adoption by all of the 62 Clinical and Translational Science Award (CTSA) Program hubs.

### Invenio next generation repository features<sup>5</sup>:

- Open-source and scalable Python-based code
- Flexible data model based on JSON
- Supports preservation by allowing minting of persistent identifiers (DOI), and by running fixity checks and exporting archival packages if archiving elsewhere
- Leverages the power of Python by offering connections to packages through REST APIs.

### menRva's customizations for the FSM and CD2H audiences:

- Social media functions: news feed, follow other researchers
- Co-PI record approval functions
- Batch uploads
- HTML signposting

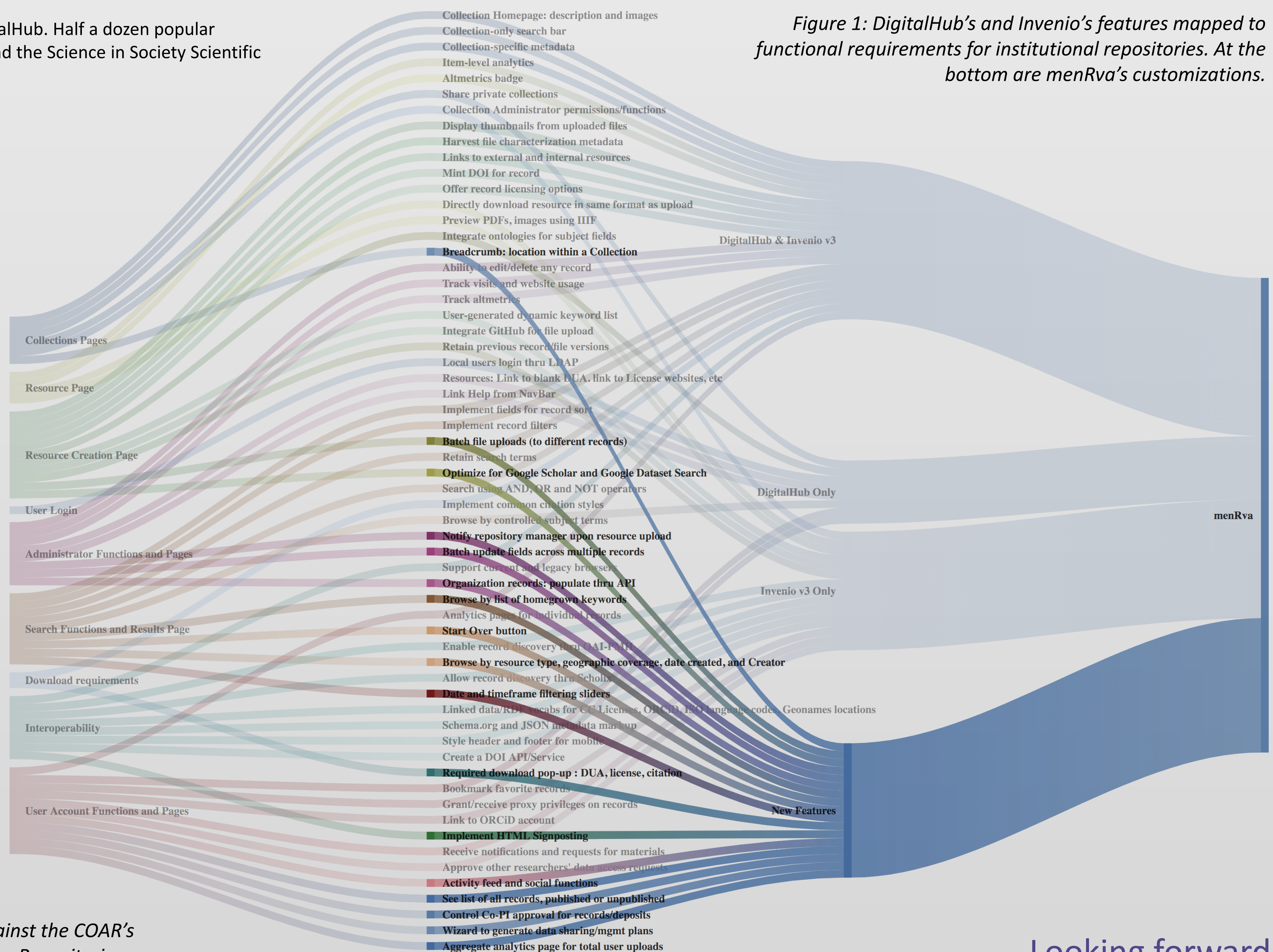


Figure 1: DigitalHub's and Invenio's features mapped to functional requirements for institutional repositories. At the bottom are menRva's customizations.

## Looking forward

To be an effective tool and to keep up with developments in the repository technology landscape, institutional repositories should not be thought of as static, one-time implementations, but rather as works in progress. Galter Library's development of the Invenio-based repository will last over the next few years, but in the meantime DigitalHub will grow and evolve and continue to house the research products of the Feinberg School of Medicine. Requirements we have tagged as Parity+ are ones that benefit DigitalHub as well as menRva, since an improvement to DigitalHub was inspired as a result. Similarly, Galter Library continues to study the repository software landscape and to watch for developments from COAR and the repository community. By striving for ever-increasing accessibility and interoperability, Galter Library will take the repository beyond discoverability and into the next generation.

## References

1. Adapted from Force 11. "The FAIR Data Principles." <https://www.force11.org/group/fairgroup/fairprinciples>
2. COAR. Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group, November 28, 2017. <https://www.coar-repositories.org/news-media/publications/>
3. COAR Guiding Principles for Next Generation Repositories. <http://ngr.coar-repositories.org/>
4. CTSA Program National Center for Data to Health. National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number U24TR002306. [https://projectreporter.nih.gov/project\\_info\\_description.cfm?aid=9568055&icde=36248767](https://projectreporter.nih.gov/project_info_description.cfm?aid=9568055&icde=36248767)
5. Invenio. "Features." <https://invenio-software.org/#features>

This work was supported by the National Institutes of Health's National Center for Advancing Translational Sciences, grants U24TR002306 and UL1TR001422.