SciDataCon 2025



Contribution ID: 73

Type: Presentation

FAIR for Now and into the Future: Building Blocks for Long-Term Data Stewardship in a Shared Data Repository Service

Tuesday 14 October 2025 12:25 (11 minutes)

Funding bodies and publishing venues increasingly require researchers to deposit and share their data in order to support rigorous, responsible, and reproducible science. Rising to the occasion, libraries have been expanding their scope to support research data as a scholarly resource and are increasingly recognized as providers of research data management and repository services. These trends are driven by the FAIR principles of findability, accessibility, interoperability, and reusability, which aim to enhance data quality to improve discovery and reuse. Data that is FAIR today may not be FAIR tomorrow, however, and ensuring that research data can be found, accessed, understood, and used into the future requires stable infrastructure and ongoing and informed interventions by the stewards of that data. Preserving research data to support these principles for the long term reveals several challenges, however, including developing policies and procedures, determining the various stakeholders that should be involved and the roles that they have in these processes, considering needs across the heterogenous and evolving landscape of research data (e.g. big data, sensitive data, disciplinary practices and protocols) and, by extension, research communities, as well as managing the associated costs and capacity required for curation, repository management, storage, ongoing preservation, and more.

This presentation will outline efforts at Borealis to develop technical building blocks and foster communitydriven initiatives in support of research data preservation in Canada. Borealis, the Canadian Dataverse Repository, is a multi-disciplinary, bilingual, national data repository service provided by Scholars Portal at the University of Toronto Libraries in partnership with regional academic library consortia and the Digital Research Alliance of Canada. With almost 24,000 published datasets, the service supports over 80 institutions and research organizations across Canada, and each institution or research organization manages their own collection and provides local support to their researchers. Borealis is library-led, community-based infrastructure, and the preservation of deposited data is a responsibility shared between Borealis as the repository service and each participating institution as an expert on their institutional context, research communities, and data collections.

To lighten the load for minimum preservation, Borealis has developed technical building blocks for the repository as a whole, including ongoing backup processes and monthly fixity checks for all deposited files, as well as fixity remediation workflows developed in line with best practices. These building blocks are complemented by preservation-friendly features provided in the repository software. Export and integration options are also available for further preservation processing and management beyond Borealis for institutions interested in engaging in additional research data preservation activities. A 2022 survey of institutional administrators further identified support for preservation planning and workflows as a primary gap in the research data services landscape. Responding to these needs, Borealis relaunched a community initiative to update and develop resources and documentation related to policy and service modelling. This building block is based on CoreTrustSeal requirements to directly support participating institutions seeking to apply for CoreTrustSeal certification as well as those interested in using the requirements to benchmark and plan their services. Bridging the technical and community, Borealis is also undertaking a format analysis project to support institutions with file format management and preservation planning.

By providing an overview of the various preservation building blocks developed for Borealis, this presentation will demonstrate the extensive work and coordination by the service provider, collection administrators and depositors required to ensure the long-term stewardship of research data. Collaborating on projects with the community to draw on our collective knowledge and experiences also provides extensive communal benefits for building local capacity and ensuring data can continue to be FAIR into the future.

Primary author: SHI, Julie (Scholars Portal)

Presenter: SHI, Julie (Scholars Portal)

Session Classification: Presentations Session 3: Rigorous, responsible and reproducible science in the era of FAIR data and AI

Track Classification: SciDataCon2025 Specific Themes: Rigorous, responsible and reproducible science in the era of FAIR data and AI