

DATA CURATION

What Is Data Curation?

Data curation¹ is the active management of research data as it is created, maintained, used, archived, shared, and reused. It is an iterative process that adds value to scholarship by optimising datasets for current use as well as future discovery and reuse. The ultimate goal of data curation is to create data sets that epitomize the FAIR² guiding principles of Findability, Accessibility, Interoperability, and Reusability.

Why Curate Data?

Data curation provides benefits to a broad range of stakeholders:

Research Communities improve transparency when data can be properly validated and replicated.

Research Funders (and the public) improve return on investment, as research products are more easily discovered and reused in new scholarship.

Individual Researchers experienced easier data management and collaboration, as well as increased citations for publications and published datasets.

Data Repository Operators benefit from higher quality products that are more discoverable, more interrelated, and better prepared for long-term preservation.

Who Are Data Curators?

Data curators have various roles and may interact with data in varying capacities throughout the research lifecycle:

- Research team members maybe (in) formally assigned a variety of data curation responsibilities
- Data curation specialists may work within organizations that collect, organize, and share data

¹ A fuller definition of data curation can be found in the [CASRAI glossary](#).

² See Wilkinson et al., doi:10.1038/data.2016.18

- Information professionals working in libraries and other research support units
May consult on practices and curate data for deposit into data repositories

What Do Data Curators Do?

Data curators apply their disciplinary knowledge, information management skills, and software expertise to improve data FAIRness. Data curators perform a broad range of activities that include:

- Preparing for data collection and documentation
- Creating documentation and metadata to explain and contextualize data
- Implementing secure storage and version control for files and code
- Preparing datasets for deposit in a data repository or archive
- Performing quality assurance through metadata inspection, file audit, and code review
- Transforming file formats to enhance reuse and long-term preservation
- Augmenting metadata and linking datasets to increase discoverability

Creating maintaining using archiving sharing reusing = data curation life cycle

Data Curation Lifecycle



Creating



Maintaining



Using



Archiving



Sharing



Reusing

How Is Portage Addressing Data Curation Needs?

The Portage curation expert group³ (CEG) aims to develop a national approach to data curation in Canada through a variety of activities⁴, including:

- Promoting best practices in preparing data and metadata for research, dissemination, and preservation

³ Learn more on the [Portage CEG webpage](#).

⁴ Refer to the Data Curation Networks Cross-Institutional Staffing Model for a more detailed list of data curation activities. <http://hdl.handle.net/11299/188654>

PRIMER

- Developing training materials and workshops to enhance data curation expertise across Canada
- Building a community of practice for Canadian data curators
- Promoting the adoption of standard data curation processes across Canadian data repositories
- Envisioning a nationally coordinated curation Network that leverages the unique skills and domain - specific knowledge of curators to meet research data curation needs across Canada



Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

Portage Network | portage@carl-abrc.ca | portagenetwork.ca

CONTACT AT YOUR INSTITUTION: