

Primer:

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 optimizing datasets for current use as well as future discovery and reuse. The ultimate goal of data curation is to create datasets 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

experience 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 may be (in)formally assigned a variety of data curation responsibilities
- Data curation specialists may work within organizations that collect, organize, and share data
- Information professionals working in libraries and other research support units may consult on practices and curate data for deposit into data repositories

¹ A fuller definition of data curation can be found in the CASRAI glossary, https://dictionary.casrai.org/data_curation

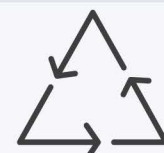
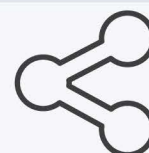
² See Wilkinson et al., 2016. doi:10.1038/sdata.2016.18

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 activities³ 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 Lifecycle

HOW IS PORTAGE ADDRESSING DATA CURATION NEEDS?

The Portage Curation Expert Group⁴ (CEG) aims to develop a national approach to data curation in Canada through a variety of activities, including:

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

³ Refer to the Data Curation Network's Cross-Institutional Staffing Model for a more detailed list of data curation activities.
<http://hdl.handle.net/11299/188654>

⁴ Learn more on the Portage CEG webpage, <https://portagenetwork.ca>



This guide was created by the Portage Curation Expert Group and can be modified and re-used freely under the CC-BY license.