Part Three

Data Wrangling

CURATE

Workflow model created by the Data Curation Network.

Check files and read documentation

Understand the data

Request missing information

Augment documentation and metadata

Transform file formats for reuse

Evaluate for FAIRness

Check

- Review the content of the data files (open files in viewer or application, run code, etc.)

- Review documentation (Metadata, readme files, instructions, related articles)

Understand

- Are any of the tags or metadata ambiguous? Is there data missing? If there is code, does it run?

- Is all the documentation where it should be?

- Is there enough documentation and is the data clearly arranged enough that another researcher with the same background could be reasonably expected to be able to reuse the data?

Understand

- Recommended content
- General information
- Data and file overview
- Sharing and access information
- Methodological information • Data-specific information

https://data.research.cornell.edu/content/readme

Request

- Evaluate gaps in information revealed in the Check and Understand steps.
- Decide on priority of information
- Contact researcher (preferably via email) to ask about missing information
- Keep it simple and to the point
- Get permission to do the work yourself.

Workflow model created by the Data Curation Network.

Augment

- Makes changes, add data, any other steps suggested by the new data that the researcher has agreed to
- Create metadata based on the original and revised documentation
- Standardize - For example ISO 8601 (YYYY-MM-DD Format)
- Compile curation log - Everything done to the data, correspondance with researcher, etc.

Workflow model created by the Data Curation Network.

Title

Virginia water central

Tag	Attribute	Language	Data
Title			Virginia water central
Publisher			Virginia Water Resources Research Center, Virginia
Date			2007 -
Format			PDF
Language			English
Subject			Water -- Pollution -- Law and legislation -- Virginia -- Periodicals
Subject			Water resources development -- Virginia -- Periodicals
Subject			Water-supply -- Virginia -- Periodicals

Update

Cancel

Basic Metadata

- Author
- Title
- Keywords
- Abstract
- Language
- Citation
- Any links necessary
- Licensing information
- Funding information

Transform to Archival Formats

- Open formats: CSV, TXT, PDF instead of DOCX, XSLX, etc.
- Lossless rather than lossy
- Widely adopted
- Independant (not tied to a particular application, OS, or hardware)
- Library of Congress Recommended Formats

Workflow model created by the Data Curation Network.

Evaluate FAIRness

Findable

Accessible

Interoperable

Reusable

Workflow model created by the Data Curation Network.

Examples

Code File

Your Turn

https://drive.google.com/open?id=1dNke_YUMpVv77XU0SW8m6vMuy-VOy7y-