Dec 11th, 10:40 AM - 11:00 AM

Checklist of DCN CURATE Steps

Lisa Johnston
University of Minnesota - Twin Cities, ljohnsto@umn.edu

Jake R. Carlson
University of Michigan - Ann Arbor, jakecar@umich.edu

Wendy Kozlowski
Cornell University Library, wak57@cornell.edu

Heidi Imker
University of Illinois at Urbana-Champaign, imker@illinois.edu

Robert Olendorf
The Pennsylvania State University, rko5039@psu.edu

See next page for additional authors

Follow this and additional works at: https://openscholarship.wustl.edu/data-curation-workshop-2017

Part of the Library and Information Science Commons

Presenter Information

Lisa Johnston, Jake R. Carlson, Wendy Kozlowski, Heidi Imker, Robert Olendorf, and Cynthia Hudson-Vitale

Checklist of CURATE Steps and FAIRness Scorecard

(published in the [DCN Planning Phase Report (2017](#))

C   **Check** files and read documentation (risk mitigation, file inventory, appraisal/selection)

U   **Understand** the data (or try to), if not… (run files/environment, QA/QC issues, readmes)

R   **Request** missing information or changes (tracking provenance of any changes and why)

A   **Augment** metadata for findability (DOIs, metadata standards, discoverability)

T   **Transform** file formats for reuse (data preservation, conversion tools, data viz)

E   **Evaluate** for FAIRness (licences, responsibility standards, metrics for tracking use)

Notes

CHECK Step

CURATE Action	Curator Checklist

Check data files and read documentation
    ● Review the content of the data files (e.g., open and run the files or code).
    ● Verify all metadata provided by the author and review the available documentation.
	❏ Files open as expected
	    ❏ Issues _____
	❏ Code runs as expected
	    ❏ Produces minor errors
	    ❏ Does not run and/or produces many errors
	❏ Metadata quality is rich, accurate, and complete
	    ❏ Metadata has issues _____
	❏ Documentation Type (circle)
	Readme / Codebook / Data Dictionary / Other: _____
	    ❏ Missing/None
	    ❏ Needs work

Notes:

UNDERSTAND Step

CURATE Action	Curator Checklist

Understand the data (or try to)
● Check for quality assurance and usability issues such as missing data, ambiguous headings, code execution failures, and data presentation concerns.
● Try to detect and extract any "hidden documentation" inherent to the data files that may facilitate reuse.
● Determine if the documentation of the data is sufficient for a user with similar qualifications to the author's to understand and reuse the data. If not, recommend or create additional documentation (e.g., a readme.txt template).
	Varies based on file formats and subject domain. For example….
	
	Tabular Data Questions (Microsoft Excel)
	❏ Organization of data well-structured
	  ❏ Not rectangular
	  ❏ Split tables into separate tabs
	❏ Headers/codes clearly defined
	  ❏ Define headers
	  ❏ Clarify codes used _____
	  ❏ Clarify use of "blanks"
	  ❏ Clarify units of measurement
	❏ Quality control clearly defined
	  ❏ Unclear quality control
	  ❏ Update/add Methodology

Notes:

REQUEST Step

CURATE Action	Curator Checklist
Request missing information or changes ● Generate a list of questions for the data author to fix any errors or issues.	*Narrative describing the concerns, issues, and needed improvements to the data submission.*

University of Michigan sample email to researcher:

Dear [name of the person Identified as the contact for the data set as stated in the DBD metadata],

Thank you for depositing your data set, [title of the data set] to the library's Deep Blue Data repository.

After we receive a data set, we review it to ensure that the data sets we host are as complete, accessible and understandable as possible. We have reviewed your data set and have the following recommendations for you:

- Recommendation #1
- Recommendation #2
- Recommendation #3
- Recommendation #4

We look forward to hearing your response to our questions and requests for additional information.

Please do let us know if you have any questions about or recommendations. We would be happy to talk with you over the phone or meet with you in person to discuss our review of your data should you wish to do so.

Sincerely,

[Name of Liaison]

Notes:

AUGMENT Step

CURATE Action	Curator Checklist
Augment the submission ● Enhance metadata to best facilitate discoverability. ● Create and apply metadata for the data record, including descriptive keywords. ● When appropriate, structure and present metadata in domain-specific schemas to facilitate interoperability with other systems.	❏ Discoverability sufficient 　❏ Recommend (circle one) 　 full-text index / file rename / 　 file reorder / file descriptions / 　 zip files into one archive 　 Other _____ ❏ Keywords Sufficient 　❏ Suggestions _____ ❏ Linkages Sufficient 　❏ Link to report/paper 　❏ Link to related data sets 　❏ Link to source data 　❏ Link to other _____

Notes:

TRANSFORM Step

CURATE Action	Curator Checklist

Transform file formats
● Identify specialized file formats and their restrictions (e.g., Is the software freely available? Link to it or archive it alongside the data).
● Transform files into open, non-proprietary file formats that broaden the potential audience for reuse and ensure that preservation actions might be taken by the repository in later steps. Retain original files if data transfer is not perfect.
	❏ Preferred file formats in use
	  ❏ Recommend conversion
	    from _____
	    to _____
	  ❏ Retain original formats
	❏ Software needed is readily available
	  ❏ Unclear version of software
	  ❏ Unclear software used
	❏ Visualization of data easily accessible
	  ❏ Recommend graphical representation _____
	  ❏ Recommend web-accessible surrogate _____

Cornell's List of Preservation Format Recommendations
http://guides.library.cornell.edu/ecommons/formats

Notes:

EVALUATE Step

CURATE Action	Curator Checklist
Evaluate and rate the overall data record for FAIRness.* • Score the dataset and recommend ways to increase the FAIRness of the data and become "DCN approved."	Findable - ❏ Metadata exceeds author/ title/ date, ❏ Unique PID (DOI, Handle, PURL, etc.). ❏ Discoverable via web search engines. Accessible - ❏ Retrievable via a standard protocol (e.g., HTTP). ❏ Free, open (e.g., download link). Interoperable - ❏ Metadata formatted in a standard schema (e.g., Dublin Core). ❏ Metadata provided in machine-readable format (OAI feed). Reusable - ❏ Data include sufficient metadata about the data characteristics to reuse ❏ Contact info displayed if the direct assistance of the author needed. ❏ Clear indicators of who created, owns, and stewards the data. ❏ Data are released with clear data usage terms (e.g., a CC License).

* Rubric evaluating the FAIR principles are based on the scoring matrix by Dunning, de Smaele, & Böhmer (2017).

Notes: