

Washington University in St. Louis Washington University Open Scholarship

IASSIST & DCN - Data Curation Workshop

Workshop Schedule

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](#)

Johnston, Lisa; Carlson, Jake R.; Kozlowski, Wendy; Imker, Heidi; Olendorf, Robert; and Hudson-Vitale, Cynthia, "Checklist of DCN CURATE Steps" (2017). *IASSIST & DCN - Data Curation Workshop*. 4.

<https://openscholarship.wustl.edu/data-curation-workshop-2017/schedule/Schedule/4>

This Handout is brought to you for free and open access by the Conferences and Symposia at Washington University Open Scholarship. It has been accepted for inclusion in IASSIST & DCN - Data Curation Workshop by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

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 |
|--|--|
| <p>Check data files and read documentation</p> <ul style="list-style-type: none"> • 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. | <ul style="list-style-type: none"> <input type="checkbox"/> Files open as expected <ul style="list-style-type: none"> <input type="checkbox"/> Issues _____ <input type="checkbox"/> Code runs as expected <ul style="list-style-type: none"> <input type="checkbox"/> Produces minor errors <input type="checkbox"/> Does not run and/or produces many errors <input type="checkbox"/> Metadata quality is rich, accurate, and complete <ul style="list-style-type: none"> <input type="checkbox"/> Metadata has issues _____ <input type="checkbox"/> Documentation Type (circle) Readme / Codebook / Data Dictionary / Other: _____ <ul style="list-style-type: none"> <input type="checkbox"/> Missing/None <input type="checkbox"/> Needs work |

Notes:

UNDERSTAND Step

| CURATE Action | Curator Checklist |
|--|--|
| <p>Understand the data (or try to)</p> <ul style="list-style-type: none"> ● 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). | <p><i>Varies based on file formats and subject domain. For example....</i></p> <p>Tabular Data Questions (Microsoft Excel)</p> <ul style="list-style-type: none"> <input type="checkbox"/> Organization of data well-structured <ul style="list-style-type: none"> <input type="checkbox"/> Not rectangular <input type="checkbox"/> Split tables into separate tabs <input type="checkbox"/> Headers/codes clearly defined <ul style="list-style-type: none"> <input type="checkbox"/> Define headers <input type="checkbox"/> Clarify codes used _____ <input type="checkbox"/> Clarify use of “blanks” <input type="checkbox"/> Clarify units of measurement <input type="checkbox"/> Quality control clearly defined <ul style="list-style-type: none"> <input type="checkbox"/> Unclear quality control <input type="checkbox"/> Update/add Methodology |

Notes:

REQUEST Step

| CURATE Action | Curator Checklist |
|--|--|
| <p>Request missing information or changes</p> <ul style="list-style-type: none"> • Generate a list of questions for the data author to fix any errors or issues. | <p><i>Narrative describing the concerns, issues, and needed improvements to the data submission.</i></p> |

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 |
|---|--|
| <p>Augment the submission</p> <ul style="list-style-type: none"> ● 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. | <ul style="list-style-type: none"> <input type="checkbox"/> Discoverability sufficient <ul style="list-style-type: none"> <input type="checkbox"/> Recommend (circle one) full-text index / file rename / file reorder / file descriptions / zip files into one archive Other _____ <input type="checkbox"/> Keywords Sufficient <ul style="list-style-type: none"> <input type="checkbox"/> Suggestions _____ <input type="checkbox"/> Linkages Sufficient <ul style="list-style-type: none"> <input type="checkbox"/> Link to report/paper <input type="checkbox"/> Link to related data sets <input type="checkbox"/> Link to source data <input type="checkbox"/> Link to other _____ |

Notes:

TRANSFORM Step

| CURATE Action | Curator Checklist |
|--|--|
| <p>Transform file formats</p> <ul style="list-style-type: none"> • 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. | <ul style="list-style-type: none"> <input type="checkbox"/> Preferred file formats in use <ul style="list-style-type: none"> <input type="checkbox"/> Recommend conversion from _____ to _____ <input type="checkbox"/> Retain original formats <input type="checkbox"/> Software needed is readily available <ul style="list-style-type: none"> <input type="checkbox"/> Unclear version of software <input type="checkbox"/> Unclear software used <input type="checkbox"/> Visualization of data easily accessible <ul style="list-style-type: none"> <input type="checkbox"/> Recommend graphical representation _____ <input type="checkbox"/> 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 |
|---|--|
| <p>Evaluate and rate the overall data record for FAIRness.*</p> <ul style="list-style-type: none"> Score the dataset and recommend ways to increase the FAIRness of the data and become “DCN approved.” | <p>Findable -</p> <ul style="list-style-type: none"> <input type="checkbox"/> Metadata exceeds author/ title/ date, <input type="checkbox"/> Unique PID (DOI, Handle, PURL, etc.). <input type="checkbox"/> Discoverable via web search engines. <p>Accessible -</p> <ul style="list-style-type: none"> <input type="checkbox"/> Retrievable via a standard protocol (e.g., HTTP). <input type="checkbox"/> Free, open (e.g., download link). <p>Interoperable -</p> <ul style="list-style-type: none"> <input type="checkbox"/> Metadata formatted in a standard schema (e.g., Dublin Core). <input type="checkbox"/> Metadata provided in machine-readable format (OAI feed). <p>Reusable -</p> <ul style="list-style-type: none"> <input type="checkbox"/> Data include sufficient metadata about the data characteristics to reuse <input type="checkbox"/> Contact info displayed if the direct assistance of the author needed. <input type="checkbox"/> Clear indicators of who created, owns, and stewards the data. <input type="checkbox"/> 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: