

Washington University in St. Louis
Washington University Open Scholarship

IASSIST & DCN - Data Curation Workshop

Workshop Schedule

Dec 12th, 10:45 AM - 11:30 AM

Evaluate Presentation

Wendy Kozlowski
Cornell University Library, wak57@cornell.edu

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

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



Part of the [Library and Information Science Commons](#)

Kozlowski, Wendy and Imker, Heidi, "Evaluate Presentation" (2017). *IASSIST & DCN - Data Curation Workshop*. 6.
<https://openscholarship.wustl.edu/data-curation-workshop-2017/schedule/Schedule/6>

This Presentation 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.



E

Evaluate

C ⇒ U ⇒ R ⇒ A ⇒ T ⇒ E



Evaluate for FAIRness

- **Evaluate** and rate the overall data record for FAIRness.*
- Score the dataset and recommend ways to increase the FAIRness of the data.



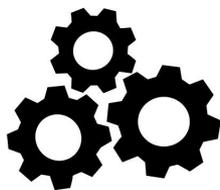
F
indable



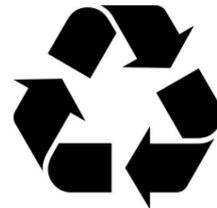
A
ccessible



I
nteroperable



R
eusable





Findable

*To be **findable (F)** or discoverable, data and metadata should be richly described to enable attribute-based search*

- ❑ (meta)data are assigned a globally unique and eternally persistent identifier
- ❑ data are described with rich metadata
- ❑ (meta)data are registered or indexed in a searchable resource
- ❑ metadata specify the data identifier



Accessible

To be broadly **accessible (A)**, data and metadata should be retrievable in a variety of formats that are sensible to humans and machines using persistent identifiers

- ❑ (meta)data are retrievable by their identifier using a standardized communications protocol
- ❑ the protocol is open, free, and universally implementable
- ❑ the protocol allows for an authentication and authorization procedure, where necessary
- ❑ metadata are accessible, even when the data are no longer available



Interoperable

*To be **interoperable (I)**, the description of metadata elements should follow community guidelines that use an open, well defined vocabulary.*

- ❑ (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- ❑ (meta)data use vocabularies that follow FAIR principles
- ❑ (meta)data include qualified references to other (meta)data



Reusable

To be **reusable (R)**, the description of essential, recommended, and optional metadata elements should be machine processable and verifiable, use should be easy and data should be citable to sustain data sharing and recognize the value of data.

- ❑ meta(data) have a plurality of accurate and relevant attributes
- ❑ (meta)data are released with a clear and accessible data usage license
- ❑ (meta)data are associated with their provenance
- ❑ (meta)data meet domain-relevant community standards



Degree of FAIR compliance can vary....





Findable (defined by metadata (PID included) and documentation)

- * No PID nor metadata/documentation
- ** PID without or with insufficient metadata
- *** Sufficient/limited metadata without PID
- **** PID with sufficient metadata
- ***** Extensive metadata and rich additional documentation available



Accessible (defined by presence of user license)

- * Metadata nor data are accessible
- ** Metadata are accessible but data is not accessible
(no clear terms of reuse in license)
- *** User restrictions apply (i.e. privacy, commercial interests,
embargo period)
- **** Public access (after registration)
- ***** Open access unrestricted



Interoperable (defined by data format)

- * Proprietary (privately owned), non-open format data
- ** Proprietary format, accepted by Certified Trustworthy
Data Repository
- *** Non-proprietary, open format = 'preferred format'
- **** As well as in the preferred format, data is standardised using a
standard vocabulary format (for the research field to which
the data pertain)
- ***** Data additionally linked to other data to provide context



Exercise:

As a table:

15 minutes: Review the final dataset, as currently visible [[LINKS](#) on next page]

2 minutes per table: Overall assessment, key things missing, parts that might not be relevant to this dataset

Data from two 2017 J. Mechanisms Robotics Papers about the Steinkamp Hopper

<https://ecommons.cornell.edu/handle/1813/46309>

Data from: Embryogenesis in the plant parasitic nematode *Heterodera glycines* is independent of host-derived stimulation. https://doi.org/10.13012/B2IDB-6946735_V2

Understanding Ecosystem Services Adoption by Natural Resource Managers and Research Ecologists: Survey Data

https://deepblue.lib.umich.edu/data/concern/generic_works/wd375w30z?locale=en

Supporting Data for “Renewable, Degradable, and Chemically Recyclable Cross-Linked Elastomers”

<https://doi.org/10.13020/D6V599>

Four-year-old Children Align their Preferences with those of their Peers DataSet

<https://doi.org/10.7936/K7KP810V>

Data and ArcPython script for Pastoralist Participation (PastPart) Model

<https://doi.org/10.7936/K7V986GW>



Additional references

1. Are the FAIR Data Principles fair?

Alastair Dunning, Madeleine de Smaele, Jasmin Bohmer
(paper)

<https://zenodo.org/record/321423#.WXnq3BPyt0s>

2. Assessing the FAIRness of Datasets in Trustworthy Digital Repositories: A Proposal

Peter Doorn, Ingrid Dillo
(slides)

[https://indico.cern.ch/event/588219/contributions/2384979/attachments/1426152/2188462/Dillo Door n - Assessing FAIRness CERN Geneva 13-03-2017-3.pdf](https://indico.cern.ch/event/588219/contributions/2384979/attachments/1426152/2188462/Dillo_Door_n_-_Assessing_FAIRness_CERN_Geneva_13-03-2017-3.pdf)