THE UNIVERSITY of TENNESSEE KNOXVILLE

MEASURING FAIR PRINCIPLES TO INFORM FITNESS FOR USE

CAROLYN HANK.
(with Wade Bishop)
School of Information Sciences
University of Tennessee

20 February 2018
12th International Digital Curation Conference

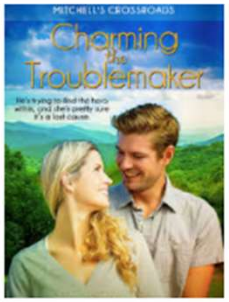
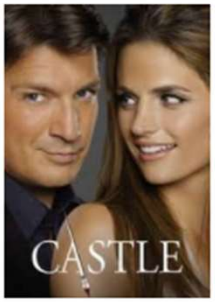
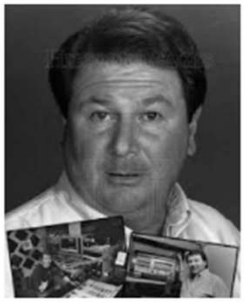



Then Came You
BECKY WADE



WICHITA




Charming the Troublemaker

CASTLE






STEEL
DEAD

Bishop


a little on our collaboration history

focus on producers and data manager/curators

- Data Curation Profiling of Biocollections
  – ASIST 2016; IDCC 2017
- Job Analyses of Earth Sciences Data Managers
  – Preview to iConference 2018 (next month in UK)

now, turning focus on consumers/re-users

To be findable:	F1. (meta)data are assigned a globally unique and eternally persistent identifier. F2. data are described with rich metadata. F3. (meta)data are registered or indexed in a searchable resource. F4. metadata specify the data identifier.
To be accessible:	A1. (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 the protocol is open, free, and universally implementable. A1.2 the protocol allows for an authentication and authorization procedure, where necessary. A2 metadata are accessible, even when the data are no longer available.
To be interoperable:	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles. I3. (meta)data include qualified references to other (meta)data.
To be re-usable:	R1. meta(data) have a plurality of accurate and relevant attributes. R1.1. (meta)data are released with a clear and accessible data usage license. R1.2. (meta)data are associated with their provenance. R1.3. (meta)data meet domain-relevant community standards.

FAIR Data Principles

- FAIR represents a concise, domain-independent, high-level set of data principles that may be applicable in a number of areas and cater to answering the questions both humans and machines will have while discovering and evaluating data prior to use (Wilkinson et al., 2016).
  - Findable
  - Accessible
  - Interoperable
  - Re-usable

Beyond FAI… R? = Fitness for use

- To determine suitability for a particular application or purpose a user must know details about the data, including data quality, scale, interoperability, cost, metadata, syntactic and semantic heterogeneity, and others (Chrisman, 1984; Veregin, 1999).
- "These properties are multifaceted and cover various aspects related to data objects, access services, and data management processes such as the level of annotation, curation, peer review, and citability or machine readability of datasets" (RDA WG site).

Assessing fitness for use

- A framework with the most *vital* facets of fitness for use from a re-user's perspective could outline considerations for the functionality and design of data/metadata, and also the tools used to locate, access, and re-use.
- Operationalizing **re-use** is required and one way this can be done is speaking with actual re-users.
- The purpose of this work is to better understand how re-users discover and evaluate data.
- This perspective likely differs from other aspects that make data "curatable."

Interview Schedule

Job-related demographics

1. What is your current job title?
2. How many years in total have you been working in your current job?
3. How many years in total have you been working with earth science data?
4. Describe your work setting?
5. Please indicate your credentials and degrees.
6. Please provide any other educational or training you have received that is applicable to performing your job.

Interview Schedule (continued)

Think of a recent search for data (or more). The following questions will determine how you discovered and evaluated that data for *fitness for use*.

Findability

7. How did you find the data?
8. Did the data have a persistent identifier (i.e., a long-lasting unique reference to an objects location) (e.g., DOI; PURL)?
9. Did the data have metadata?
10. Did the metadata help you locate the data?

To be findable:	**F1. (meta)data are assigned a globally unique and eternally persistent identifier.** **F2. data are described with rich metadata.** **F3. (meta)data are registered or indexed in a searchable resource.** **F4. metadata specify the data identifier.**

Interview Schedule (continued)

Accessibility

11. How did you access the data?
12. Was the data in an open format (e.g., Public Domain, Attribution License, and so forth)?
13. Was the data free?
14. Did the data have use constraints (e.g., limitations of use)?
15. Was the metadata accessible?

To be accessible	A1. (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 the protocol is open, free, and universally implementable. A1.2 the protocol allows for an authentication and authorization procedure, where necessary. A2 metadata are accessible, even when the data are no longer available.

Interview Schedule (continued)

Interoperability

16. Was the data in a useable format?
17. How was the data encoded?
18. Was it using encoding common to other data used in your research (i.e., same format)?
19. Was the data using shared a controlled vocabularies, or a data dictionaries, and/or other common ontologies?
20. Was the data machine-actionable (e.g., to be processed without humans)?

To be interoperable	**I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.** **I2. (meta)data use vocabularies that follow FAIR principles.** **I3. (meta)data include qualified references to other (meta)data.**

Interview Schedule (continued)

Reusability

20. Were there any issues with the data that impacted reuse of the data (e.g., resolution)?
21. Did the data geographic scale used impact reuse of the data?
22. Did the coordinate systems used impact reuse of the data?
23. Did the metadata provide sufficient information for data reuse?

To be re-usable:	R1. meta(data) have a plurality of accurate and relevant attributes. R1.1. (meta)data are released with a clear and accessible data usage license. R1.2. (meta)data are associated with their provenance. R1.3. (meta)data meet domain-relevant community standards.

Closing

24. Please provide any other feedback about this project or data fitness for use.

This project's potential implications

1) Creating a user-centered methodology to build other data principles beyond FAI;
   - Data can be FAI, but R requires more research;
2) Contributing new knowledge of how scientists access and use science data; and
3) Producing a framework that apprises how best to meet the re-users' needs (i.e., fitness for use).

Next Steps

- Interviews with 22 re-users of data
  - NOAA-Marine, NOAA-Seismic & USGS

Citations

Chrisman, N. R. (1984). Part 2: Issues and problems relating to cartographic data use, exchange and transfer: The role of quality information in the long-term functioning of a geographic information system. *Cartographica: The International Journal for Geographic Information and Geovisualization, 21*(2–3), 79–88. doi: 10.3138/7146-4332-6J78-0671.

Wilkinson, M. D., Verborgh, R., Bonino da Silva Santos, L. O., Clark, Tim, Swertz, Morris A., Kelpin, Fleur D. L., . . . Dumontier, Michel. (2017). Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Computer Science, 3*, e110. doi: 10.7717/peerj-cs.110

Veregin, H. (1999). Data quality parameters. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind. (Eds.), Geographical information systems—Volume I, Principles and technical issues. (2nd ed., pp. 177–189). New York: Wiley.

THANK YOU

Contact

Wade Bishop
bbisho13@utk.edu

Carolyn Hank
chank@utk.edu