Libraries 4 Research Data (L4RD) Interest Group

P13 Meeting: Philadelphia, PA

- Introduction
- "User Stories and Agile Project Management in Developing Repository Software" - Dr. Devan Ray Donaldson, Indiana University
- Project updates
  - Engaging Researchers with Research Data
  - 10 FAIR Things
  - FAIR Lessons in a Carpentry Style
  - Revising 23 Things
  - Support Your Data Update (and other announcements)
- Discussion about future directions of the group

https://rd-alliance.org/

Current Co-chairs: Andi Ogier, Birgit Schmidt, Juliane Schneider, Marta Teperek
Past Co-chairs: Kathleen Shearer, Michael Witt, Wolfram Horstmann
Wiki: https://www.rd-alliance.org/node/1633/all-wiki-index-by-group

- P2 Washington: First meeting as a BOF
- P3 Dublin: Research Data Skills in Libraries
- P4 Amsterdam: Research Data Solutions in Libraries
- P5 San Diego: Organizational Models for Data Services
- P6 Paris: Developing and Adapting to Research Data Policies
- P7 Tokyo: Applying Global Information-sharing and collaboration to Local Practice
- P8 Denver: International Data Week
- P9 Barcelona: Bringing Research Data Management into the Library Mainstream
- P10 Montréal: Realities and Assessment of Library Data Services
- P11 Berlin: From data services to research partnerships: Libraries facilitating knowledge creation
- P12 Botswana: Engaging with Researchers and FAIR Principles
- P13 Philadelphia: Project Updates

http://tinyurl.com/y47pkwh8

Devan Ray Donaldson, Ph.D.

User Stories and Agile Project Management in Developing Repository Software

drdonald@indiana.edu

@hoosierdevan

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Digital Curation Course Overview

Preserving and providing long-term access to digital materials over time is a Grand Challenge. They require constant and ongoing maintenance. This course:

1. *provides an overview of research, policy and current practices in curating and preserving digital data, and*
2. *gives students practical experience working with digital materials, digital curation environments, and digital curation plans.*

Course Goals and Objectives

- Define digital curation.
- Distinguish between different types of curation strategies, explaining their respective merits.
- Explain why file format is an important aspect of digital curation.
- Define preservation metadata and discuss their importance to digital curation.
- Describe important dimensions of data quality.
- Discuss digital repositories and explain why trust is an important factor in digital repositories.
- Develop data management plans for repositories.
- Apply digital forensics methods to data in ways that are relevant to how staff who manage digital resources in collecting institutions apply them.
- Create user stories and use agile project management tools to develop repository software.
- Discuss the importance of data sharing.
- Support domain scientists with management of their data.
- Conduct original digital curation research.

Course Goals and Objectives

- Define digital curation.
- Distinguish between different types of curation strategies, explaining their respective merits.
- Explain why file format is an important aspect of digital curation.
- Define preservation metadata and discuss their importance to digital curation.
- Describe important dimensions of data quality.
- Discuss digital repositories and explain why trust is an important factor in digital repositories.
- Develop data management plans for repositories.
- Apply digital forensics methods to data in ways that are relevant to how staff who manage digital resources in collecting institutions apply them.
- Create user stories and use agile project management tools to develop repository software.
- Discuss the importance of data sharing.
- Support domain scientists with management of their data.
- Conduct original digital curation research.

Course Goals and Objectives

- Define digital curation.
- Distinguish between different types of curation strategies, explaining their respective merits.
- Explain why file format is an important aspect of digital curation.
- Define preservation metadata and discuss their importance to digital curation.
- Describe important dimensions of data quality.
- Discuss digital repositories and explain why trust is an important factor in digital repositories.
- Develop data management plans for repositories.
- Apply digital forensics methods to data in ways that are relevant to how staff who manage digital resources in collecting institutions apply them.
- **Create user stories and use agile project management tools to develop repository software.**
- Discuss the importance of data sharing.
- Support domain scientists with management of their data.
- Conduct original digital curation research.

Course Goals and Objectives

- Define digital curation.
- Distinguish between different types of curation strategies, explaining their respective merits.
- Explain why file format is an important aspect of digital curation.
- Define preservation metadata and discuss their importance to digital curation.
- Describe important dimensions of data quality.
- Discuss digital repositories and explain why trust is an important factor in digital repositories.
- Develop data management plans for repositories.
- Apply digital forensics methods to data in ways that are relevant to how staff who manage digital resources in collecting institutions apply them.
- **Create user stories and use agile project management tools to develop repository software.**
- Discuss the importance of data sharing.
- Support domain scientists with management of their data.
- Conduct original digital curation research.

IU Data Repository Assignment

- Data repositories are large database infrastructures that collect, manage, and store data sets for data analysis, sharing, preservation, and reporting.
- Indiana University Libraries is currently developing a new open-source data repository to address researchers' needs on campus.
- In this assignment, you will visit the test version of the data repository, examine it and answer questions about it (see Syllabus, Pages 3-4)
- IU Data Repository Assignment is due on Wednesday, February 27, 2019 and is worth 10% of your final grade

IU Data Repository Assignment

- Data repositories are large database infrastructures that collect, manage, and store data sets for data analysis, sharing, preservation, and reporting.
- Indiana University Libraries is currently developing a new open-source data repository to address researchers' needs on campus.
- In this assignment, you will visit the test version of the data repository, examine it and answer questions about it (see Syllabus, Pages 3-4)
- IU Data Repository Assignment is due on Wednesday, February 27, 2019 and is worth 10% of your final grade

IU Data Repository Assignment

- You will visit the test version of the data repository (http://carbon.dlib.indiana.edu:8610/data) answer the following questions:
  - What are the features of the data repository?
  - How do I access data?
  - How do I deposit data?
  - How do I cite data?
  - What software platform does the data repository use?
  - Are there file size limits?
  - What metadata can be included with data sets?
  - What are the strengths of the data repository?
  - What are the weaknesses of the data repository?

User Stories Assignment

- Goal: For students to gain experience with the process that librarians use to create repository software, including creating user stories and use of agile project management.
- User stories are informal, natural language descriptions of features. They are written from the perspective of end-users and other stakeholders (clients, managers or developers). User stories have this standard structure. As a 1, I want 2, so that 3:
  - Who are we building it for, who the user is? — As a …
  - What are we building, what is the intention? — I want…
  - Why are we building it, what value it bring for the user? — So that…

User Stories Assignment

- Goal: For students to gain experience with the process that librarians use to create repository software, including creating user stories and use of agile project management.
- User stories are informal, natural language descriptions of features. They are written from the perspective of end-users and other stakeholders (clients, managers or developers). User stories have this standard structure. As a 1, I want 2, so that 3:
  - Who are we building it for, who the user is? — As a …
  - What are we building, what is the intention? — I want…
  - Why are we building it, what value it bring for the user? — So that…

User Stories Assignment

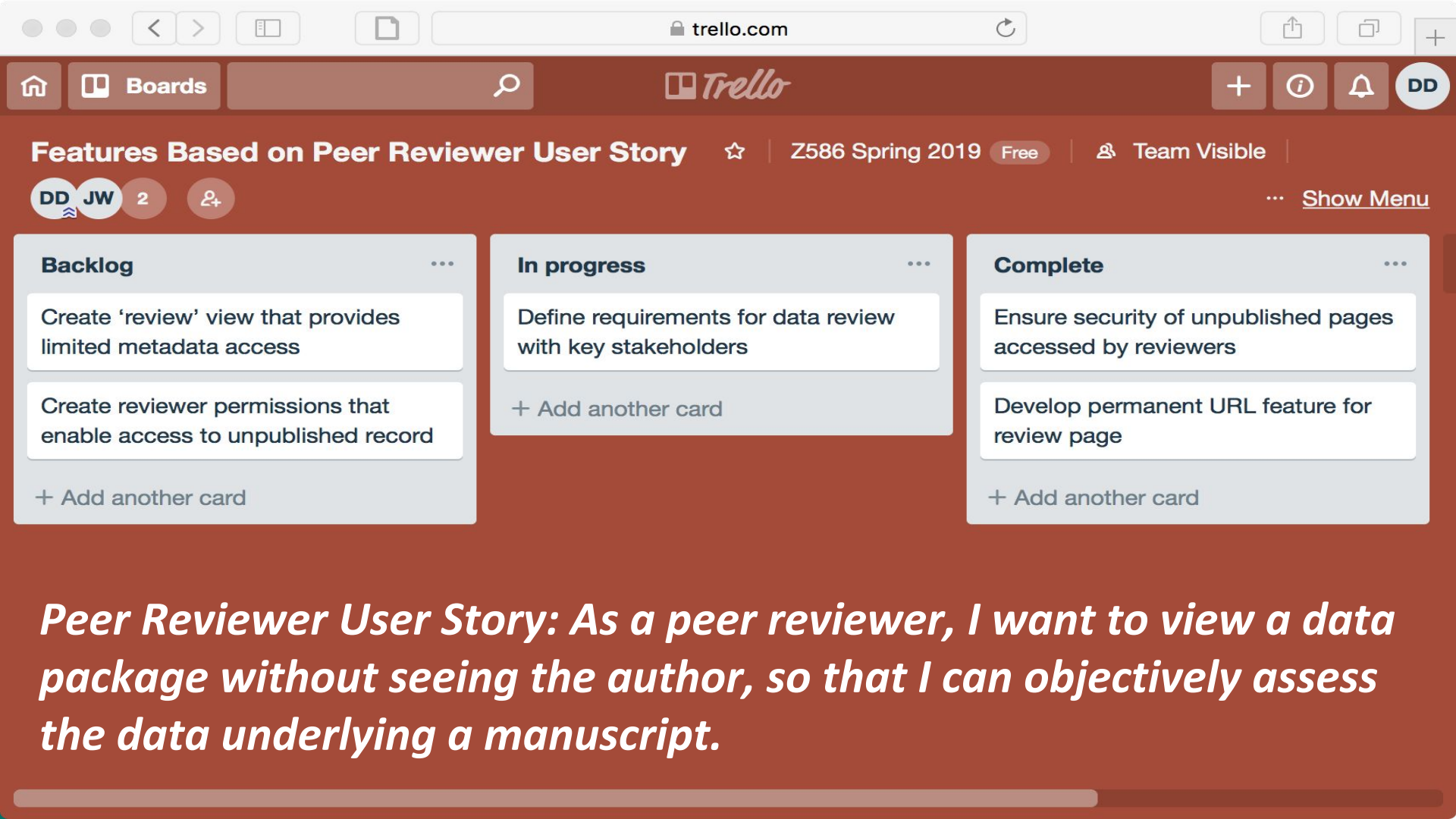
- User Story Example: *As a "client" [e.g., #1], I "want to have easy access to previous metadata versions for a Digital Object Identifier" [e.g., #2], "so that I can make comparisons between versions or investigate potential errors" [e.g., #3].*

User Stories Assignment

- For this assignment, students were broken into groups and were assigned to write a user story from the perspective of one of the following:
  - A data producer
  - A data consumer
  - A data repository manager
  - A grant funder
  - An office of sponsored research employee
  - An Institutional Review Board committee member

User Stories Assignment

- For this assignment, the students:
  - used information from their earlier examination of the test repository,
  - used any information that they learned in class based on course readings and discussions, and
  - emailed and/or met with people at IU and elsewhere who fit these user groups, explaining the nature of this project and asking them about their data needs and data repository needs.



Peer Reviewer User Story: As a peer reviewer, I want to view a data package without seeing the author, so that I can objectively assess the data underlying a manuscript.

User Stories Assignment

- Students delivered a 10 to 15-minute PowerPoint presentation describing your work on this assignment
- The presentation included the following:
  - A title slide listing all group members
  - A slide with their user story
  - A slide explaining why they chose the user story and a description of the sources they used to develop it
  - A slide explaining how they turned their user story into a list of features
  - A slide with a screenshot of their features listed in Trello, including an explanation of why they sorted each feature in the lane they chose.

User Stories

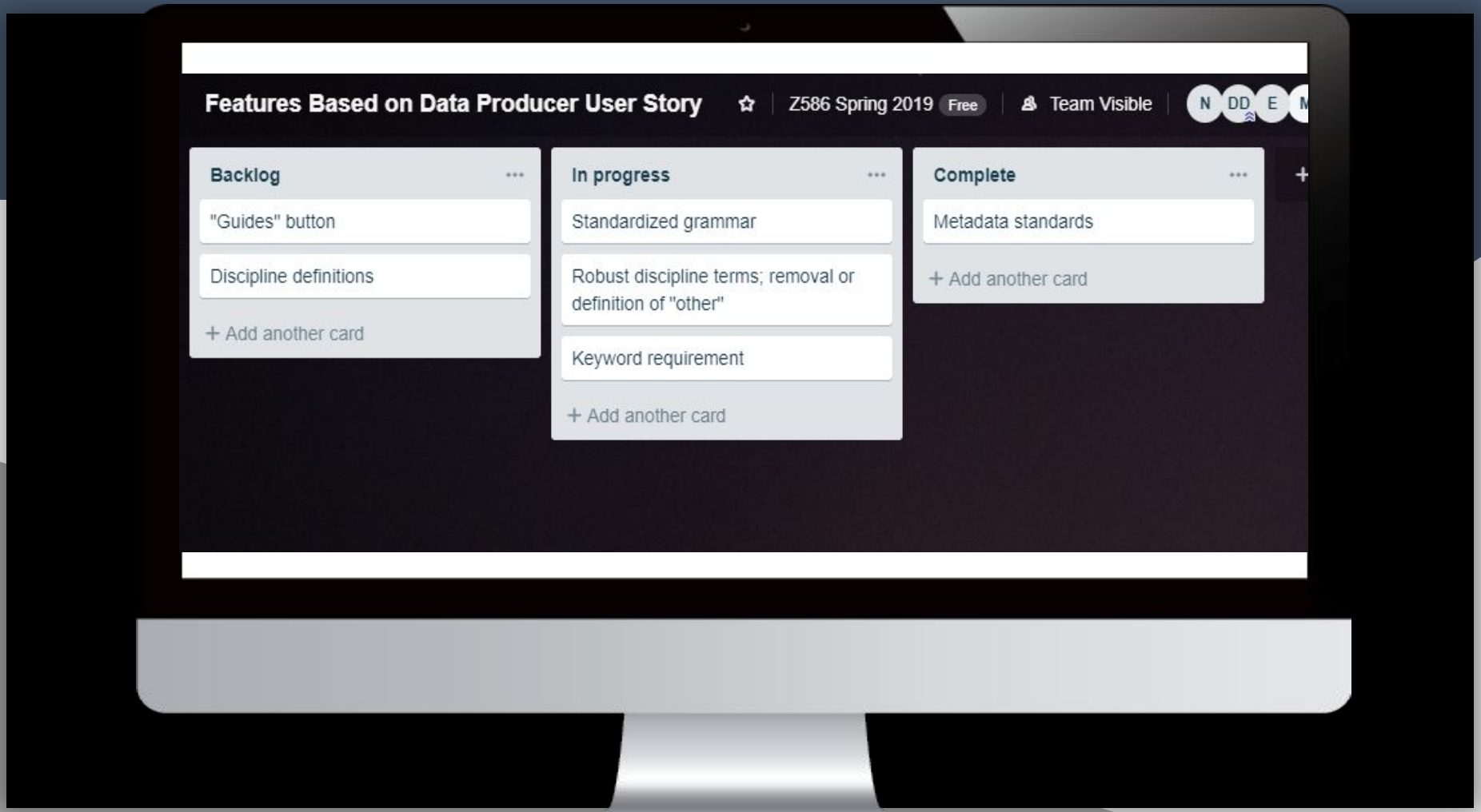
- A data producer: “As **data producers**, we want to build **structured metadata** to offer researchers **easy access to data for their projects**.”
- A data consumer: “As a **data consumer**, I want to be able to **sort by type of Creative Commons license applied to datasets** so that **I can find data to be used for my own projects.**”
- A data repository manager: “As **digital repository managers**, we want to **implement site analytics** so that **we ensure the trustworthiness, functionality, and efficiency of all parts of Chimera-backed data**.”

User Stories

- A grant funder: “As **a grant funder** I want **to be able to easily access metadata based on funding agency** so that **researcher compliance with grant requirements can be evaluated**.”
- An office of sponsored research employee: “As **Associate Director of the Office of Research Administration, I want federally funded researchers to use the Chimera Data Repository as an organizational tool and storage resource** so that **they meet data sharing compliance for NSF grant benefactors and are able to prepare NIH funded research for PubMed Central publication**.”

User Stories

- An Institutional Review Board committee member: "**As an** IRB committee member, **I want** to ensure research data is securely transmitted, stored, and shared **so that** research subjects' privacy and confidentiality is protected."


Features Based on Data Producer User Story
Z586 Spring 2019
Free
Team Visible
N
DD
E
Backlog
"Guides" button
Discipline definitions
+ Add another card
In progress
Standardized grammar
Robust discipline terms; removal or definition of "other"
Keyword requirement
+ Add another card
Complete
Metadata standards
+ Add another card

Features Based on Data Consumer User Story

Z586 Spring 2019 Free | Team Visible | M DD

Backlog

- Create search filters that aggregate CC licenses of a specific type
- Add section for additional licenses that may be applied to data

+ Add another card

In progress

- Develop robust set of search tools that meet the needs of our users

+ Add another card

Complete

- Apply CC license to every dataset uploaded to repository
- List available licenses on data upload page for users

+ Add another card

Backlog

- Create a tombstone record system for removed documents
- Increase file format acceptability
- Provide remote or tape backups for disaster recovery services
- Establish automated processes through scripting
- Run analytics on access points for the repository to track how individuals find and use data

+ Add another card

In progress

- Running quality control checks such as checksum creation and fixity checks, and creation of archival backups
- Completing necessary software updates
- Migrating files to other stable formats (Normalization for some file formats)
- Maintaining awareness of changes to storage medium and ability to make updates as necessary
- Facilitating the creation of DOIs

+ Add another card

Complete

- Provided a way of collecting metadata
- Established user browse and search to make data findable and usable
- Established technical services liaison to assist users with user interface technical issues and questions
- Chose Samvera, which includes Google Analytics, as a repository framework
- Implemented individual activity tracking

+ Add another card

Features Based on Grant Funder User Story

Z586 Spring 2019 Free | Team Visible | AR DD M

Backlog

- Searchable by funding agency
- Features that encourage sharability (record or data)
- long term preservation or repository succession plan

+ Add another card

In progress

- Publicly accessible data
- Grant number metadata
- Transparent access costs

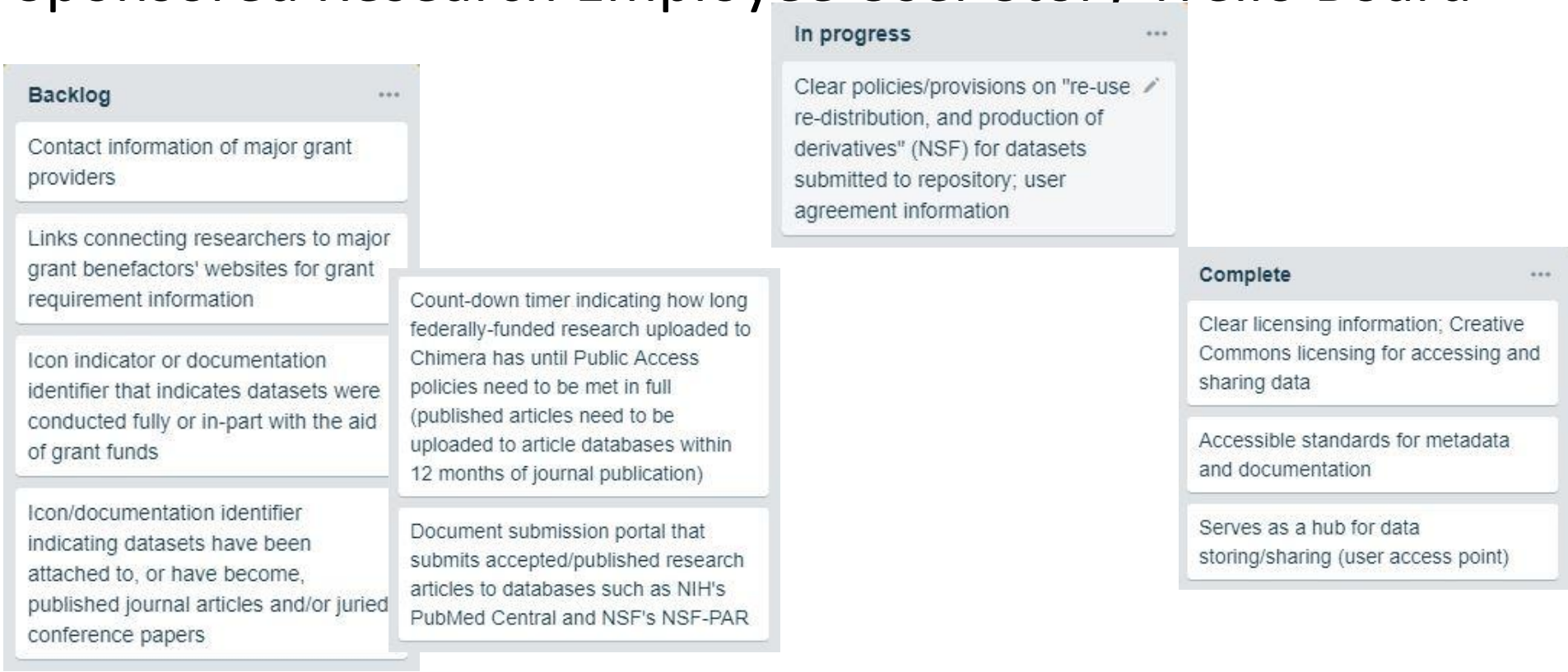
+ Add another card

Complete

- Funding agency metadata
- All relevant data sets, primary, physical, and supporting information, in the same place.
- Copyright/use information
- Protection of confidential data

+ Add another card

Sponsored Research Employee User Story Trello Board



Features Based on IRB User Story

Z586 Spring 2019 Free | Team Visible | M DD K 3 | Invite

Backlog

- Develop a stewardship plan for sensitive research data (does NOT include: sensitive identifiable human subject research data)
- Limited review level that allows limited metadata access for sensitive research data
- Include in required metadata an indication of sensitive research data. Depositor selects between: No sensitive research data/Coded data/De-Identified Data/Anonymous Data
- Develop data auditing process to periodically ensure no sensitive identifiable human subject research data is stored in the repository

+ Add another card

In progress

- Clear explanations of different kinds of sensitive research data in the "Conditions for Deposit" Statement and "Work That Cannot Be Accepted" Statement
- Vetting process for depositors with sensitive research data

+ Add another card

Complete

- Establish a "Mint a DOI" process, where data cannot be altered after the DOI is minted
- Integrate checksums to detect any changes in stored data
- Conditions for Deposit Statement
- Work That Cannot Be Accepted Statement
- Link out to university resources on sensitive data and IRB compliance
- IRB members can search repository by study name and principle investigator name

+ Add another card

+ Add another list

Conclusion

- The **IU Data Repository Assignment** and the **User Stories Assignment** helped the students learn about data repositories with **hands-on experience**
- The User Stories Assignment was **mutually-beneficial to Librarians and LIS students**
- This project underscores the potential **benefit of Library-LIS collaboration**
- **Collaborate with LIS programs at your institution or nearby OR collaborate with ME :0)**

Devan Ray Donaldson, Ph.D

User Stories and Agile Project Management in Developing Repository Software

drdonald@indiana.edu

@hoosierdevan

INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Engaging Researchers with Research Data

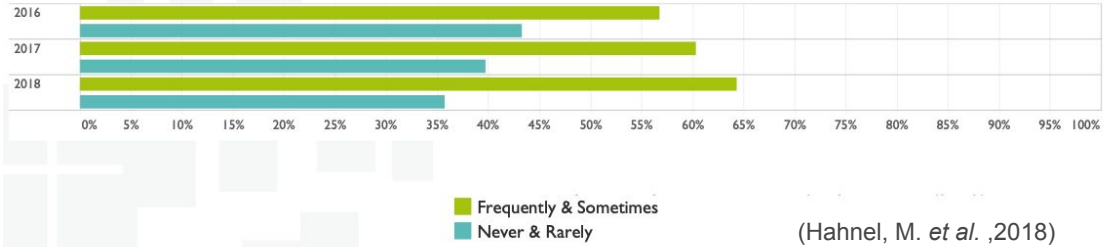
Helene N. Andreassen[1], Raman Ganguly[2] & Andrea Medina-Smith[3]
[1]UiT The Arctic University of Norway, [2]University of Vienna, [3]National Institute of Standards and Technology

Libraries for Research Data IG
RDA 13th Plenary, Philadelphia, April 2-4, 2019

Background

Gap between perceived benefit of data sharing and actual practice

Fig 1. How often researchers have made their data openly available



(Hahnel, M. *et al.* ,2018)

It is the job of RDM service providers to help shrink this gap via engagement with researchers.

Goals

Help institutions increase engagement with RDM among researchers

- Gather RDM activities from research institutions
- Categorise and synthesise activities
- Publish activities in order to allow evaluation and reuse by interested parties
- Format of publication: textbook and web-based portal

Project group and activity

- Project leader: Marta Teperek (TU Delft)
- Steering board to advise on the project
- 30 + project members from 3 continents
- January-February 2019: Survey invitation sent out
- March 2019: Data analysis discussions

Data + Analysis

- Contacted 60 funders, 80 scientific institutions, and used 28 mailing lists
- 216 responses, 88 described their activities
- 50 unique scientific institutions
- Team is categorizing the responses by activity and purpose
- Post categorization, will request more information via email

Website

- Presenting information about the project, team, data and data analysis
- The data will be distributed over a repository and linked to the website
- Find a place on RDA for the website
- Data will be in a git repository
- The website is under construction and will come soon

Next steps

- Split data analysis into qualitative (case studies) and quantitative branches
- Full report hopefully to be presented in Helsinki @ RDA 14th plenary
- Dissemination: Results will be disseminated via a textbook, and a searchable website
- RDA: Join the BoF Research Data Management in Institutions initiative?

Survey Still Open!

- Collect more case studies to increase database
- One activity per survey entry to facilitate categorisation

https://www.1ka.si/a/193487

RDA Researcher Engagement Project

An international & multi-institutional project

Steering Board

Lauren Cadwallader, Julien Colomb, Maria Cruz, Mary Donaldson, Lambert Heller, Rosie Higman, Elli Papadopoulou, Vanessa Proudman, James Savage, Marta Teperek

Project group members

Helene N. Andreassen, Daniel Bangert, Miriam Braskova, Lauren Cadwallader, John Chodacki, Julien Colomb, Philipp Conzett, Maria Cruz, Mary Donaldson, Biswanath Dutta, Esther Fernandez, Joshua Finnell, Raman Ganguly, Patricia Henning, Amy Hodge, Stein Høydalsvik, Greg Janée, Lynda Kellam, Gabor Kismihok, Iryna Kuchma, Narendra Kumar Bhoi, Young-Joo Lee, Leif Longva, Andrea Medina-Smith, Solomon Mekonnen, Remedios Melero, Rising Osazuwa, Elli Papadopoulou, Fernanda Peset, Josiline Phiri, Piyachat Ratana, Gerry Ryder, James Savage, Souleymane Sogoba, Magdalena Szuflita-Żurawska, Ralf Toepfer, Ellen Verbakel, Irena Vipavc Brvar, Jacquelynne Waldron, Anna Wałek, Yan Wang, Iza Witkowska, Joanne Yeomans, Grzegorz Bulczak, Szymon Andrzejewski

References

Cruz, Maria & Colomb, Julien (2019). “Researcher Engagement With Data Management – What Works?” Retrieved from https://www.rd-alliance.org/blogs/researcher-engagement-data-management

Science D, Hahnel M, Fane B, Treadway J, Baynes G, Wilkinson R, et al.. The State of Open Data Report 2018. *figshare*; 2018. https://doi.org/10.6084/m9.figshare.7195058.v2

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., & Astell, M. (2018). *Practical challenges for researchers in data sharing*. https://doi.org/10.6084/m9.figshare.5975011

Wouters, P., & Haak, W. (2017). *Open data: The researcher perspective*. Retrieved from https://www.elsevier.com/about/open-science/research-data/open-data-report

Ready, Set, Go!

Join the Top 10 FAIR Data Things Global Sprint!

Webinar - 20 November 2018
Sprint - 29-30 November 2018

Top 10 FAIR Data Global Sprint 29-30 November 2018

Organised by:

Library Carpentry, **Australian Research Data Commons** and the Research Data Alliance **Libraries for Research Data Interest Group**

In collaboration with

FOSTER Open Science, OpenAire, RDA Europe, Data Management Training Clearinghouse, California Digital Library, Dryad, AARNet, DANS, and Centre for Digital Scholarship at Leiden University Library.

See: https://librarycarpentry.org/blog/2018/10/top-ten-fair-announcement/

Global sprint - what and why?

What is the purpose of the Sprint?

To create a wide range of *Top 10 FAIR Data Things* by research disciplines and/or themes.

What is a *Top 10 FAIR Data Things* resource?

"Things" is a neat concept for creating packaged content on any topic. Each “Thing” is a self-directed learning activity for anybody who wants to know more about FAIR research data. The *Top 10 FAIR Data Things* resources we create during the Sprint can be used by the research community to understand FAIR in different discipline and theme contexts as well as providing some initial steps to consider.

Example

Medical & health Thing 2: Issues in research data management

SHARE

Research data is critical to solving the big questions of our time. So what are some of the issues we face in managing research data?

Activity 1 Considerations in data management

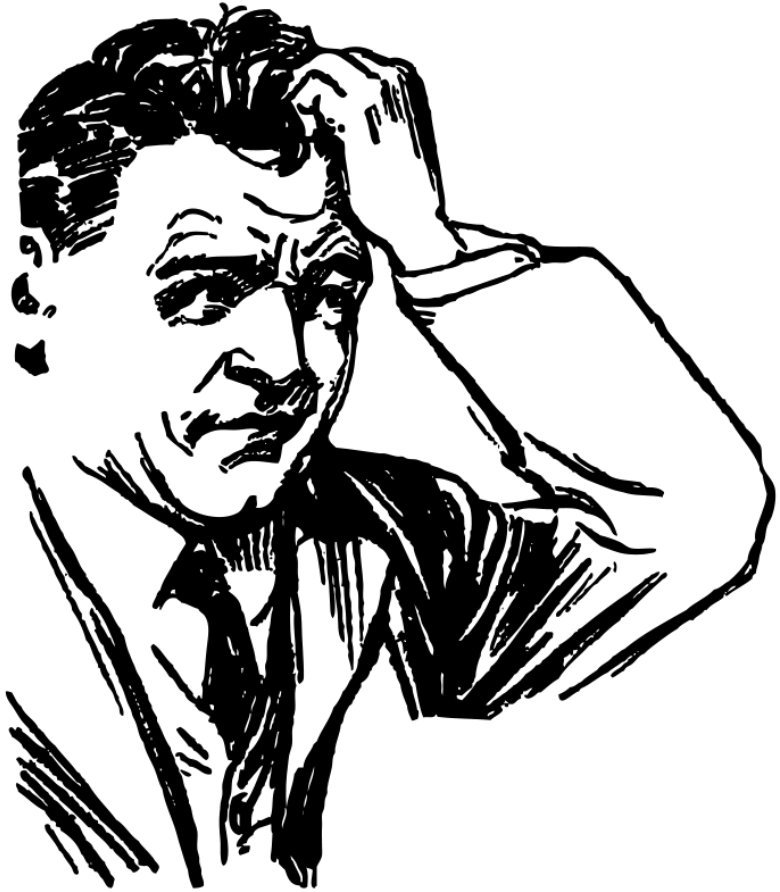
Research data is for everyone. Governments and Universities all around Australia and the world are now encouraging researchers to better manage their data so others can use it.

Research data might be critical to solving the big questions of our time, but so much data are being lost or poorly managed.

1. Take just a minute and browse over some ways Queensland Government Data ☐ is being used by businesses, families, travellers, farmers.
2. This 4.40mins cartoon ☐ put together by the New York University Health Sciences Library, is about what happens when a researcher hasn't managed their data (at all). What could possibly go wrong?!
3. As you watch the cartoon jot down the data management mistakes which interest or appall you.
4. Now, scan through the dot points in the 'Consider the following' section of the University of the Sunshine Coast's LibGuide ☐ which provides advice for researchers on how to manage their data.

Consider how just ONE of the data disasters depicted in the cartoon could have been avoided.

https://www.ands.org.au/working-with-data/skills/23-research-data-things/10-medical-and-health-things



https://commons.wikimedia.org/wiki/File:Headscratcher.png

… but what does it mean in a discipline?

Find Out! https://librarycarpentry.org/Top-10-FAIR/

Oceanography

Research Software

Research Libraries

Research Data Management Support

International Relations

Humanities: Historical Research

Geoscience

Biomedical Data Producers, Stewards, and Funders

Biodiversity

Australian Government Data/Collections

Archaeology

Next steps

Library Carpentry Global Sprint!!!

May 30-31

https://librarycarpentry.org/blog/2019/03/lc-mozilla-global-sprint/

Develop the Library Carpentry materials which includes the 10 FAIR Things outputs

Additional Updates - John Chodacki (California Digital Library - CDL)

Library Carpentry Instructor Training

an open instructor training session is happening in Portland, Oregon

May 5-6 – Travel Support is available

https://forms.gle/98pRUdaxC9RLk7ts8

Or tinyurl.com/libcarpportland

UC3

Additional Updates - John Chodacki (California Digital Library - CDL)

FSCI - Force11 Scholarly Communications Institute

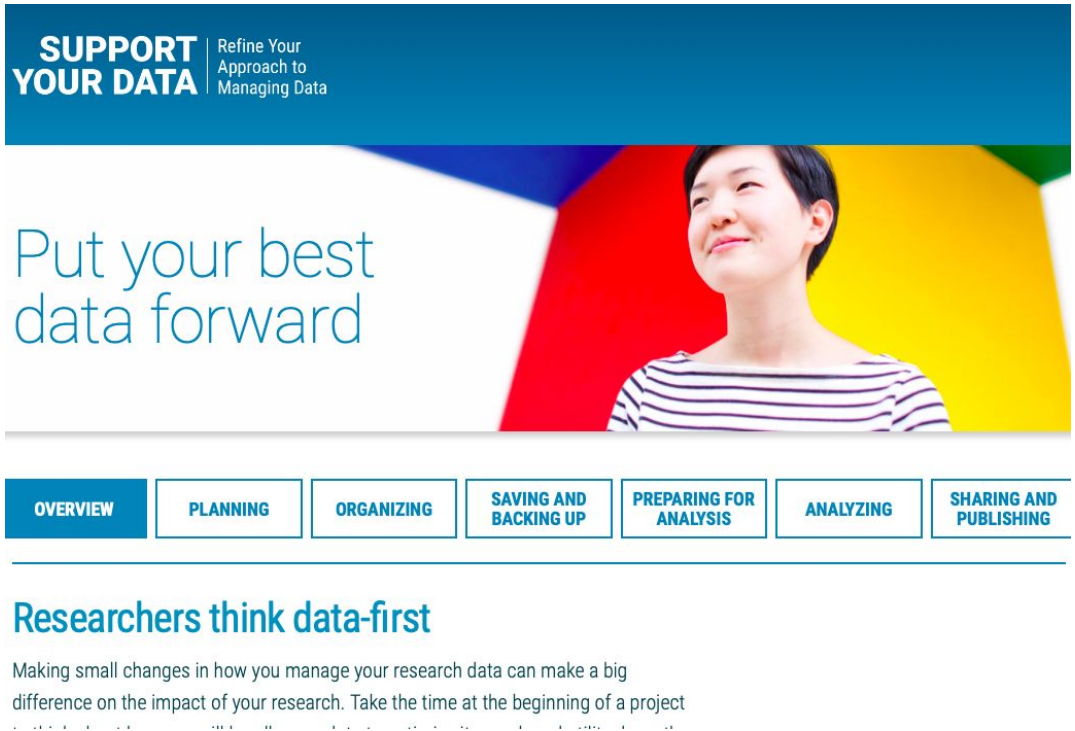
www.force11.org/fsci/2019

UC3

Additional Updates - John Chodacki (California Digital Library - CDL)

Support Your Data: messaging written by researchers, for researchers, aimed at helping researchers do better RDM (with input by this L4RD RDA Groups)

Website launching in May 2019 at researchdata.org (not live yet)



UC3

Additional Updates - John Chodacki (California Digital Library - CDL)

Wikicite (ad hoc presentation)

You can see info here:
http://wikicite.org

You can play with Scholia here:
https://tools.wmflabs.org/scholia/

UC3