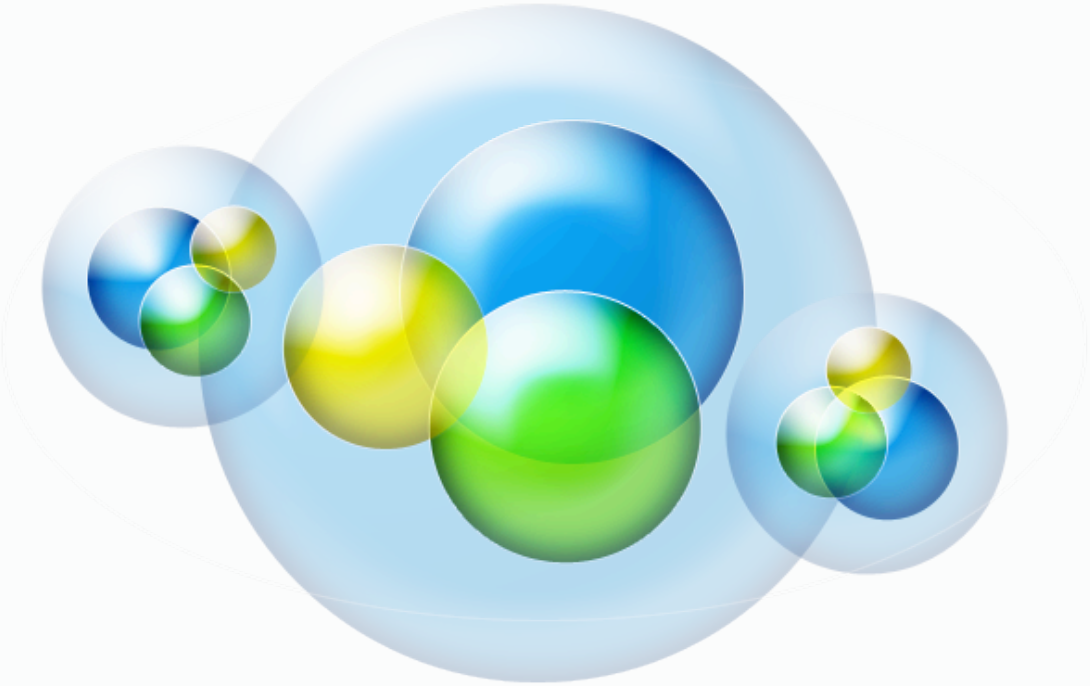


# Top 10 FAIR Data & Software Things

**Let's get practical!**

# Juliane Schneider


Juliane_Schneider@hms.harvard.edu
Team Lead, www.eagle-i.net
Twitter: @JulianeS
ORCID: 0000-0002-7664-3331

# Organizers




February 1, 2019

Lesson Open Access

# Top 10 FAIR Data & Software Things

Christopher Erdmann; Natasha Simons; Reid Otsuji; Stephanie Labou; Ryan Johnson; Guilherme Castelao; Bia Villas Boas; Anna-Lena Lamprecht; Carlos Martinez Ortiz; Leyla Garcia; Mateusz Kuzak; Paula Andrea Martinez; Liz Stokes; Tom Honeyman; Sharyn Wise; Josh Quan; Scott Peterson; Amy Neeser; Lena Karvovskaya; Otto Lange; Iza Witkowska; Jacques Flores; Fiona Bradley; Kristina Hettne; Peter Verhaar; Ben Companjen; Laurents Sesink; Fieke Schoots; Erik Schultes; Rajaram Kaliyaperumal; Erzsébet Tóth-Czifra; Ricardo de Miranda Azevedo; Sanne Muurling; John Brown; Janice Chan; Niamh Quigley; Lisa Federer; Douglas Joubert; Allissa Dillman; Kenneth Wilkins; Ishwar Chandramouliswaran; Vivek Navale; Susan Wright; Silvia Di Giorgio; Mandela Fasemore; Konrad Förstner; Till Sauerwein; Eva Seidlmayer; Ilja Zeitlin; Susannah Bacon; Katie Hannan; Richard Ferrers; Keith Russell; Deidre Whitmore; Tim Dennis

# 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.**

# Disciplines Covered:

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**
**Music**

# Basic Word Analysis

| Metadata | Citations/ PIDs | Licensing | Vocabulatires | Standards /Formats |
|---|---|---|---|---|
| 6** | 11 | 9 | 6 | 5 |

**The word ‘metadata’ was used

Others:

Privacy

Preservation

Containers

Funder requirements

APIs and Applications

Linked Data

# Identifiers/PIDs by Letter

- F: 7
- A: 2
- I: 2
- R: 0

# FAIR Letters vs. Concepts

| By Letter | Concepts |
|---|---|
| Oceanography | Research Data Management Support |
| Research Software | Research Libraries |
| Humanities: Historical Research | International Relations |
| Geoscience | Biomedical Data Producers, Stewards and Funders |
| Biodiversity | Archaeology |
| Australian Government Data/ Collections | Music |

# Oceanography

- **Findable**
  - **Thing 1: Data repositories**
  - **Thing 2: Metadata**
  - **Thing 3: Permanent Identifiers**
  - **Thing 4: Citations**
- **Accessible**
  - **Thing 5: Data formats**
  - **Thing 6: Data Organization and Management**
  - **Thing 7: Re-usable data**

# Oceanography

- **Interoperable**
  - **Thing 3: Permanent Identifiers**
  - **Thing 6: Data Organization and Management**
  - **Thing 2: Metadata**
  - **Thing 10: APIs and Apps**
- **Reusable**
  - **Thing 8: Tools of the trade**
  - **Thing 9: Reproducibility**
  - **Thing 10: APIs and Apps**

# Thing 6: Data Organization & Management

## Activity 1:

## Considerations for basic data organization and management

## Group Discussion 1:

- Is your data file structure something that a new lab member could easily learn, or are datasets organized in a more haphazard fashion?
- Do you have any documentation associated describing how to navigate your data structures?

## Group Discussion 2:

- Talk about where/how you are currently storing data you are working with. Would another lab member be able to access all your data if needed?

# Thing 6: Data Organization & Management

## Activity 2:

## Identifying vulnerabilities

- **Scenario 1:** Your entire office/lab building burns down overnight. No one is harmed, because no one was there, but all electronics in the building perish beyond hope of repair. The next morning, can you access any of your data?
- **Scenario 2:** The cloud server you use (everything from Google Drive to GitHub) crashes. Can you still access your most up to date data?

## Discussion 1:

- From either of the two scenarios, can your data survive a disaster? What are some of the things that you think you are doing incorrectly to prevent data loss?

# Research Libraries

- **Thing 1: Why should librarians care about FAIR?**
- **Thing 2: How FAIR are your data?**
- **Thing 3: Do you teach FAIR to your researchers?**
- **Thing 4: Is FAIR built into library practice and policy?**
- **Thing 5: Are your library staff trained in FAIR?**

# Research Libraries

- **Thing 6: Are digital libraries FAIR?**
- **Thing 7: Does your library support FAIR metadata**
- **Thing 8: Does your library support FAIR identifiers**
- **Thing 9: Does your library support FAIR protocols**
- **Thing 10: Next steps for your library in supporting FAIR**

# Thing 2: How FAIR are your data?

The FAIR Principles are easily understood in theory but more challenging when applied in practice. In this exercise, you will be using the Australian Research Data Commons (ARDC) Data self-assessment tool to assess the 'FAIRness' of one of your library's datasets.

**Activity:**

1. Select a metadata record from your library's collection (e.g. your institutional repository) that describes a published dataset. If your library or institution doesn't have a repository for research data, choose one from re3data.org
2. Open the ARDC FAIR Data Assessment tool and run your chosen dataset against the tool to assess its 'FAIRness'.

**Consider:**

- How FAIR was your chosen dataset?
- How easy was it to apply the FAIR criteria to your dataset?
- What things need to happen in order to improve the 'FAIRness' of your chosen dataset?

**Want more?**

Try your hand at other tools like the CSIRO 5 star data rating tool and the DANS FAIR data assessment tool.

OpenRefine

# ANDS FAIR Assessment Tool

## FAIR self-assessment tool

Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

This self-assessment tool has been designed predominantly for data librarians and IT staff, but could be used by software engineers developing FAIR Data tools and services, and researchers provided they have assistance from research support staff.

You will be asked questions related to the principles underpinning Findable, Accessible, Interoperable and Reusable. Once you have answered all the questions in each section you will be given a ‘green bar’ indicator based on your answers in that section, and when all sections are completed, an overall 'FAIRness' indicator is provided.

Please be aware that additional explanatory information is provided within the tool. The (i) information button provides an overview of each of the FAIR high-level elements (Findable, Accessible, Interoperable and Reusable). Additionally, each question is hyperlinked, leading users to explanatory information and links to wider resources on related topics.

### Findable

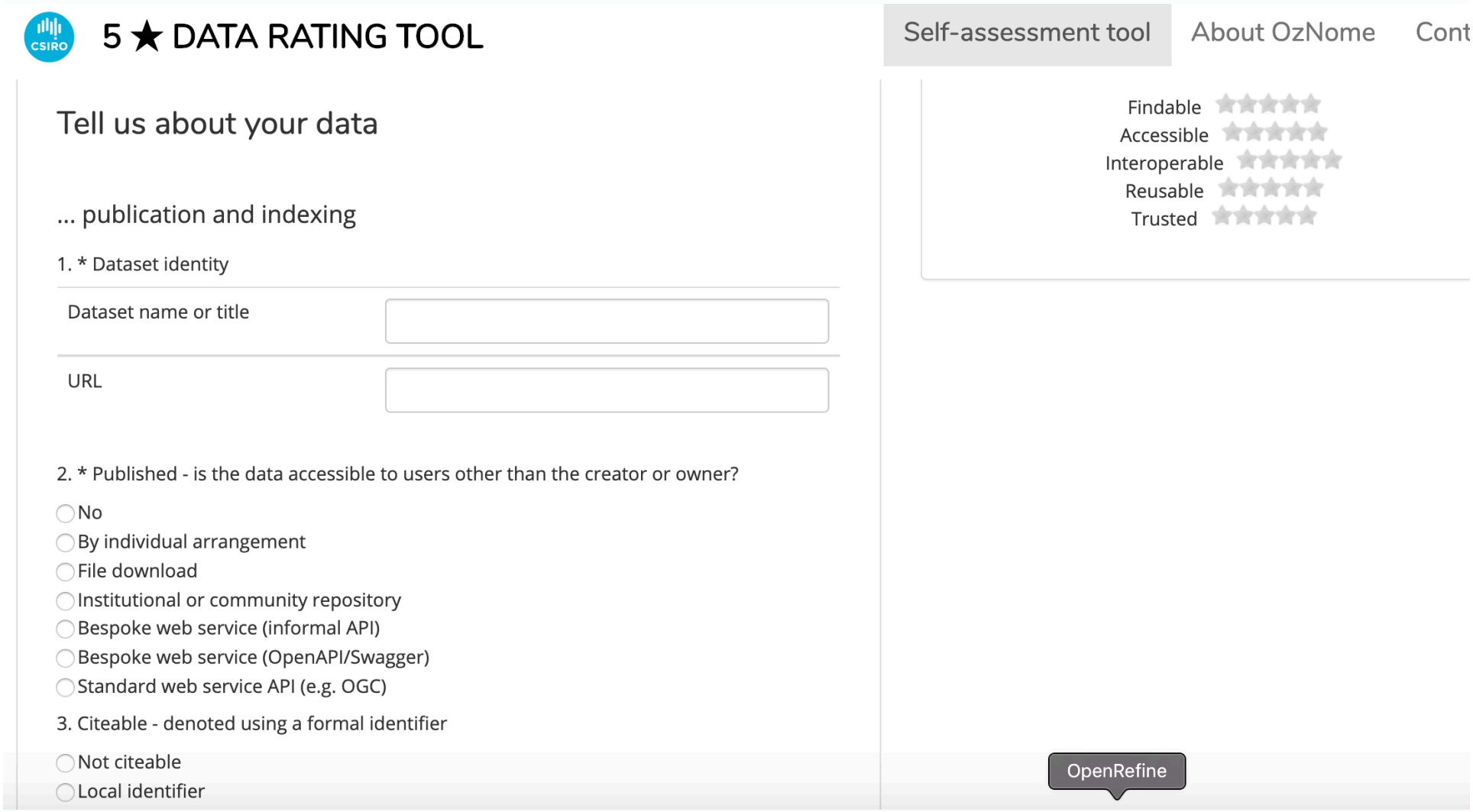
**Does the dataset have any identifiers assigned?** — No identifier

**Is the dataset identifier included in all metadata records/files describing the data?** — No

**How is the data described with metadata?** — The data is not described

**What type of repository or registry is the metadata record in?** — The data is not described in any repository

# CSIRO 5-Star Data Rating



# Thing 8: Does your library support FAIR identifiers?

The FAIR data principles call for open, standardised protocols for accessing data via a persistent identifier. Persistent identifiers are crucial for the findability and identification of research, researchers and for tracking impact metrics. So how well does your library support persistent identifiers?

**Activity:**

1. Find out how well your library supports ORCIDs and DOIs:

- Do your library systems support the identification of researchers via an ORCID? Do you authenticate against the ORCID registry? Do you have an ORCID?
- Do your library systems, such as your institutional repository, support the issuing of Digital Object Identifiers (DOIs) for research data and related materials?

**Consider:**

- What other types of persistent identifiers do you think your library should support? Why or why not?

**Want more?**

If your library supports the minting of DOIs for research data and related materials, is there more that you could do in this regard? Check out A Data Citation Roadmap for Scholarly Repositories and determine how much of the roadmap you can check off your list and how much is yet to do.

# Pre-FAIR Considerations

- **Which communities do your research data belong to?**
- **What are the standards for that community?**
- **Are there overlaps?**
- **Are there geographical considerations? (HIPAA privacy rules in the United States)**