DataCite

CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

FAIR WORKFLOWS

Towards metadata completeness

FAIR Workflows

Xiaoli Chen

9. 11. 2022

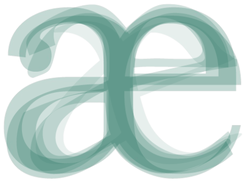
DataCite Open Hour

Agenda

- Metadata considerations in the context of the project
- Capturing metadata
- Enriching metadata

TEMPLETON WORLD
CHARITY FOUNDATION

DataCite
CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE



Max-Planck-Institut
für empirische Ästhetik

University of California
CDL
California Digital Library

ARDC
Australian Research Data Commons

DRYAD

ChronosHub

CEDAR

Crossref

CENTER FOR
OPEN SCIENCE

What makes a workflow FAIR

FAIR Entities

- Uniquely identified resources associated to a project
  - Researcher (ORCID iD)
  - Research organization (ROR ID)
  - Funding agency (ROR ID) and grant (Grant-ID)

FAIR Practices

- Sharing various types of interim outputs
  - Data Management Plan, Pre-registration, Protocol, Preprint, Code, Dataset, etc.

FAIR Supporting Structures

- Tools and platforms that integrate PIDs and metadata workflow
  - DMP authoring tools, Metadata templates, Data repositories, Notebooks, Collaborative research platforms, etc.

FAIR Outputs

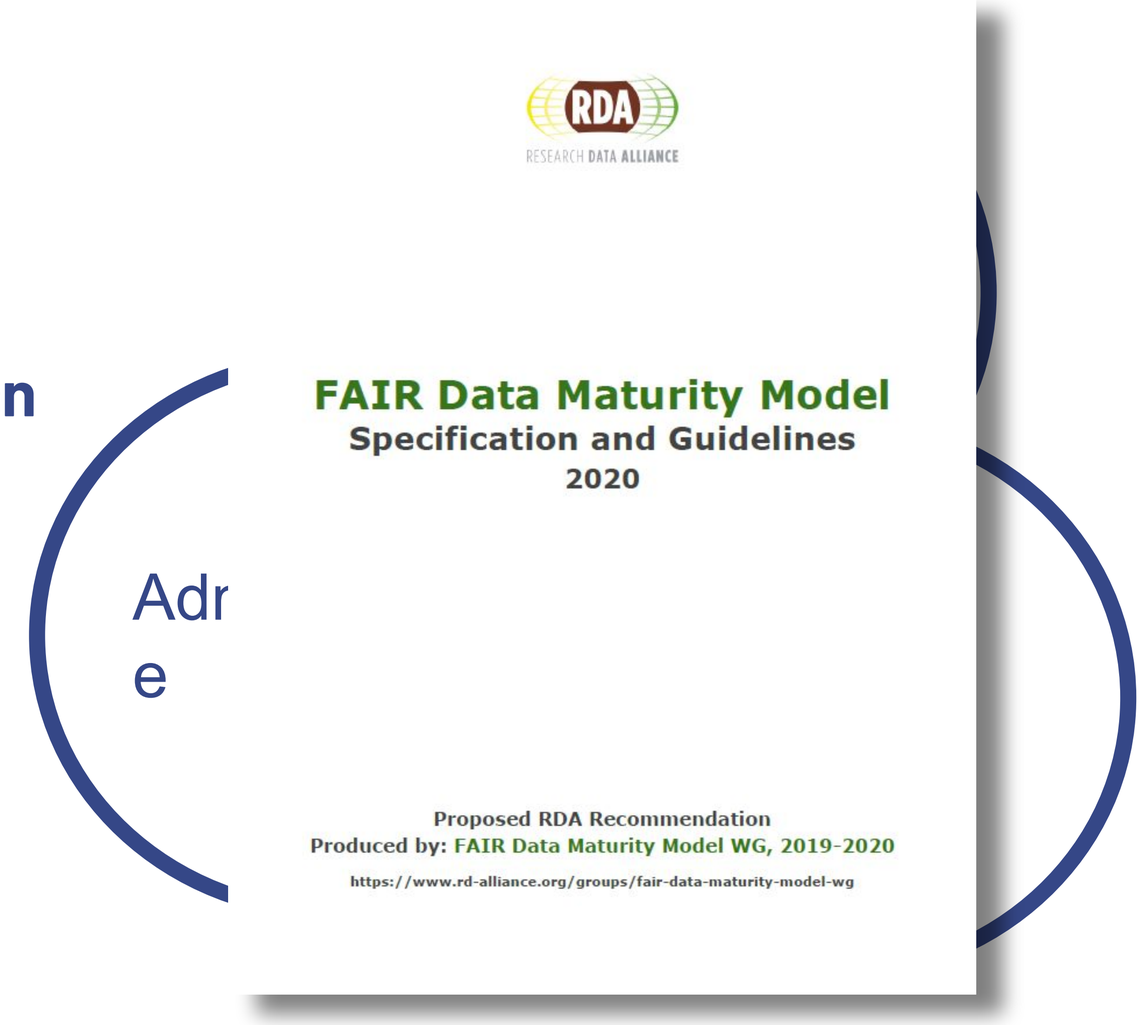
- Assigning PIDs to outputs with rich metadata annotation
  - Essential descriptive and connection metadata
    - Connection between inputs and outputs
    - Relations between outputs
  - Domain specific metadata
    - Disciplinary ontological information
    - Experimental setup

Identifier	Name
FsF-F1-01D	**Data** is assigned a globally unique identifier.
FsF-F1-02D	**Data** is assigned a persistent identifier.
FsF-F2-01M	**Metadata** includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.
FsF-F3-01M	**Metadata** includes the identifier of the data it describes.
FsF-F4-01M	**Metadata** is offered in such a way that it can be retrieved by machines.
FsF-A1-01M	**Metadata** contains access level and access conditions of the data.
FsF-A1-02M	**Metadata** is accessible through a standardized communication protocol
FsF-A1-03D	**Data** is accessible through a standardized communication protocol
FsF-A2-01M	**Metadata** remains available, even if the data is no longer available.
FsF-I1-01M	**Metadata** is represented using a formal knowledge representation language.
FsF-I1-02M	**Metadata** uses semantic resources.
FsF-I3-01M	**Metadata** includes links between the data and its related entities.
FsF-R1-01MD	**Metadata** specifies the content of the data.
FsF-R1.1-01M	**Metadata** includes license information under which data can be reused.
FsF-R1.2-01M	**Metadata** includes provenance information about data creation or generation.
FsF-R1.3-01M	**Metadata** follows a standard recommended by the target research community of the data.
FsF-R1.3-02D	**Data** is available in a file format recommended by the target research community.

Devaraju, Anusuriya, Huber, Robert, Mokrane, Mustapha, Cepinskas, Linas, Davidson, Joy, Herterich, Patricia, L'Hours, Herve, de Vries, Jerry, & White, Angus. (2020). FAIRsFAIR Data Object Assessment Metrics (0.3). Zenodo. https://doi.org/10.5281/zenodo.3934401

What does complete mean?

- **Completeness is one aspect of quality**
  - Accuracy
  - Provenance/authority
  - Accessibility
  - Timeliness
  - …
- **Completeness is only meaningful when measured against a standard**
  - Community recommendation
  - Use case requirements
  - Local context



FAIR Workflows Project

Potential workflow

4. Research

Research conducts the research study

5. Protocols and research methods

Researcher shares and deposits

3. Submit Data Management Plan

Researcher submits DMP

esearch data, instruments deposited

rcher deposits data and
ments

2. Application Approved

Grant approved and grant ID assigned

esearch article published

suitable journal and submit
as per funder policy

1. Grant Application

Researcher registers grant proposal in the TWCF system

8. FAIR Workflow interface

Interface that enables researchers have greater confidence in scientific studies - linked DMPs, research outputs, instruments, experimental methods and protocols, event data (usage and citations) and visualization

This project was made possible through the support of a grant from Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation, Inc.

Capturing Metadata

Put in place mechanism and practices for metadata submission and capture.

Researchers

- Plan for output management routine
- Build protocols for not only individuals but also teams, foster lab-wide FAIR culture
- Preserve and share research outputs
- Select tools and platforms with intention.

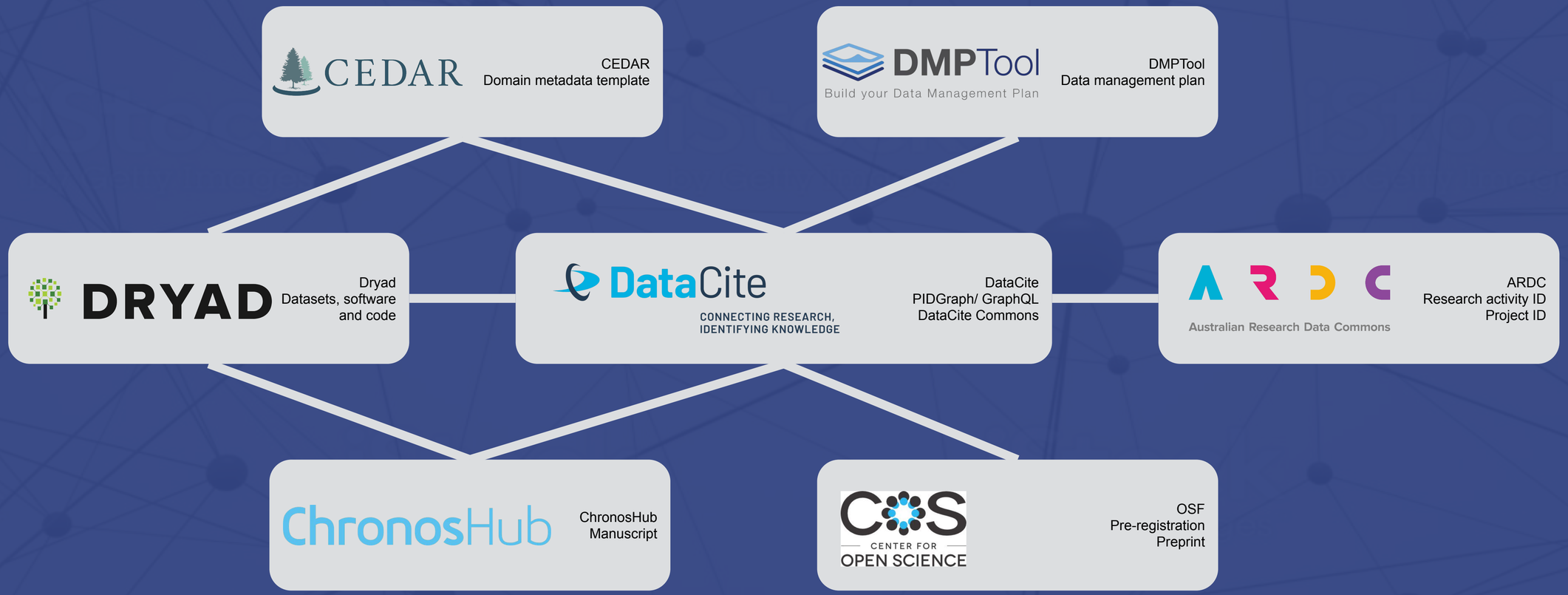
Repositories

- Adopt PIDs
- Use standardized semantic resources
- Build workflows that capture rich metadata and continuously enrich metadata over time
- Foster an open environment and interoperability by providing API endpoints

Tools and platforms

- Adopt PIDs when possible
- Integrate PID enabled features

System Integrations



Enriching Metadata

Improve the quality of metadata entered / submitted to the metadata commons.

Semantic resources

- Ontologies
- Classification
- Controlled list
- Standardized format
- ...

Comprehensive Crosswalk

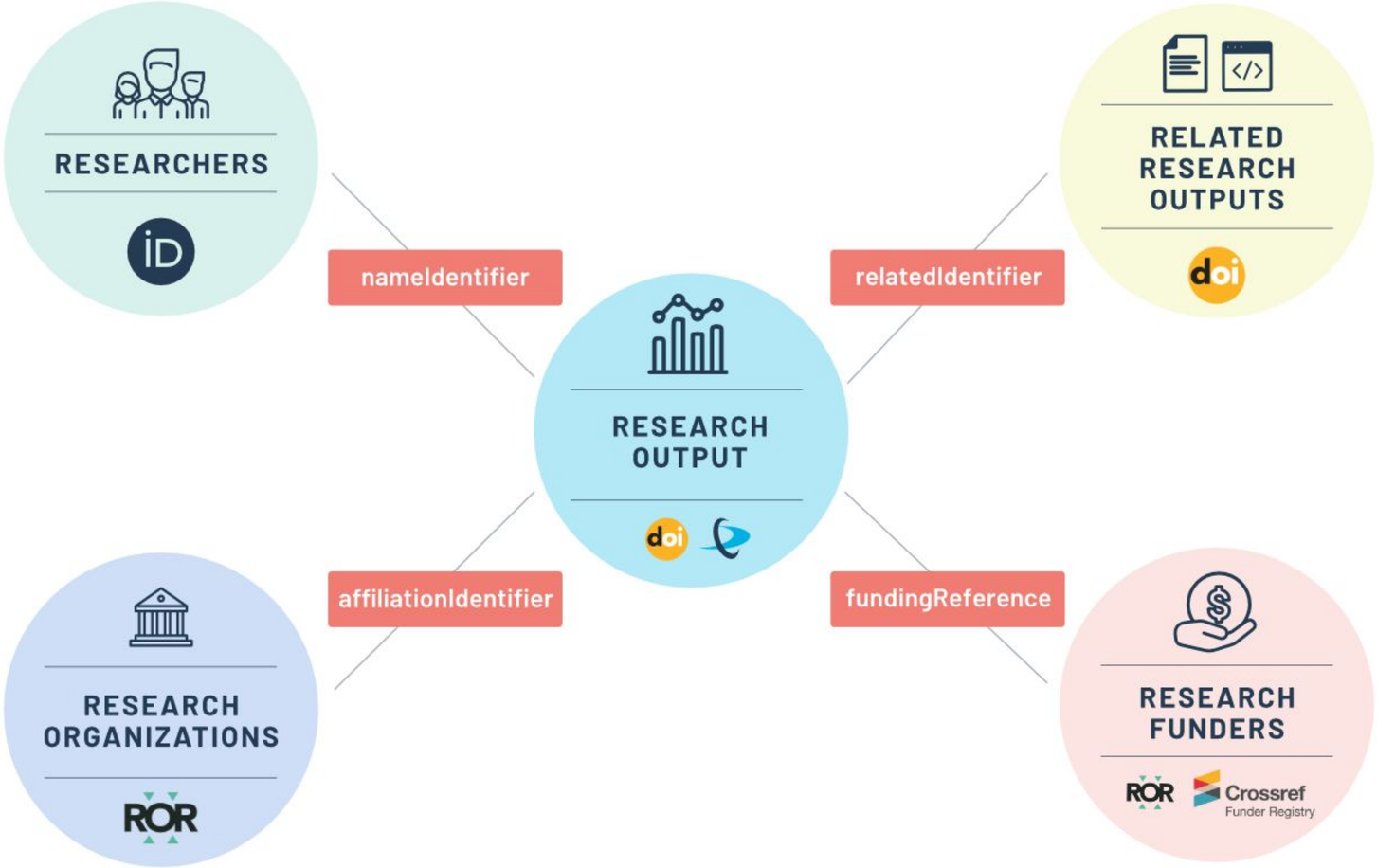
- Maximize number of submittable metadata fields when registering resource
- Introduce new metadata fields

Connection metadata

- Formulate connection information into standardized metadata
- Share connection metadata

DataCite Connection Metadata

Connect DataCite DOIs to every part of research ecosystem



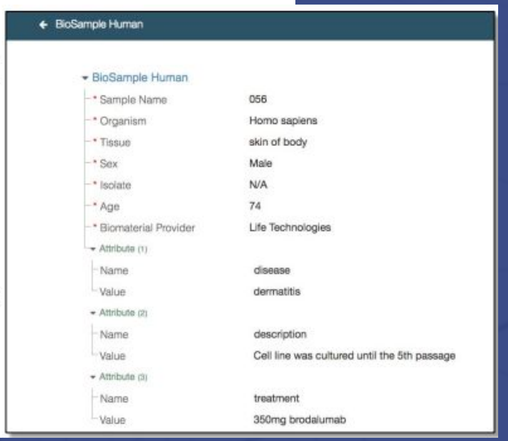
Domain specific Metadata Template

If we want to have FAIR data, we need good metadata. Good metadata need:

- **Reporting guidelines**—like MIAME—to provide a uniform structure
- **Ontologies** to provide controlled terms
- **Technology** to make it easy to author good metadata in the first place

A **metadata template** can ensure compliance with all investigator-controlled FAIR principles, including:

- Making metadata "rich"
- Using metadata vocabularies that follow the FAIR principles
- Meeting domain-relevant community metadata standards



FsF-I1-02M
Metadata uses semantic resources

FsF-R1-01MD
Metadata specifies the content of the data

FsF-R1.3-01M
Metadata follows a standard recommended by the target research community of the data

FsF-R1.3-02D
Data is available in a file format recommended by the target research community

Musen, 2021, There's No Easy Way Out: Making Data FAIR Requires Better Metadata. Information Technology Can Help. Implementing FAIR Workflows Project Partners Workshop (Presentation)

Devaraju, Anusuriya, Huber, Robert, Mokrane, Mustapha, Cepinskas, Linas, Davidson, Joy, Herterich, Patricia, L'Hours, Herve, de Vries, Jerry, & White, Angus. (2020). FAIRsFAIR Data Object Assessment Metrics (0.3). Zenodo. https://doi.org/10.5281/zenodo.3934401

FAIR WORKFLOWS

Thanks!

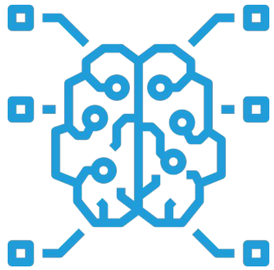
Project advisory committee

Internal advisors	**Adrian Burton**	Australian Research Data Commons
	Jennifer Kemp	Crossref
	Nici Pfeiffer	Center for Open Science
External advisors	**Russell Poldrack**	OpenNeuro
	Franco Pestilli	University of Texas
	Rachael Kotarski	British Library
	Helena Ledmyr	International Neuroinformatics Coordinating Facility
	Dylan Roskams-Edris	Tanenbaum Open Science Institute
	Jean-Baptiste Poline	Brain Imaging Centre Neuroinformatics, McGill University
	Bryan Lawrence Caron	NeuroHub (Lead PI), NeuroDataScience Director, INCF chair of scientific council

Project work packages

WP1

Workflow development



WP2

Application in research

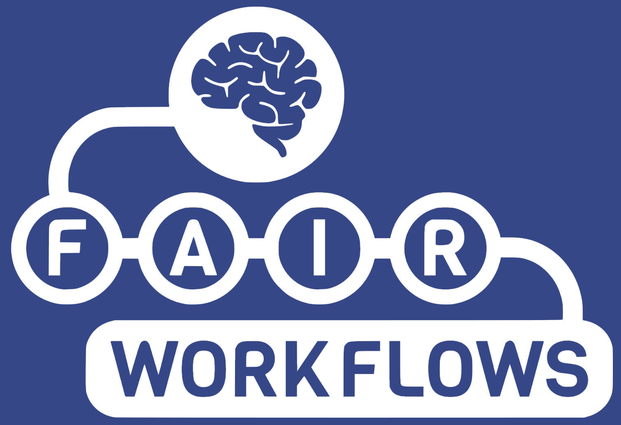
WP3

PID graph & dashboard

WP4

Adoption & dissemination

Immediate
Next steps



Keep in touch!

Email

@DataCite

slack

webpage

DataCite

Chapter cover 02

“In order to make it more dynamic, we highlight pages as well as we highlight text.”