



FAIR Bytes

The favourite lunch hour talk about all things FAIR

@GOFAIR Festival 2021

 [0000-0003-1378-5495](tel:0000-0003-1378-5495)

 [@yalahowy](https://twitter.com/yalahowy)



Origins of FAIR Bytes

Library Carpentry: FAIR Data and Software

FIXME: home page introduction

✦ Under Design

This lesson is currently in its early design stage; please check [the design notes](#) to see what we have so far. Contributions are very welcome: we would be particularly grateful for exercises and for commentary on the ones already there.

✦ Prerequisites

FIXME

Schedule

	Setup	Download files required for the lesson
00:00	1. Introduction	What does the acronym "FAIR" stand for, and what does it mean? How can library services contribute to FAIR research?
00:00	2. Findable	What is a persistent identifier or PID? What types of PIDs are there?
00:00	3. Accessible	Key question

FAIR Bytes in response to the rapidly evolving landscape of FAIR.

Providing a dynamic, flexible, inclusive and approachable resource for the community.



What is FAIR Bytes?

Lunch hour

Biweekly

Two time zones

Presentation keynote including discussions and bite sized exercises

Hands-on “Mini sprints”:

- Create training material with schema
- Templates
- FAIR cards
- Translations
- Conversation summaries

Who is it for?

- 
- Researchers
 - Data Stewards
 - Librarians
 - Research Data Managers
 - Information specialists
 - Publishers
 - Higher education representatives
 - Software engineers/developers
 - Data protection
 - Research engineers

Everyone!!!

Why FAIR Bytes?

1. Provide a framework for FAIR conversations
2. To highlight the latest in FAIR landscape
3. Address FAIR from a pragmatic how-to approach
4. Providing FAIR guidance for researchers for their workflows
5. Publish conversation summary
6. Develop top things/simple rules on a ground level of application e.g. how to set up an experiment, day-to-day activities in a lab, etc...
7. Update and add content to bite-sized FAIR lesson material
8. Advancing schema approaches to making training material more FAIR
9. Making FAIR more accessible to broader audiences beyond RDM & the libraries
10. Exploring FAIR practices from a global perspective

Topics

CSV on the web

Datasets and Provenance information in articles

Making training material more FAIR

FAIR & Software features

Is FAIR global enough?

Topics

- FAIR in workflows (R, Python)
- FAIR in data stewardship workflows
- Making your data FAIR
- FAIR in disciplines
- PIDs and connecting identifiers
- FAIR in ML and AI
- FAIR & GDPR
- Securing Metadata quality for future use
- Indigenous data collection and CARE principles
- FAIR principles and knowledge/data accessibility

CSV on the web



Radosław Drożdżewski / CC-BY-SA-4.0

CSV on the Web (CSVW) helps you build sidecars for spreadsheets!

Sidecar – a functional addition

- Motorcycle sidecar: carry best friends
- Kubernetes sidecar: support related work
- ~~Unstructured~~ README file

Given mydata.csv:

- Can serve JSON-LD sidecar at `mydata.csv-metadata.json`
- Use vocabulary terms from <https://www.w3.org/ns/csvw#> to provide extra information
 - Example: describing columns
 - Example: elaborating on a column

More info: <https://zenodo.org/record/4737519>

Datasets and Provenance information in articles

Research Papers

CMIP6 Data Citation of Evolving Data

Authors: Martina Stockhause ✉, Michael Lautenschlager

Abstract

Data citations have become widely accepted. Technical infrastructures as well as principles and recommendations for data citation are in place but best practices or guidelines for their implementation are not yet available. On the other hand, the scientific climate community requests early citations on evolving data for credit, e.g. for CMIP6 (Coupled Model Intercomparison Project Phase 6). The data citation concept for CMIP6 is presented. The main challenges lie in limited resources, a strict project timeline and the dependency on changes of the data dissemination infrastructure ESGF (Earth System Grid Federation) to meet the data citation requirements. Therefore a pragmatic, flexible and extendible approach for the CMIP6 data citation service was developed, consisting of a citation for the full evolving data superset and a data cart approach for citing the concrete used data subset. This two citation approach can be implemented according to the RDA recommendations for evolving data. Because of resource constraints and missing project policies, the implementation of the second part of the citation concept is postponed to CMIP7.

<http://doi.org/10.5334/dsj-2017-030>

Lessons from Climate Research

How do we ensure the provenance of datasets in an article?

Data Superset Citation: to give credit to the data creators and to provide statistics on data usage in literature

Data Subset Citation: used for the identification of the part of the data underlying an article

Making training material more FAIR



Integration of
metadata
schemas to
support the
reusability of
training material

100 lines (98 sloc) 4.23 KB

Raw Blame

```
1 lessons:
2 -
3 # Library Carpentry: The UNIX Shell
4 # Mandatory Level
5 description: "This Library Carpentry lesson introduces librarians to the Unix Shell. At the conclusion of the lesson you will: understand the basics of the U
6 keywords: "Unix Shell, command line, terminal, automation, filesystem, text mining, data mining, loop (programming language), free text, data cleaning" # Com
7 name: "Library Carpentry: The UNIX Shell"
8
9 #Recommended Level
10 about:
11 - "https://en.wikipedia.org/wiki/Unix_shell"
12 - "https://en.wikipedia.org/wiki/Automation" # ontology URI here
13 abstract: "Learners are introduced to the Unix-style command line interface, and basic shell navigation, as well as the use of loops and pipes for linking sh
14 audience:
15 - "Library related roles"
16 - "Information related roles"
17 author:
18 - "Library Carpentry"
19 -
20   name: "Library Carpentry"
21   email: "team@carpentries.org"
22 identifier:
23 - "http://doi.org/10.5281/zenodo.3266085"
24 inLanguage:
25 - "en-US"
26 learningResourceType:
27 - "tutorial"
28 license:
29 - "https://creativecommons.org/CC-BY_4.0"
30 # Tools mentioned/used
31 mentions:
```

Example: https://github.com/LibraryCarpentry/librarycarpentry.github.io/blob/main/_data/lessons.yml

More: <https://schema.org/LearningResource>

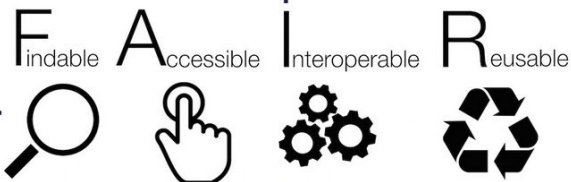
FAIR & Software features

Findable

- Capturing metadata (metadata prompts, schemas integration, links to ontologies)
- Persistent Identifiers

Accessible

- Languages, both human and computer languages



Interoperable

- Open export formats
- Json/HTML export options
- Open API

Reusable

- Data organization
- DMP
- README files
- Repository links



Is FAIR global enough?

COAR Community Framework for Good Practices in Repositories

Public version 1 – October 8, 2020

The community framework will be reviewed annually. Please send feedback to office@coar-repositories.org

Purpose

The purpose of the framework is to assist repositories to evaluate and improve their current operations based on a set of applicable and achievable good practices.

Currently, there are a number of existing frameworks and evaluation criteria that were developed to assist repositories in assessing certain facets of their operations (such as discovery, access, reuse, integrity, quality assurance, preservation, privacy, and sustainability), but these criteria are spread across different organizations and are often relevant for only one region or one type of repository.

The aim of this work was to bring together relevant criteria into a **global, multidimensional framework for assessing best practices** that can be adopted and used by different types of repositories (publication, institutional, data, etc.) and in different geographical and thematic contexts.

<https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/>

Lessons from COAR community

**Providing a global framework
for the Implementation of FAIR
principles and best practices**

Join the conversation!

Google form:

<https://forms.gle/WBRbGSsReKnpMYfK7>



TASK FORCE



Chris Erdmann

Assistant Director for Data
Stewardship American
Geophysical Union (AGU)



Sara El-Gebali

Research Data Manager
(MDC)



Donny Winston

President
Polyneme LLC



Annajiat Alim Rasel

Senior Lecturer
School of Data and Sciences
Brac University



Nabil Ksibi

Engagement Lead Global
Consortia (ORCID)

Special thanks: Naoe Tatara, Nina Exner, Agata Bochynska