Advancing FAIR:

Connecting Stakeholders, Communities & Resources

https://gofair.us/

Overview

- FAIR Principles
- GO FAIR Initiative
  - Technologies; Data Stewardship
- GO FAIR US!
  - Activities and growth
  - How to get involved
- Discussion / Q+A

FAIR is…

A set of principles that describe the attributes data need to have to enable and enhance reuse, by humans and machines.

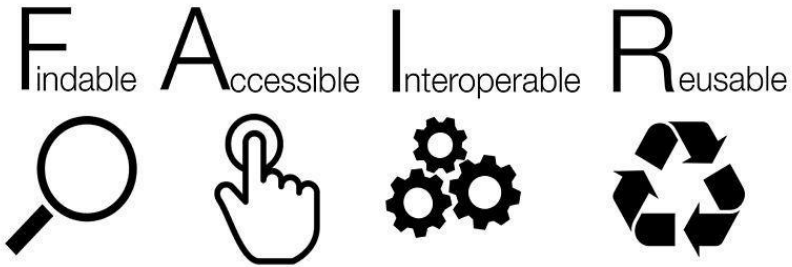


Image CC-BY-SA by Sangya Pundir

FAIR PRINCIPLES

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

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;

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;

Reusable:

R1. (meta)data are richly described with 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 detailed provenance;

R1.3. (meta)data meet domain-relevant community standards;

Luiz Bonino

DOI 10.1038/sdata.2016.18

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016).

GO FAIR US

FAIR as an Aim:

The Principles are:

- An aspiration, a journey.
- Ambiguous.
- A spectrum.
- Domain respectful / specific
- Implementable with today's protocols and standards.
- A small part of indicators.
- A framework for prompting organizational change
- Work in progress

The Principles are not:

- A standard.
- Strict.
- One size fits all.
- One domain
- Inventing new protocols.
- Technology specific
- Anything to do with quality.
- Synonymous with open.
- An architecture
- "Tablets of stone"

CC BY

From: Carole Goble “FAIR History and Future”

GO FAIR US

GO FAIR Initiative

Pillars

- GFISCO (GO FAIR International Support & Coordination Office)
- National Offices
  - France, Brazil, Germany, U.S.
- Mobilization around Pillars

GO FAIR US

Data Stewardship

Improving data value across its lifecycle

- Linking Research Data Management with Planning for long-term use
- Application of standards and semantics technologies
- Adding value to support new kinds of use
- Training for new roles and adding new skills
- Facilitating Policy

GO FAIR Technologies: Design-as-we-go

The Three-Point FAIRification Framework

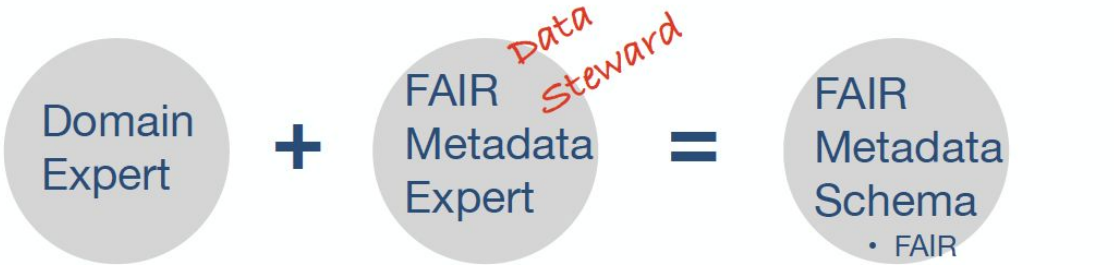
How a systematic process of
FAIRification can power convergence

GO FAIR US

Metadata for Machines Workshops

https://www.go-fair.org/today/making-fair-metadata/

Metadata for Machine Workshops assist domain experts who want to produce FAIR resources with the creation of machine-actionable metadata that **satisfies** and **adheres** to the FAIR Principles.



Erik Schultes

International FAIR Convergence Symposium

Slides: https://osf.io/37vkp/

GO FAIR US

FAIR Implementation Profiles & Matrix

FAIR Implementation Matrix

Note: This matrix is a working tool to collect current and planned choices and detected challenges of the GO FAIR communities (mostly INs). This is a living and growing document and awaits

On the OSF this document: https://osf.io/n7uwp/

Red indicates waist of hourglass

*Blue is an Implementation **Choice***

*Orange is Implementation **Challenge***

Green highlight indicates a service provided by the IN or spin-off

Blank cell is not relevant for IN

FAIR Principle	Services	Digital Object	Most used	C2CAMP	OPEDAS	PHT (infra)	Rare-Diseases	GERDI	NOMAD
	central to all	DOIP	DOIP/HANDLES	DOIP	DOIP	DOIP	DOIP	-	curenly private, DOIP to come
	central to all	Metadata format	RDF/JSON/XML	many primary (databse, XML)	RDF	RDF/JSON/XML	RDF, XML, JSON-LD	DataCite plus community-specific extensions	database, xml
	central to all	Metadata access protocol	OAI-PMH	many use OAI-PMH	LDP/FDP	REST/SOAP	LDP/FDP, OAI-PMH, RESTful	OAI-PMH	OAI-PMH to come
	central to all	Metadata core elements	Dublin core	not applicable	TBD on M4M	TBD on M4M	TBD on M4M; DC, etc	TBD	NOMAD set
	Technology	Data Format		many different	RDF for interop.	RDF/JSON/XML/Binaries	RDF/JSON/XML/Binaries/Tabular data/VCF/BAM/Phenopackets/etc.	Agnostic, JSON supported for extended functionality	get lotsmof dat formats, reduction to format in archive
	Technology	Data Access Protocols (MR/A)		now mostly HTTP/REST, in future DOIP	LDP/FDP	PHT-standard	Discussion with implementation partners. (CASTOR/MOLGENIS) Looking at internet standard protocols like OpenID Connect, SAML and related.	*Determined by data source, communicated via metadata*	*https, in future DOIP perhaps*

GO FAIR US

FDPs: FAIR Data Points

- FDP is a standards-based, web-accessible store of *information about data sets*
- Components include
  - Defined “API”
  - Reference implementation
  - API client

https://www.fairdatapoint.org/

Launch of Data Together

GO FAIR, together with the Committee on Data (**CODATA**), Research Data Alliance (**RDA**), and World Data Systems (**WDS**) (the four major international data organizations), announced joint commitment to work together to

> “ optimize the global research data ecosystem and to identify the opportunities and needs that will trigger federated infrastructures to service the new reality of data-driven science.”

Data Together Statement

GO FAIR US

GO FAIR US

Advance FAIR Capacity in the US:

- National leadership and cross-community engagement
- Train FAIR Data stakeholders
- Extend Implementation Networks (IN) into US
- Technical services - building and support

Goals:

- Build connections among FAIR stakeholders
- Foster community discussions on the topic of FAIR
- Update communities on developments in FAIR tech
- Offer training and a space to share expert knowledge
- Liaise with GO FAIR International Support and Coordination Office (GFISCO).
- Increase awareness of GO FAIR Implementation Network (IN) activities in the US and worldwide.
  - DSCCs

GO FAIR US

Workshop : Advancing FAIR and GO FAIR in the U.S.

- Facilitate community of practice
- Improve understanding of FAIR technologies
  - how to teach this to others
- Preparation to support FAIR data management and policies in multiple contexts
- Collaboration with GFISCO & the South Big Data Innovation Hub

Funding - NSF Award #1937953

Expertise and Area Leads at GO FAIR US

Christine: Head of GF-US
- FAIR and AI/ML
- Institutional planning
- FAIR/RDM for Research Computing
- Infrastructure side of FAIR data ecosystem
- Geosciences, metabolomics, data science, CS / HPC

Nancy:
- FAIR implementation tools using metadata & linked data technologies
- Educator Community Engagement / Community of Practice facilitation
- Geosciences, social science and digital humanities

Juliane:
- Metadata and Ontologies for discovery
- Data curation / RDM (and training)
- RDA

Melissa:
- Research Policy
- Data Use & Users / practices (earth/life sci; neurosci)
- OSN; BD Hubs; ADSA; GO FAIR; CODATA

Chris:
- Community Development / Coordination
- Training & Skills Development (The Carpentries)
- Research Software/Data (Software & Data Citation Principles)

GO FAIR US

Virus Outbreak Data Network (VODAN)

The VODAN Implementation Network is one of the joint activities carried out by **CODATA, RDA, WDS, and GO FAIR** (Link to the **Data Together Statement**).
Read the full statement on **Data Together COVID-19 Appeal and Actions**.

Active GO FAIR Implementation Network

The spread of the virus causing the COVID-19 outbreak is far from over. During this epidemic and in earlier occasions, we have seen severely suboptimal data management and data reuse. Moreover, access to the immensely valuable data of past and current epidemics is not always equally accessible for different affected populations and countries. For instance, the data from the past Ebola epidemics are very difficult to find, to access, and if accessible, they are not interoperable, *let alone reusable*. Under the urgent need to harness machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreaks, we need to do better and ensure that data are FAIR (in this sense also meaning **F**ederated, **AI**-**R**eady).

Purpose of the Implementation Network

This time, we can do better. We now have the technical ability, as well as the commitment from experts in a series of affected countries, to make the SARS CoV-2 virus data FAIR, meaning that they are Findable, Accessible, Interoperable and thus Reusable by both humans and machines, during this epidemic of COVID-19. The technical components that make this possible can remain in place, waiting in ready state *for potential future infectious disease outbreaks*.

VODAN-Africa

Univ. de Sousse, Tunisia's FDP in action!
https://fdp.uc.rnu.tn/

VODAN-Africa – for Africa by African partners

VODAN-Africa is made up of health ministries, universities with government and foundation support.

Groundbreaking initiatives:

COVID-19 effects people in clusters, requires data for prevention, direction of resources, evaluation of approaches and adjustments

Metadata for Machines workshops

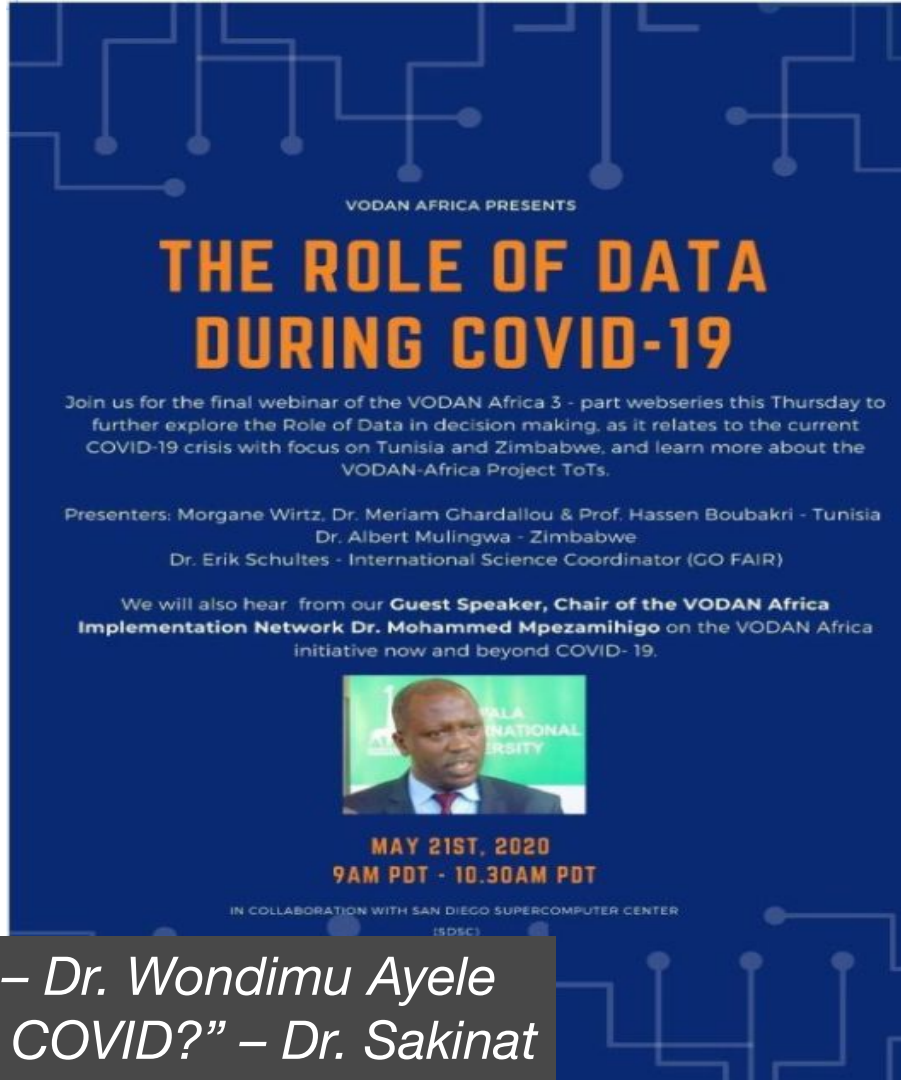
Train the Trainers (ToT) New skill-based capacities: new data steward workforce

FAIR data points Infrastructure for future crisis

"Until every country is safe, no one is safe." – Dr. Reginald Nalugala

VODAN-Africa Highlights

- **Internet unreliable in Uganda, Ethiopia, can't assume WFH for data stewards**
- **Tunisia – tracking and protecting those who rely on daily income; COVID another reason to marginalize already marginalized groups, neighboring conflict**
- **Social data needed: Alcohol abuse, domestic violence**
  "What comes after COVID?"
- **Liberia – kids at home, no technology, no remote school**



"Data save lives!" – Dr. Wondimu Ayele

"How would I meet these people without COVID?" – Dr. Sakinat

Force11 Scholarly Communication Institute

FSCI'20

- Four FAIR-related training sessions
  - Materials will be publicly available soon!
- Members of our team led 2 of these:
  - FAIR Data in the Scholarly Communications Life Cycle
  - Advancing FAIR Data Stewardship: Fostering Institutional Planning and Service Development
- Significant interest in FAIR by information professionals around the globe
  - "What does this mean for my organization?"
  - "What does this mean for current RDM and curation services?"
  - "How to we teach about this?"
  - "How do we implement this?"

GO FAIR US

Tools for the Community

FAIR Data Stewardship Plan Template for Organizations and Institutions

FSCI 2020 Course: T16- Advancing FAIR Data Stewardship: Fostering Institutional Planning and Service Development

Table of Contents

7. Priorities and Resource Wishlist

Review the goal, assessment, opportunity, and training plan.

1. List the actions these suggest and record a timeframe using the worksheet.
2. For each time frame, collapse into 10 items. This can be done by categorizing like things into one item.
3. Order based on priority.

Priorities Wish List Worksheet

Action to Take	Short-term (0-12 mos)	Medium Term (1-2 Years)	Longer Term (3+ years)

Short-Term Priorities

Item	Resources available	Resources needed
1.		
2.		
3.		
...		

Medium-Term Priorities

Item	Resources available	Resources needed
1.		

GO FAIR US

Earth Science Information Partners (ESIP) activity to date: FAIR Implementation Profile working session at ESIP Summer 2020

Goal: Move toward a FAIR data ecosystem by:

- Exploring what the ESIP community knows about ***implementing*** the FAIR Data Principles
- Introducing the GO FAIR "tools" for documenting agreements on FAIR related community standards:
  - FAIR Convergence Matrix
  - FIPs (FAIR Implementation Profiles)
  - FIP Wizard:
    - https://fip-wizard.ds-wizard.org/knowledge-models

Takeaways:

- People see the value of knowing how others' have implemented FAIR, and want to know more about machine actionable metadata
- GO FAIR tools and services can be very helpful to ESIP in implementing FAIR
- ESIP partners should contribute more to the GO FAIR tools like FAIRsharing.org directory and create more FAIR Implementation Profiles!

GO FAIR US

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
00:00	4. Interoperable	What does interoperability mean? What is a controlled vocabulary, a metadata schema and linked data? How do I describe data so that humans and computers can understand?
00:00	5. Reusable	Key question
00:00	6. Assessment	How can I assess the FAIRness of myself, organisation, service, community… ? Which FAIR assessment tools exist to understand how FAIR you are?
01:00	7. Software	Key question

https://librarycarpentry.org/lc-fair-research/index.html

GO FAIR US

Next steps for the FAIR data & software lesson

OCTOBER 4, 2020 - CHRIS ERDMANN, KRISTINA HETTNE, LIZ STOKES

On 2020 August 27-28, 29 participants from 21 institutions and 11 countries took part in the CarpentryCon @ Home - Library Carpentry FAIR Data & Software lesson sprint. The ultimate goal of the sprint was to develop a shared FAIR lesson that communities across the world can benefit from. Overall, participants made 35 commits, including 347 additions and 62 deletions, and closed 9 issues out of 16 opened during the sprint. Lesson development also occurred outside of GitHub, in Etherpads and Google Docs, and is in the process of being merged. *Thanks to all the participants and their contributions!*

https://gofair.us/news/2020-10-04-FAIR-lesson-next-steps.html

IN PARTNERSHIP WITH UCLA LIBRARY

FSCI 2020

FORCE11 Scholarly Communication Institute | AUGUST 3 – 13, 2020 | ONLINE

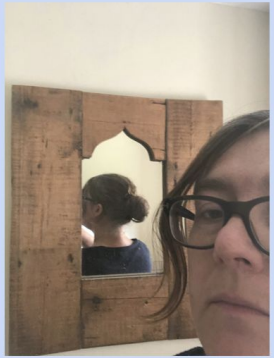
FAIR Data in the Scholarly Communications Lifecycle

Course W21 - Session one

Natasha Simons
ARDC AUSTRALIA

Chris Erdmann
RENCI UNITED STATES

Daniel Bangert
UGOE/RDA GERMANY



Fiona Murphy
More Brains, UK

http://doi.org/10.5281/zenodo.3987052

How to support FAIR?

FAIR is a shared responsibility

FAIR needs an ecosystem of services

FAIR needs people

Open (research) needs FAIR

GO FAIR US

Recommendations

(1) **funders** and **institutions** should consider FAIR alignment and data sharing as part of research assessment, among other criteria;

(2) **services** should support domain-specific ontologies by identifying disciplines that lack ontologies and enriching existing registries of ontologies;

(3) **repositories** should support FAIR data by developing tools, such as APIs, sharing best practices, and undergoing FAIR-aligned certification; and

(4) **institutions** should support FAIR awareness and implementation by establishing data stewardship programs providing simple and intuitive training for researchers.

Koers, H., Bangert, D., Hermans, E., van Horik, R., de Jong, M., & Mokrane, M. (2020)

GO FAIR US

Get Involved

Connect with GO FAIR US:

- Google Group Email List. GO FAIR US's main communication channel. Sign up to receive periodic updates.
- Slack Channel. Sign up to the GO FAIR US Slack to chat with community members.
- Twitter. Follow GO FAIR US on Twitter, learn about what community members are doing, and get updates.
- Go FAIR US News & Events. Stay up-to-date on news and events of interest to the Go FAIR US community.
- GitHub Organization. Submit issues/pull requests to Go FAIR US resources on GitHub.
- Contact. Send GO FAIR US general inquiries.

Deeper ways to engage:

- Join an Implementation Network (IN).
- Contact us to present on your FAIR initiative to the GO FAIR community.
- Submit a guest event and/or news blog post.

Webinars

- 15 Oct - Basics
- 10 Dec - Intro to GO FAIR US
- 14 Jan 2021 - FAIR & Research Software

Learn More

- Join VODAN: tinyurl.com/join-vodan
- VODAN-Africa: https://www.vodan-totafrica.info/
- Models of Infectious Disease Agent Study (MIDAS): https://midasnetwork.us/covid-19/

  June 26, 12 PM ET

  Confronting the Unknown: Immunopathology verses Protective Immunity in SARS-CoV-2

GO FAIR US

Further resources

Recommendations for Services in a FAIR data ecosystem
https://doi.org/10.1016/j.patter.2020.100058

Cultural Obstacles to Research Data Management and Sharing at TU Delft
http://doi.org/10.1629/uksg.484

Turning FAIR into Reality (European Commission Expert Group on FAIR Data)
https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1

FAIR in practice reference list https://doi.org/10.5281/zenodo.3898673

Engaging Researchers with Data Management: The Cookbook
http://doi.org/10.11647/OBP.0185

Evaluate your RDM offering https://sparceurope.org/evaluate-your-rdm-offering/

GO FAIR US

GO FAIR US VODAN webinar series (spring 2020)

1. "An Introduction to the Virus Outbreak Data Network (VODAN) and Related Developments with EURETOS"
    1. Presented by Barend Mons, President of CODATA and Founder of GO FAIR.
    2. Barend's presentation for VODAN-IN FAIR Data Points as a service for data-driven disease outbreak tracking and research is also available.
2. "VODAN-Africa / Training of Trainers (ToT)"
    1. Presented by Professor Mirjam van Reisen, Leiden Institute of Advanced Computer Science, Leiden University.
3. Two-part session. "Enabling the Collection and Exploration of COVID-19 Data via FAIR Principles"
    1. "VODAN: FAIR Data Work in Action"
        - Presented by Albert Mons (Phortos Consultants) and Luiz Bonino (GO FAIR).
            - World Health Organization eCRF forms.
    2. "The SARS-Cov-2-SPOKE Molecular Explorer: A Public Knowledge Graph Visualization Tool"
        - Presented by Sergio Baranzini and Sharat Israni (University of California, San Francisco).
4. Three-part session. "Fighting COVID-19 by Mining Insights from Heterogeneous Datasets"
    1. "COVID19-net"
        - Peter Rose and Ilya Zaslavsky (SDSC, University of California, San Diego).
    2. "From Data to Knowledge"
        - Presented by Iris Shen, Principal Data Scientist (Microsoft Academic Team, Microsoft).
    3. "Distant Reading for Quick Insights"
        - Presented by Natalie Meyers, E-Research Librarian (Navari Center for Digital Scholarship, Hesburgh Library) and Eric Morgan (Hesburgh Library, Notre Dame).

GO FAIR US

Great, free resource - new recordings!

International FAIR Convergence Symposium

- Virtual meeting: 30 Nov. - 4 Dec. 2020
- CODATA & GFISCO

Thank You

Chris Erdmann

Engagement, Support, Training Expert
RENCI, UNC

Christine Kirkpatrick

Division Director
Research Data Services, SDSC

Juliane Schneider

Senior Bioinformatics Analyst/Data Liaison, Sage Bionetworks

Melissa Cragin

Chief Strategist, Data Initiatives
RDS, SDSC

Nancy Hoebelheinrich

Information Analyst / Principal, Knowledge Motifs LLC

This work is supported in part by NSF awards 1937953, and 1916481.