THE FAIR PRINCIPLES

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;

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;

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;

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;

https://www.nature.com/articles/sdata201618

GO FAIR

Do you find any of the FAIR Principles

… weird?

What about this one:

F3: Metadata clearly and explicitly include the identifier of the data it describes

That's an unusually specific rule, given the generality of the other Principles! ...why?

Understanding the Objective/Purpose of each Principle helps clarify the basis of a "Maturity Indicator" we designed to measure it

FAIR Principles Explained:
https://www.go-fair.org/fair-principles/

FAIRNESS ASSESSMENT CHALLENGES

WHY TO ASSESS?

- Because everybody is talking about FAIR and my resources should be seen as FAIR, whatever this means?

- To satisfy funders requirements?

- To serve as a guideline for achieving higher levels of interoperability and reuse with clarity on the concrete benefits (help improve)?

- **Metadata and data?**

- **Only metadata?**

- **Only data?**
  - What do you mean by data?
  - In the FAIR principles, data refers to a variety of different resources, e.g., "traditional" data, services, software, APIs, vocabularies, ontologies, articles, etc.

HOW TO ASSESS?

- **Manual**
  - Takes advantage of human understandable artifacts, which are currently prevalent
  - May lead to subjective assessments and, therefore, harder to compare resources
  - Harder to scale
  - Harder to evaluate FAIR for machines, which is the main goal of the FAIR principles
- **Automatic**
  - Requires more rigor on the assessed resources
  - More likely to produce objective assessments
  - Easier to scale
  - Able to check if machines can, in fact, "work" with the (meta)data

GO FAIR

- **Need for a scoring system**
  - One score for the 4 aspects of FAIR? Does not seem useful.
  - One score per aspect (F, A, I and R)?
  - One score per principle? What about the sub-principles?
  - Is there a hierarchy among the principles? Is there an order of precedence? Or different weights?
  - Is there an acceptable minimal FAIR level? Should it be across domains and applications or domain/community-dependent?
  - Do we use a pass/fail approach or introduce intermediary compliance levels in each/some evaluation?
- **Need for a visual representation of the scores**
  - To facilitate quick perception of the FAIRness level, a visual representation of the FAIR scores is required, e.g., stars, bars, etc.

WHAT MAKES A MEASUREMENT "GOOD"?

- Clear: so that anybody can understand what is meant.

- Realistic: so that anybody can report on what is being asked of them.

- Discriminating: so that we can distinguish the degree to which a resource meets a specific FAIR principle, and can provide instruction as to what would maximize that value.

- Measurable: The assessment can be made in an objective, quantitative, machine-interpretable, scalable and reproducible manner → transparency of what is being measured, and how.

- Universality: The extent to which the MIis applicable to all digital resources.

Source: Mark D. Wilkinson

GO FAIR

We designed a set of parameters that must be considered for every MI.

The parameters are designed to help ensure that the MI you are designing is "good".

MI Identifier	FAIR MIs should, themselves, be FAIR objects, and thus should have globally unique identifiers.
MI Name	human-readable name for the MI
To which principle does it apply?	MIs should address only one sub-principle, since each FAIR principle is particular to one feature of a digital resource; MIs that address multiple principles are likely to be measuring multiple features, and those should be separated whenever possible.
What is being measured?	A precise description of the aspect of that digital resource that is going to be evaluated
Why should we measure it?	Describe why it is relevant to measure this aspect
What must be provided?	What information is required to make this measurement?
How do we measure it?	In what way will that information be evaluated?
What is a valid result?	What outcome represents "success" versus "failure"
For which digital resource(s) is this relevant?	If possible, a MI should apply to all digital resources; however, some MIs may be applicable only to a subset. In this case, it is necessary to specify the range of resources to which the MI is reasonably applicable.

GENERAL CHALLENGES

- Clarify that nobody has been asked to be 100% FAIR. Many times a lower FAIRness level is perfectly adequate.

- Consider different preparedness levels from different communities;

- How to deal with the conflicting forces that, from one side want to push the communities towards a better (and FAIRer) data landscape and, from the other side, want to preserve the *status quo* (existing "kingdoms") but labeling themselves FAIR?

- Who will define the assessment criteria?

- Who will execute the assessments based on the defined criteria?

- Should we have a unique set of assessment criteria? Or a core set for general comparison and domain-specific sets on top of the core for the specific needs of a given domain/application?

Luiz Bonino

International Technology Coordinator – GO FAIR
Associate Professor BioSemantics – LUMC

E-mail: **luiz.bonino@go-fair.org**
Skype: **luizolavobonino**
Web: **www.go-fair.org**