

“Anatomy of a FAIR Dissertation”

Michael Witt

<http://www.lib.purdue.edu/research/witt>

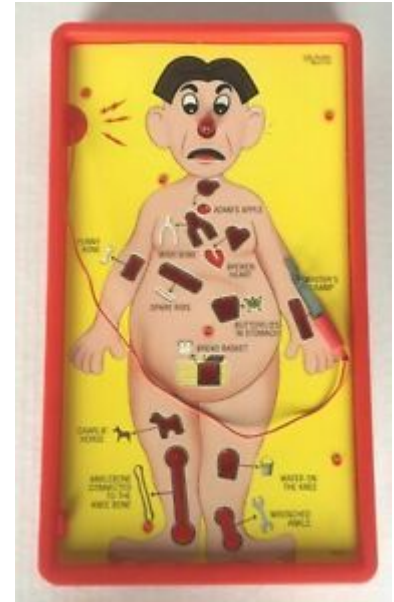
Interim Associate Dean for Research & Associate Professor of Library Science
Purdue University Libraries and School of Information Studies - USA

**22nd Symposium on Electronic Theses and Dissertations
ETD2019 @ Universidade Portucalense, Porto, Portugal
November 7, 2019**



What could / does / should a FAIR dissertation look like?

- Audience calibration & engagement
- FAIR Principles (walk-through tutorial)
- Thought experiment and discussion
- One approach for implementation



A quick word about lasers and personal safety

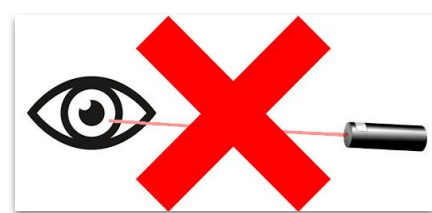
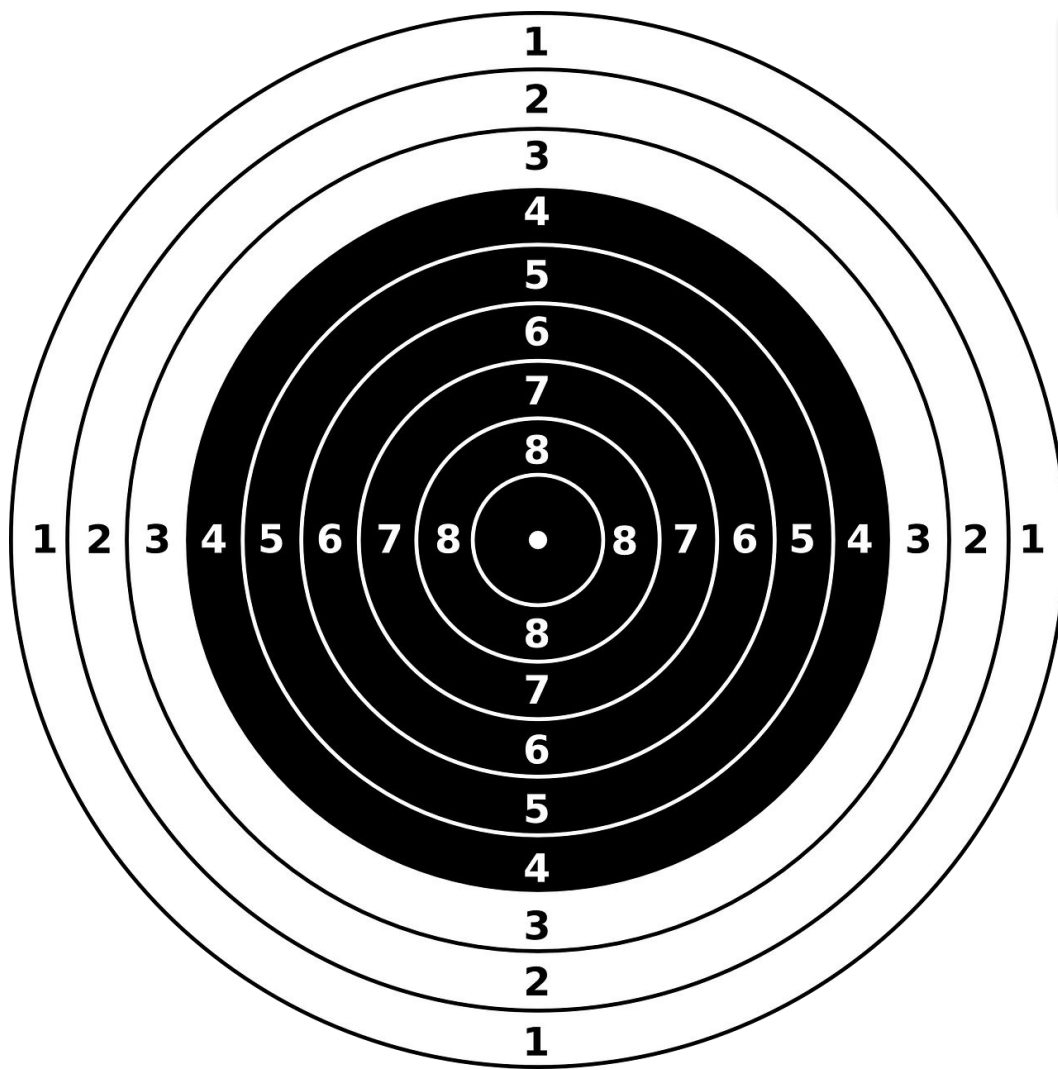
1. Please don't point them at me

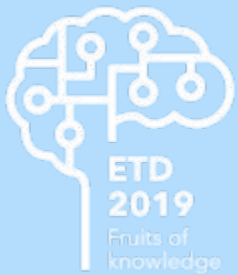
não atire em mim, por favor!

2. Please don't point them at your eyes

não atire em si mesmo!





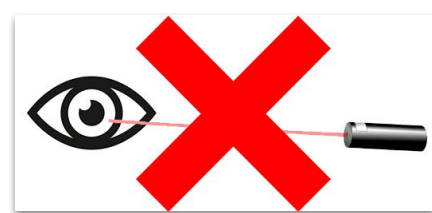


Where are we from?



$n =$

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64



SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

FAIR Guiding Principles

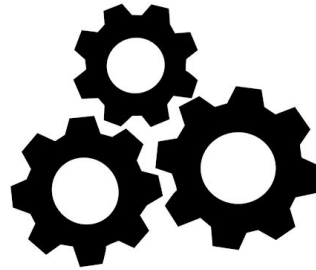
- an urgent need to improve the infrastructure supporting the reuse of scholarly data
- a concise and measurable set of principles
- specific emphasis on enhancing **the ability of machines to automatically find and use the data** (as well as humans)
- the principles apply not only to data in the conventional sense, but also to the algorithms, tools, and workflows that led to that data
- ‘machine actionable’ indicates a continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously acting, computational data explorer
- “all scholarly digital research objects ... benefit from application of these principles”

F
Findable

A
Accessible

I
Interoperable

R
Reusable



As open as possible, as closed as necessary : FAIR ≠ Open Access

FAIR metadata & data for humans & machines

What is FAIR metadata for humans?

What is FAIR data for humans?

What is FAIR metadata for machines?

What is FAIR data for machines?



To be Findable

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

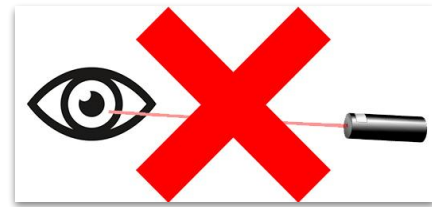
F2. data are described with rich metadata

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

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



To be Findable



F1. ETDs are assigned a globally unique and persistent identifier

F2. ETDs are described with rich metadata for discovery

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

F4. ETDs are registered or indexed in a searchable resource

No

Maybe?

Yes

To be Accessible

[A1](#). (meta)data are retrievable by their identifier using a standardized communications protocol

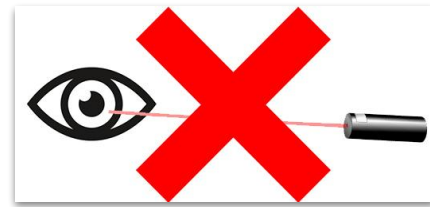
[A1.1](#) the protocol is open, free, and universally implementable

[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



To be Accessible



A1. ETDs 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 is accessible, even when the ETD is not available

No

Maybe?

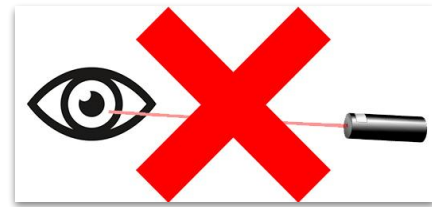
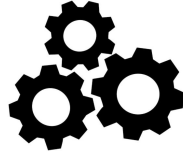
Yes

To be Interoperable

11. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
12. (meta)data use vocabularies that follow FAIR Principles
13. (meta)data include qualified references to other (meta)data



To be Interoperable



11. ETDs use a formal, accessible, shared, and broadly applicable language for knowledge representation.
12. ETDs use vocabularies that follow FAIR principles
13. ETDs include qualified references to other (meta)data

No

Maybe?

Yes

To be Reusable



R1. meta(data) are richly described with a plurality of accurate and relevant attributes

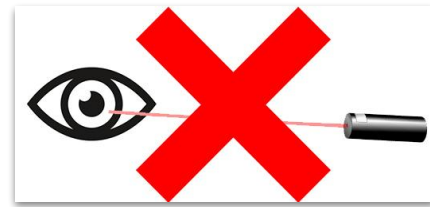
R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

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



To be Reusable



R1. ETDs are richly described with a plurality of accurate and relevant attributes for reuse

R1.1. ETDs are released with a clear and accessible data usage license

R1.2. ETDs are associated with detailed provenance

R1.3. ETDs meet domain-relevant community standards

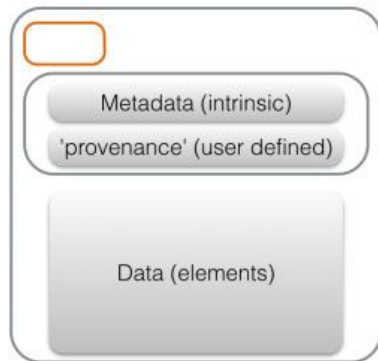
No

Maybe?

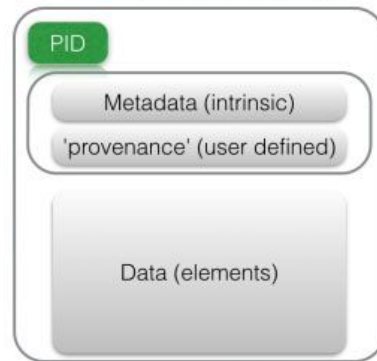
Yes

Data as increasingly FAIR Digital Objects

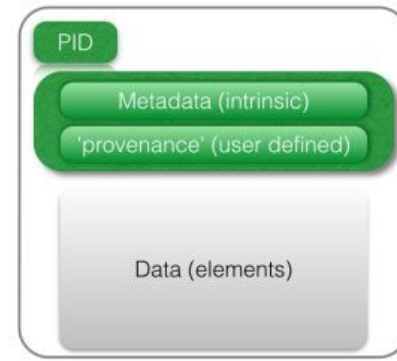
Totally UNFAIR



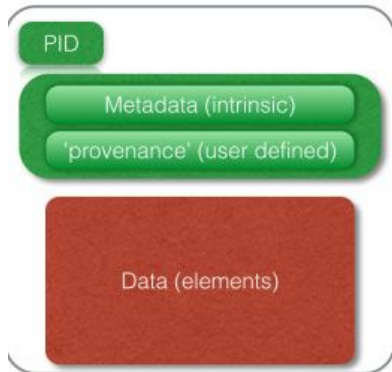
Findable
Usable for Humans



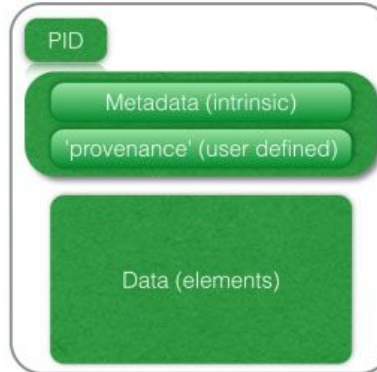
FAIR metadata



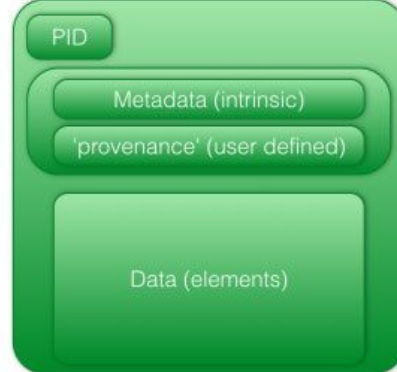
FAIR data-
restricted access



FAIR data-
Open Access



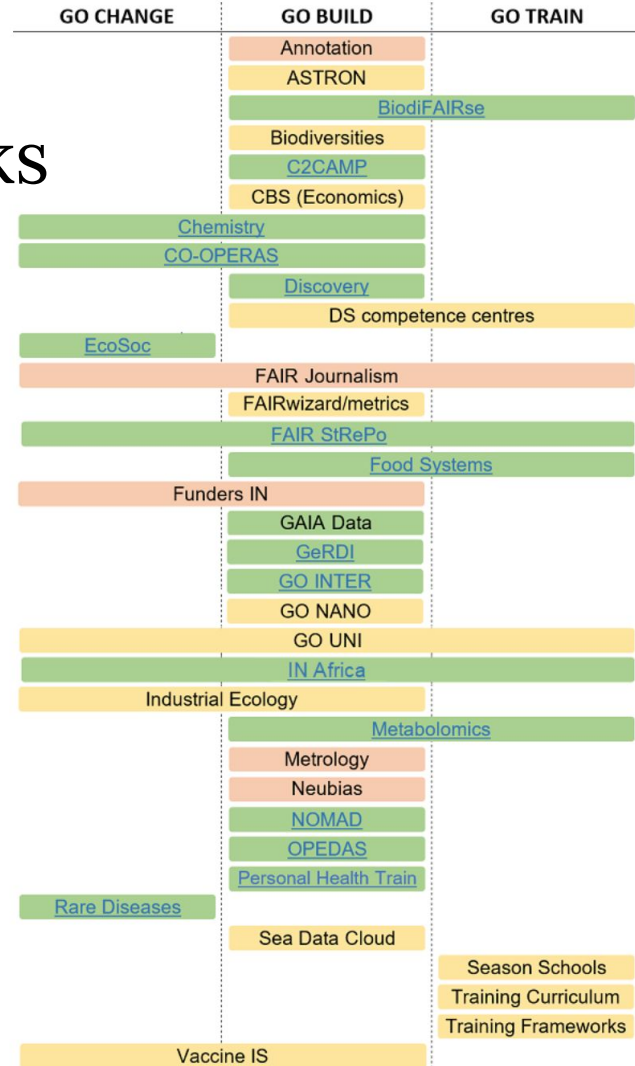
FAIR data-
Open Access/Functionally Linke

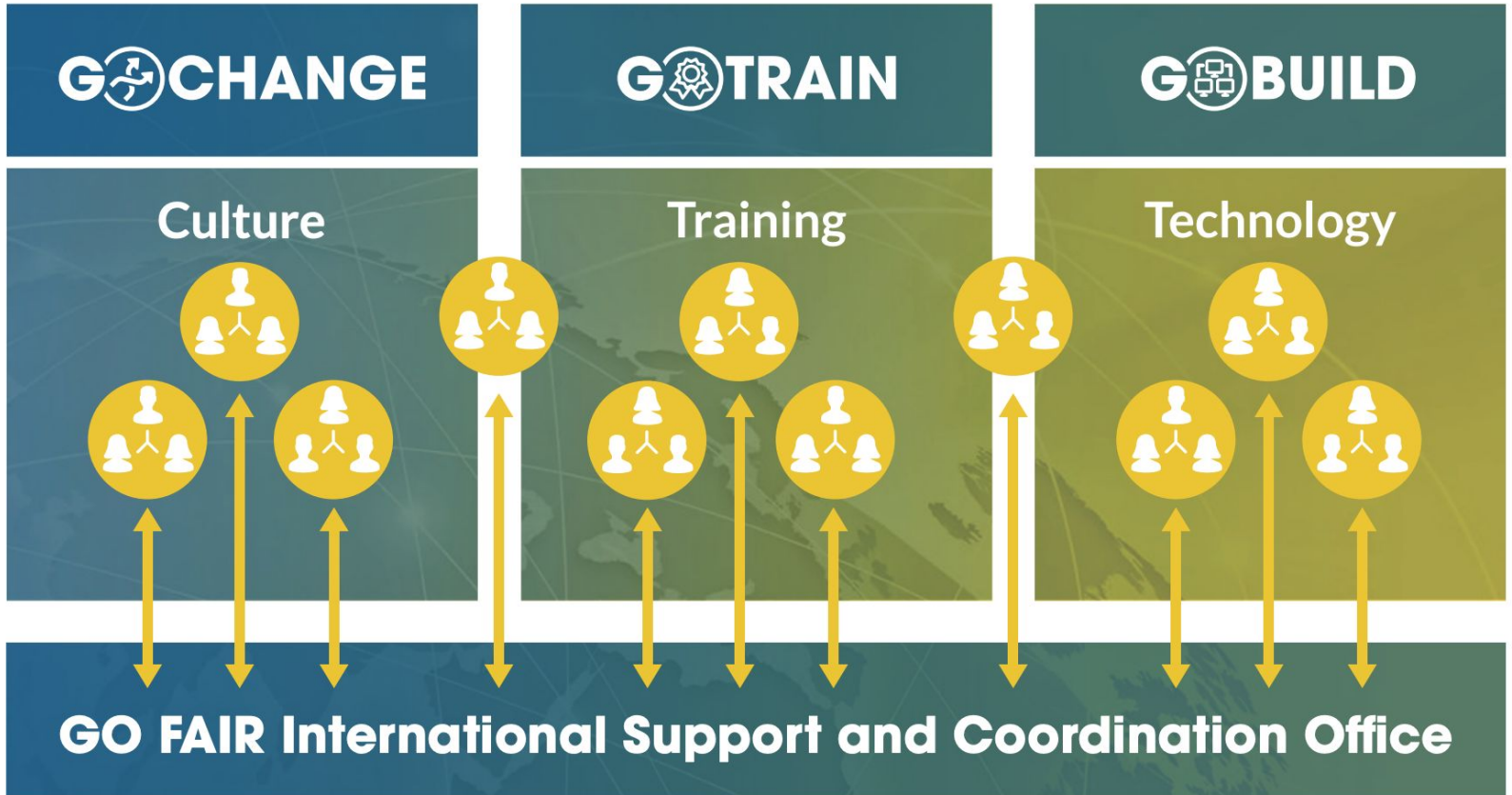


GO FAIR Implementation Networks

“A GO FAIR Implementation Network is a consortium committed to defining and creating specific materials and tools as elements of the [Internet of FAIR Data and Services](#)”

1. Have defined objectives such as building a component or a service, defining a training curriculum or a data stewardship policy
2. Comply with the [GO FAIR Rules of Engagement](#)
3. Have enough critical mass to be regarded as leaders in the field of expertise (Potential International Community Impact)





<https://www.go-fair.org>



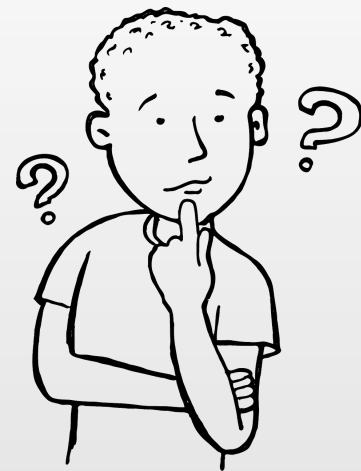


Global ETD Search

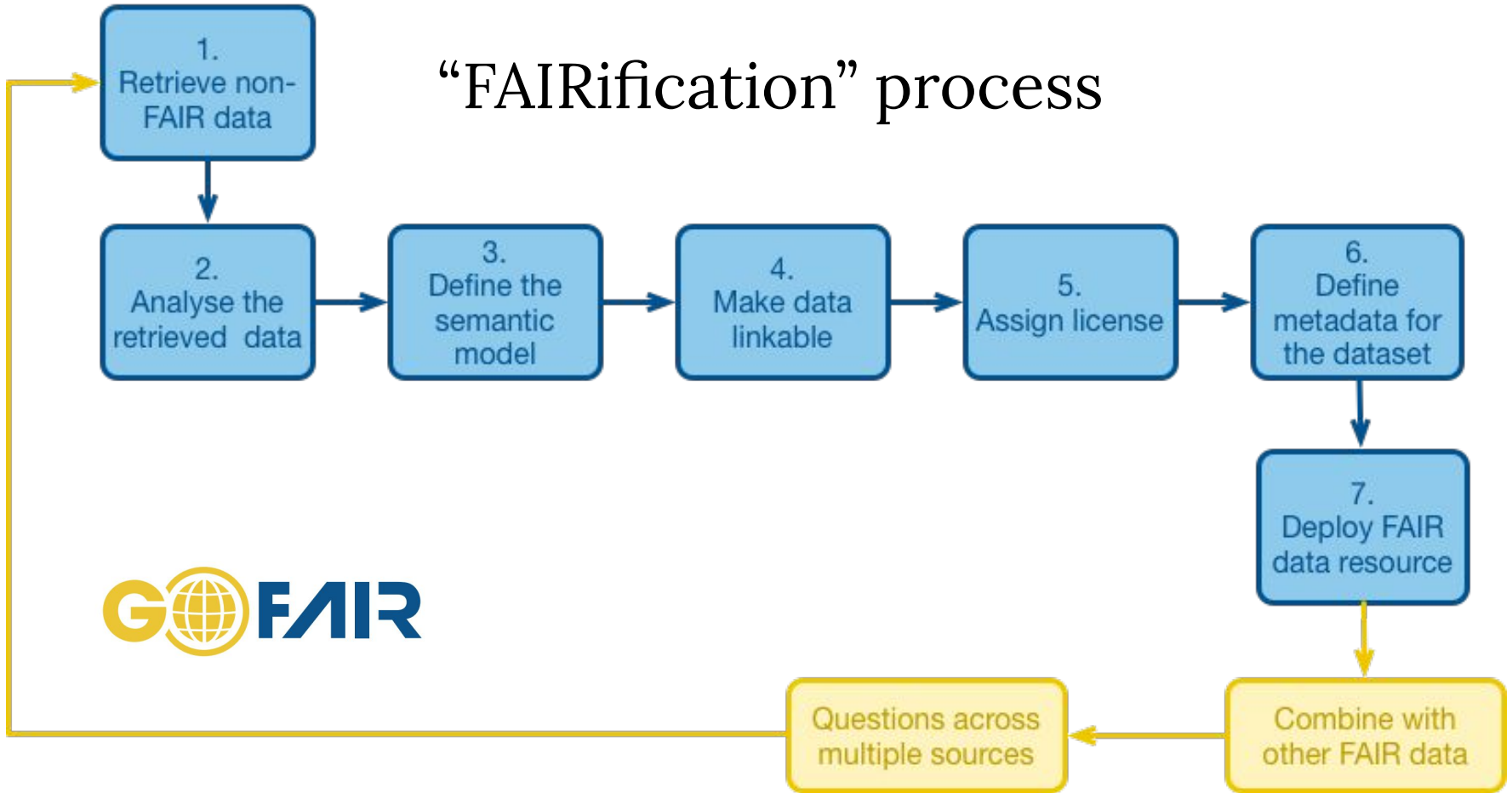
Search the 5,881,807 electronic theses and dissertations contained in the NDLTD archive:



[advanced search tips](#) ▼ [how to contribute records](#) ▶



“FAIRification” process

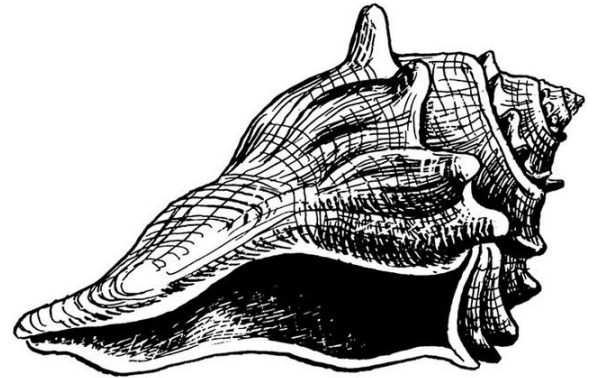


Looking to the future

shell

1. FAIRness of the dissertation ~~document~~ with its semantics and metadata
2. FAIRness of data (e.g., files, software, workflows) connected to the dissertation

```
mwwitt — -bash — 60x9
Last login: Fri Nov 1 00:14:01 on console
Lappy-Too:~ mwwitt$
```



Closing



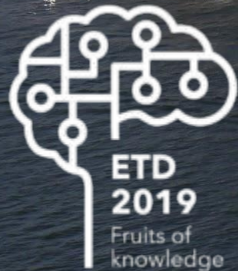
- Are the FAIR Principles applicable to ETDs?
- Could they be applied to the data, software, and other scholarly research objects that are connected to our ETDs?
- Can we conceive of our millions of ETDs together as a FAIR dataset or as a collection of FAIR datasets?
- Could use cases and new value be articulated to make the case to invest in aligning with or developing an implementation network of FAIR ETDs and/or are there other approaches to consider?

No

Maybe?

Yes

Obragado



mwitt@purdue.edu