

DTL

WHAT IS FAIR DATA?

- Findable**
- Accessible**
- Interoperable**
- Reusable**

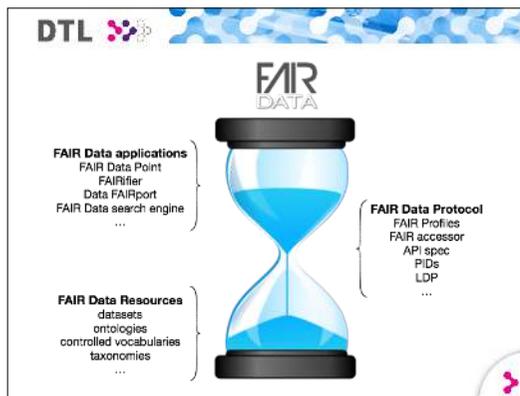
... For both humans and machines

FAIR

Humans and machines

FAIR Data aims to support existing communities in their attempts to enable valuable scientific data and knowledge to be published and used in a 'FAIR' manner.

Not include Open (but still open science)



At the meeting, much more was discussed: architecture of hourglass.

Data sources and applications connected through a narrow neck, a few necessary protocols

Compare this to how the internet was developed

DTL

FAIR METRICS

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT | SUBMIT
| ALERTS/TIPS | CONTACTS

Search []
Advanced Search

New Results

A design framework and exemplar metrics for FAIRness

World December 1, 2017

Download PDF | Share | Email | Tweet | Like it

Mark D. Wilkinson, Susanna-Assunta Sansone, Erik Scerbo, Peter Dumas, Ivan Ober, Thomas G. S. Slater, Muel D. de Meester
doi: <https://doi.org/10.1101/214499>

This article is a preprint and has not been certified by peer review [it may differ from the final published version]

Principles were deliberately vague. Everybody could agree.

Metrics make things more concrete

Everyone can self-evaluate. Evaluation turns up ideas for improvement.

And people developing this can also help.

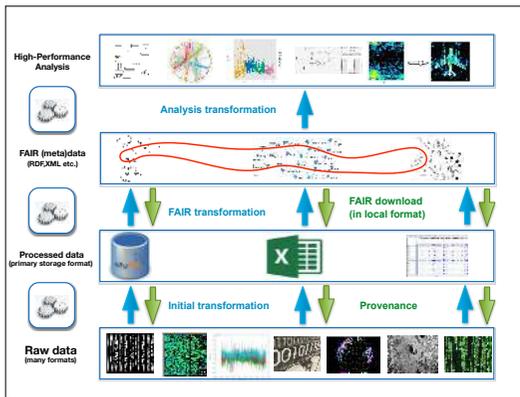
e.g. we apply this to ELIXIR core resources

DTL

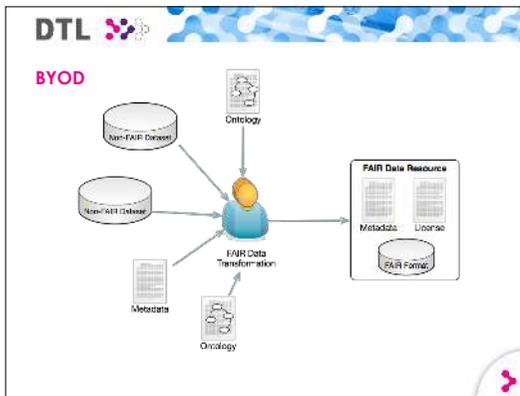
FAIR TOOLS

- BYOD – Workshops
- FAIRifier
- FAIR data point
- ORKA
- FAIR Search engine
- FAIR Data Stewardship
- FAIR Federated data analysis
- FAIR Training

Quite some practical experience by now on FAIR
 Based largely on the Linked Data assumption
 Defining the waist of the hourglass.



First: We embrace Linked Data, BUT:
 We won't store all data as RDF!
 Native (open!) formats for a lot of the "raw" data.
 Conclusions and relationships through RDF
 MAP all concepts in the raw data to Linked Data ontologies
 e.g. what does it mean if a column is labeled "blood pressure"? How

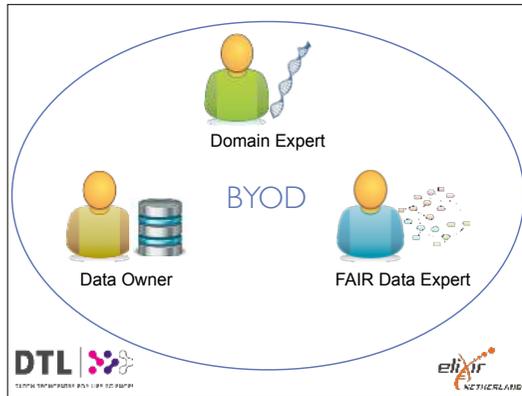


Goals:

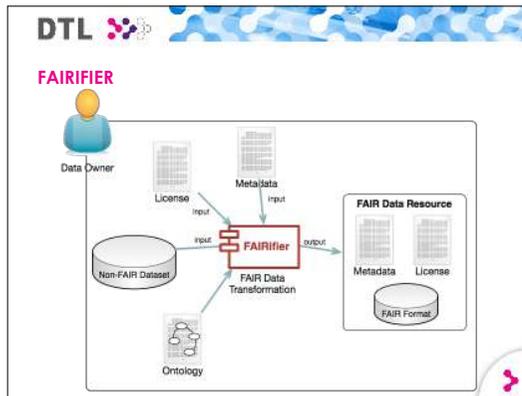
- Learn how to make data linkable "hands-on" with experts
- Create a "telling story" to demonstrate its use
- Make FAIR Data at the SOURCE [HOURGLASS requires this]

Who:

- Data owners – specialists on given datasets
- FAIR Data experts
- Domain experts

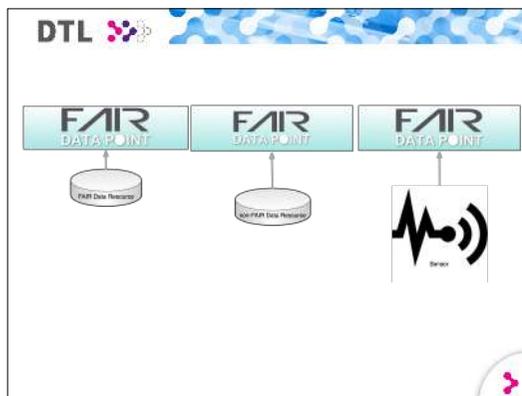


Much of the work is manual, through expertise.
 Requires all three types of expertise
 Try to find a surprisingly easy application question



Because of this we are working on a tool to automate this process as much as possible, the FAIRifier.
 Also, once a data format has been made FAIR once, it will be much easier to do this another time for another file of the same type.
 We suggest to store the "Model" with the Format. E.g. in the FAIRsharing database.

But where is the data itself going? A repository. And we worked on a platform to make those interoperable too.



FAIR Data Point: A FAIR Data System that provides access to published datasets. The datasets can be external or internal to the FAIR Data Point. Also, the source data can be a regular (non-FAIR) dataset or a FAIR Data Resource. If the source data is non-FAIR, the FAIR Data Point needs to make the necessary FAIR transformations on the fly.

DTL 

"METADATA"

Layer	Description	Example	Standard
FDP (Data repository)	Information about the FDP as a data repository	PID, title, description, license, owner, API version, etc.	OAI-PMH (extended)
Catalog	Information about the catalog of datasets offered	PID, title, description, publisher, etc.	W3C DCAT #Catalog
Dataset	Information about each of the offered datasets	AccessURL, downloadURL, format, mediaType, etc.	W3C DCAT #Dataset, #Distribution
Data record	Information about the actual data, types, identifiers, etc.	data types, domain, range, predicates, etc.	RML-Community/ domain, e.g.: DICOM, VCF,

FAIR Data Point contains standard descriptions of the data collections/data sets /versions /multiple different formats

- * Use existing standards, bring them together in a REST interface.
- * This makes it easy to make many data platforms into FDPs!

Metadata is part of the data here. How can someone else annotate



ORKA = a Prototype

Make use of Open world Assumption of RDF

Extrinsic metadata: opinions and corrections by others

Important to annotate reuse and reusability



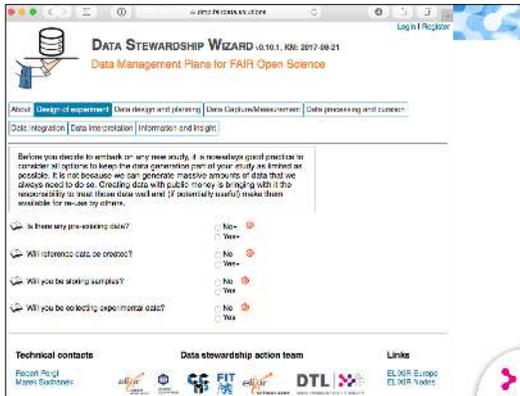
Aggregate search engine over the metadata

Purpose is to showcase, not to be the "google" yet.

And even google didn't start by reading PDFs.

This is just another application eating FAIR data.

The hourglass makes it possible for others to make added-value applications, just like on the web. And to sell those.



Switching to a different branch of FAIR application.

*My own hobby: DMP that serves the researcher

* decisions during a research project to make data as FAIR as possible.

1. Workflow to make DMP for researcher that wants to do it well, be aware of all dangers:



Where it is not legally possible or not practical to collect data in one place

Make data available as FAIR data points. Common computing environment to work with that data.

*Analyse the data in-place



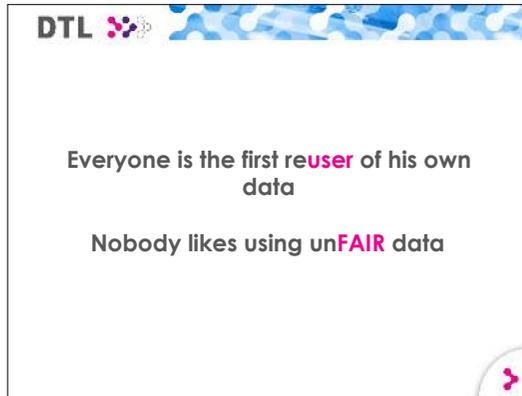
Spreading all this knowledge.

NFU D4LS WP9: Bring together expert desks // DTL broadens this to all partners

Train researchers- FAIR Data Management e.g. ZonMW

Train bioinformaticians – With BioSB research school

Train trainers – Capacity building



Data management and planning is often perceived as burden

But: Making FAIR data is not only for others, and because we have to.

First Reuser: tomorrow (you forgot what you did), next month (the reviewer asks), and next year (your postdoc leaves)

There is a lot everyone already can do to make their data more fair

But: Everyone is still waiting for more tools to make this easier (e.g. to keep provenance automatically, Electronic Lab notebook)

Lets build a FAIR world together
