



health RI

enabling data driven health

Introductie FAIR data

Jan-Willem Boiten; 9 maart 2022

Met veel slides van Rob Hooft

 janwillem.boiten@health-ri.nl



De FAIR principes – het begon met een workshop in Leiden

Lorentz center
Jointly Designing a Data FAIRPORT
Workshop: 13 - 16 January 2014, Leiden, the Netherlands

Scientific Organizers: Scott Lusher, HLW&C Amsterdam, Barend Mons, Leiden UMC

Topics:
• Towards a Modular Blueprint "Floor plan" of a Safe and Fair Data Stewardship, Trading and Routing Environment
• A Public Private Partnership to Ensure Long Term Solutions for Data in the eScience Era.

www.lorentzcenter.nl

open data
is about
MORE
THAN
DISCLOSURE
it must be
"Fair"

- Findable
- Accessible
- Interoperable
- Reusable

scientific **data**

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [scientific data](#) > [comment](#) > [article](#)

Open Access | Published: 15 March 2016

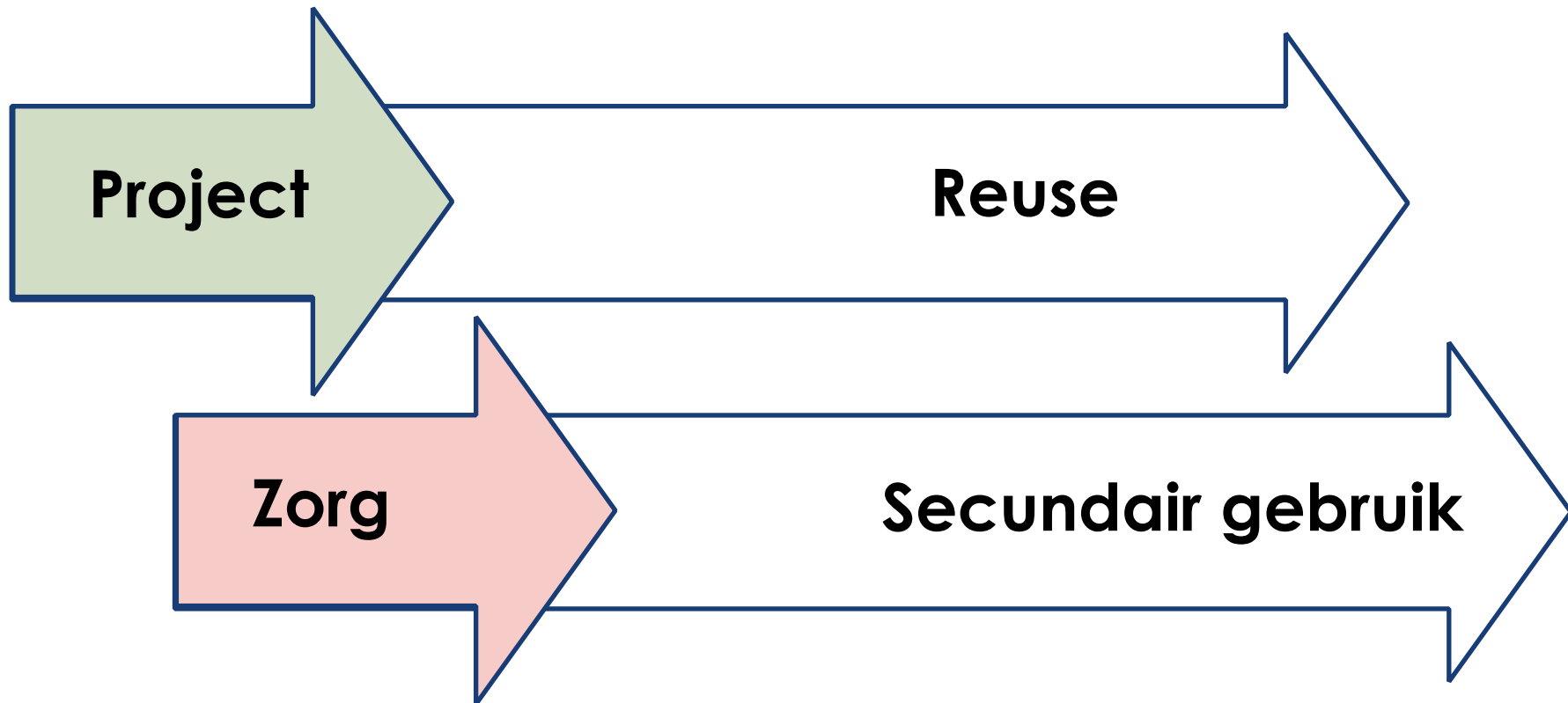
The FAIR Guiding Principles for scientific data management and stewardship

[Mark D. Wilkinson](#), [Michel Dumontier](#), ... [Barend Mons](#) [+ Show authors](#)

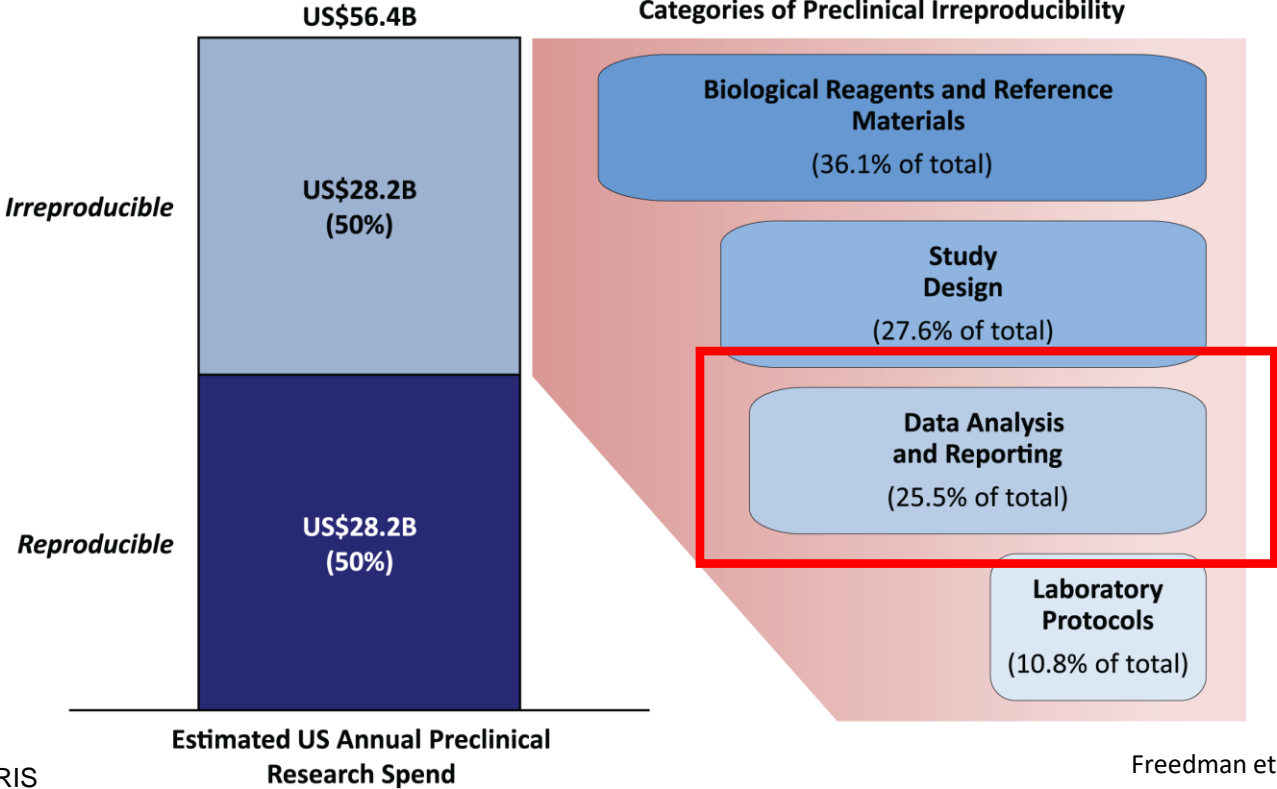
[Scientific Data](#) **3**, Article number: 160018 (2016) | [Cite this article](#)

428k Accesses | 3642 Citations | 1937 Altmetric | [Metrics](#)

Data is van oudsher beperkt in hergebruik



FAIR data – ook essentieel voor reproduceerbaarheid



Slide: Anton Ussi; EATRIS

Freedman et al PLOS Biology (2015)

75 Years of Astonishing Evolution of IT: 1946–2021

George Strawn , US National Academy of Sciences, Engineering, and Medicine

prescription not an implementation. As mentioned in the previous paragraph, currently data is “munged” manually for analysis rather than by automated processes. This results in data preparation/munging taking 80% of the time leaving only 20% for analytics (science or business). I hope/believe that in the decade of the 2020s, we will learn how to automate the process of making data FAIR, which will go a long way to making data *Fully AI-Ready* much more rapidly.

<https://doi.org/10.1109/MITP.2021.3070963>



I would like to exploit common genotype-phenotype relations between Alzheimer's Disease and Huntington's Disease...
I need to combine AD and HD data...

I can help with that!



Source: Marco Roos



???

DOES NOT COMPUTE



Here's my data, have fun!

米当局は、あなたの国籍故に、在米日本領事代表にあなたが逮捕又は拘禁されたことを通報する必要があります。領事官は通報を受けた後、あなたに電話を掛けたり、あるいはあなたを訪門することができます。あなたは領事官の援助を受ける必要はありませんが、あなたがあなたの家族との連絡、身柄の安全を確保する必要がある場合は、領事官に通報します。

Очи чёрные, очи жгучие,
Очи страстные и прекрасные,
Как люблю я вас, как боюсь я вас,
Знать увидел вас я не в добрый час.

Очи чёрные, очи пламенные
И моят они в страны дальные,
Где царит любовь, где царит покой,
Где страдания нет, где вражды запрет.

Here's my data, have fun!



Source: Marco Roos



???

DOES NOT COMPUTE



Here
data, I
fun

It is very hard to make data FAIR/interoperable afterwards:

We need to use data standards from the start

We need to supply sufficient meta data for others to interpret our data

We need to be FAIR-by-design

's my
have
n!



И менят они в страны дальные,
Где царит любовь, где царит покой,
Где страдания нет, где вражды запрет.



Source: Marco Roos

Findable

- Choose Repository
 - Domain specific
 - Institutional
 - National
 - Special
- Make sure to get and use Persistent Identifiers
- Describe what the data **is**, not only why you collected it
- Register the data in a catalogue, where others would look for it

Accessible

- Assure Longevity
- Check Legal Conditions
- Limit the Embargo
- Use standard protocols also for machine actionability
- Ensure proper ICT procedures

Interoperable

- Terminology
- Format
- Sufficient metadata

Brussels ← → Bruxelles
Cancer ← → Malignant Neoplasm

Reusable

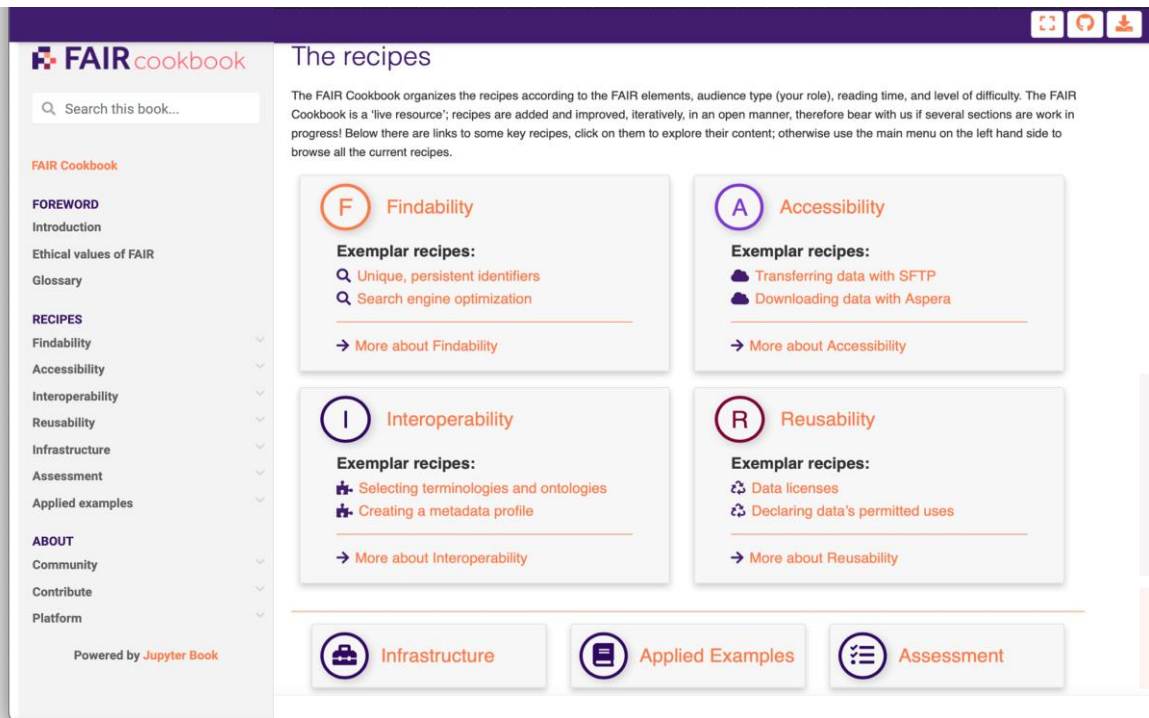
- Document provenance
- Use *minimal metadata* standards
- Choose a liberal license
- Make sure results are not only narrated

FAIRplus Dataset Maturity Levels

5	Managed Data Assets	Enterprise Level. Data at this level is optimally managed at the most granular level in an environment offering <i>data governance, master data management and reference data management</i> capabilities.
4	Semantically Typed Data	Cross-community Level. This level focuses on cross-domain interoperability and is meant to be the level required for larger harmonization and integration projects.
3	Standardised Data	Community Level. Data at this level complies with community standard domain models, terminologies and formats, and is hosted in an environment offering searching and retrieval capabilities.
2	Described Data	Project Level. All datasets generated within a project are consistently described against a locally defined schema, controlled terminologies, and hosted in an environment offering data catalogue level searching capabilities.
1	Identifiable Data	Data Object level. Data at this level is identifiable as individual generic data objects and described by generic metadata elements. Hosting environment offers limited retrieval capabilities.
0	Single Use Data	No potential for re-use beyond lifetime of the research project

Model - Ibrahim Emam

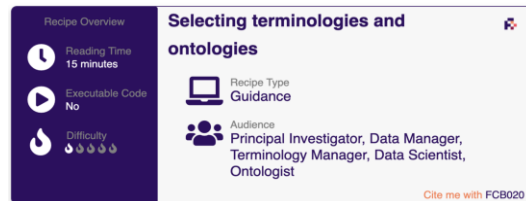
FAIR cookbook: from knowledge to recipes



The screenshot shows the FAIR Cookbook website. At the top, there's a search bar and navigation icons. The main content area is titled 'The recipes' and contains a description of the cookbook's purpose. Below this, there are four main categories: Findability (F), Accessibility (A), Interoperability (I), and Reusability (R). Each category has a list of exemplar recipes and a link to 'More about' that category. At the bottom, there are three additional categories: Infrastructure, Applied Examples, and Assessment. A sidebar on the left contains a navigation menu with sections for 'FAIR Cookbook', 'FOREWORD', 'RECIPES', and 'ABOUT'. The footer of the sidebar mentions 'Powered by Jupyter Book'.

URL: fairplus.github.io/the-fair-cookbook

Examples:



The screenshot shows a 'Recipe Overview' card for 'Selecting terminologies and ontologies'. It includes a 'Reading Time' of 15 minutes, 'Executable Code' status as 'No', and 'Difficulty' level as 3 out of 5. The 'Audience' is listed as 'Principal Investigator, Data Manager, Terminology Manager, Data Scientist, Ontologist'. A 'Cite me with FCB020' link is provided at the bottom right.

Authors:



.....and open to other contributions,
because FAIR is a 'team sport'!
fairplus-cookbook@elixir-europe.org

Tot slot....

FAIR \neq Open Science