https://www.youtube.com/watch?v=N2zK3sAtr-4

But your work is in PubMed Central and was funded by NIH.
That is true!
FAIR Analysis: Data Sharing & Management Snafu

- **Brown Bear** performed research that includes data on B Cell Function
  - published in *Science*
  - listed in PubMed Central
  - funded by NIH
FAIR Analysis: Data Sharing & Management Snafu

- **Brown Bear** performed research that includes data on B Cell Function
  - published in *Science*
  - listed in PubMed Central
  - funded by NIH

**Data Sharing Policy at Science**
http://www.sciencemag.org/authors/science-editorial-policies

**Data and Materials Availability after Publication**
After publication, **all data** and materials necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of Science. **All computer codes** involved in the creation or analysis of data must also be available to any reader of Science. After publication, **all reasonable requests** for data or materials must be fulfilled.
• Brown Bear performed research that includes data on B Cell Function
  • published in Science
  • listed in PubMed Central
  • funded by NIH

Data Sharing Policy at PubMed Central
https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/

Supplementary Data
Any supplementary data (images, tables, video, or other documents/files) that are associated with an article must be deposited in PMC with the article. This applies to supplementary data that are available at the journal’s site as well as those that may be stored in a public repository. An exception may be made for data sets that require custom software to read/use, or are very large (over 2 GB).
FAIR Analysis: Data Sharing & Management Snafu

- **Brown Bear** performed research that includes data on B Cell Function
  - published in *Science*
  - listed in PubMed Central
  - funded by NIH

---

Data Sharing Policy at NIH

[https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/](https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/)

**NIH Data Sharing Policy and Implementation Guidance**

In NIH's view, all data should be considered for data sharing. *Data should be made as widely and freely available* as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data. To facilitate data sharing, investigators submitting a research application requesting $500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.
**FAIR Analysis: Data Sharing & Management Snafu**

- **Panda Bear** reads about research in *Science* (F)
- **PB** requests a copy of the data (A)
- **BB** assumes (wishes) the data are already in the *Science* article *(assumes data are FAIR)*
- **PB** indicates that the data are not there *(they are not FAIR)*
- **BB** is not sure where the data are (a USB drive... in a box at home... so many boxes) *(F, A, I)*
- **BB** 7 months to locate *(F)* sends only copy to **PB** *(risking A)*
- **BB** has ‘complied’ with the request for data but...
  - layers of idiosyncratic coding *(specialized software is required to decode the hexadecimal notation)* *(I)*
  - software not supported since 2007 *(F, A, R)*
  - Semantically empty field names *(SAM1, SAM2, ...)* with no documentation *(not even in the *Science* publication)* *(I, R)*
- Personnel point of failure: graduate student who moves back to China *(I, R)*
Conclusion:
Despite the Data Policies of *Science*, PubMed Central and NIH, after 7 months of frustration and uncertainties, Panda Bear still cannot reuse Brown Bear’s data.
The 15 FAIR Data Principles
& some tools for better data stewardship

Erik Schultes, PhD
erik.schultes@dtls.nl

Dutch TechCenter for Life Science
Leiden University Medical Center
Leiden Center for Data Science
Data integration allows us to...

1. ask scientific questions beyond the scope of individual laboratories.
2. do new science without making new data.
So.... What is the Problem?
So.... What is the Problem?

Data Overload
Data Reuse
Data Reproducibility
Data Overload

- 21 million references (abstracts) to journal articles (1946 - present)
- 5,600 journals
- More than 700,000 references added in 2013
- 3,000 references are added each day (125 / hour)
Data Overload

TED Ideas worth spreading

Meet e-Patient Dave

Born
Jack Thomas Andraka
January 8, 1997 (age 20)
Crownsville, Maryland, U.S.

Residence
United States

Nationality
American

Fields
Cancer research, medical research, invention
Data Overload

Computer performance doubles every 18 months

DNA sequencing data doubles every 6 months
Data Reuse

As research articles age, the odds of their raw data being extant drop dramatically.

Nature news, 19 December 2013

EUDAT Summer School, 3-7 July 2017, Crete
ARS TECHNICA LIVE —

The US government is removing scientific data from the Internet

At Ars Technica Live, we talked to Lindsey Dillon, who decided to do something about it.

ANNALEE NEWITZ - 7/1/2017, 2:41 AM
The US government is removing scientific data from the Internet

In the first report of its kind, analysis of the Environmental Protection Agency (EPA) by Trump and Administrator Scott Pruitt...
How sloppy science creates worthless cures and wastes billions

New book explains everything that’s going wrong, why it matters, and what to do.

DIANA GITIG - 4/1/2017, 8:00 AM
Reproducibility Crisis
Reproducibility Crisis

Over half of psychology studies fail reproducibility test

Largest replication study to date casts doubt on many published positive results.

Monya Baker

27 August 2015

Don’t trust everything you read in the psychology literature. In fact, two thirds of it should probably be distrusted.

In the biggest project of its kind, Brian Nosek, a social psychologist and head of the Center for Open Science in Charlottesville, Virginia, and 269 co-authors repeated work reported in 98 original papers from three psychology journals, to see if they independently came up with the same results.

Brian Nosek’s team set out to replicate scores of
Reproducibility Crisis

Cancer reproducibility project releases first results

An open-science effort to replicate dozens of cancer-biology studies is off to a confusing start.

Monya Baker & Elie Dolgin

18 January 2017
Data Stewardship
Data Stewardship

A plan for maximizing the re-use of data beyond the life of the project.
A plan for maximizing the re-use of data beyond the life of the project.
Data Stewardship

- exposing versus publishing data
- open versus closed data
- licensing and fees
- consent, authorship, ownership (e.g. patient privacy)
- who paid to create the data?
- who will pay for the persistence of the data?
- persistence policy: serving (e.g. 24/7), backup, archiving
- versioning
- big data scaling: storage, compute, serving
- machine interoperation (ontology engineering, data modeling)
- human interoperation (24 EU languages)
What is... FAIR?

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;

doi: 10.1038/sdata.2016.18  EUDAT Summer School, 3-7 July 2017, Crete
The FAIR Data Principles explained

These webpages provide an actionable list of the 15 FAIR Data Principles as a simple guide when publishing data. For each principle, we give a basic definition, examples, and links to useful resources. We hope that by working through the list, anyone wishing to maximize the reusability of their data, can prioritize their efforts and make more informed choices regarding a suitable repository. We hope that this list will also focus the growing public discourse around FAIR: what is FAIR exactly, and what is it not.

Findable: Data and metadata are easy to find by both humans and computers. Machine readable metadata is essential for automatic discovery of relevant datasets and services, and for this reason are essential to the FAIRification process.

- F1: (meta) data are assigned globally unique and persistent identifiers
- 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: Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

- A1: (metadata are retrievable by their identifier using a standardized communication protocol.
- A1.1: The protocol is open, free and universally implementable
- A1.2: The protocol allows for an authentication and authorization when required
- A2: Metadata should be accessible even when the data is no longer available

Interoperable: The computer can interpret the data, so that they can be
What FAIR is not...

- FAIR is not a standard
- FAIR is not equal to ‘Semantic Web’
- FAIR is not equal to ‘Open’ or ‘Free’
- FAIR is not for humans only

doi: 10.3233/ISU-170824
What FAIR is not...

FAIR is not a standard
FAIR is not equal to ‘Semantic Web’
FAIR is not equal to ‘Open’ or ‘Free’
FAIR is not for humans only

Data is Open but not FAIR

doi: 10.3233/ISU-170824
The Idea
The Idea
The Idea
The Idea

EOSC
5%
500k

Data Commons

Scientific Organizers
- Scott Lusher, NLeSC 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.
The Idea

EOSC
5%
500k

Data Commons

G20 2016
China
Europe joins forces to create largest ever shared data repository for researchers

14 June 2017

by Benedict O’Donnell
The European Open Science Cloud (EOSC) explained by the ZBW
EOSC: From Vision To Action - EOSC Summit 12 June 2017

The EOSC Summit took place in Brussels, on the 12th of June, collecting 120 invited experts and key players in the European Open Science Cloud. The Summit was the “moment of commitment to the EOSC”, as clearly highlighted by director-general Robert-Jan Smits during the day.

Carlos Moedas, European Commissioner for Research, Science and Innovation, introduced the event with an inspiring explanation of its visionary idea on the role of the EOSC as 'The New Republic of Letters'.

Read all here <<

A final, explanatory, quote from Voltaire ended perfectly Moedas' speech, showing all of us that the process of knowledge-sharing is all but not a recent matter:


EUDAT Summer School, 3-7 July 2017, Crete
110 European stakeholders of EOSC implementation
80 scientific fields
13 research funders
15 national scientific infrastructures
23 officials from ministries of Member States
1800 watched the web stream of the event (from 54 countries)
second most popular Twitter topic in Belgium during the day: #EOSC
third most popular Twitter topic in Belgium during the day: #EOSC


EUDAT Summer School, 3-7 July 2017, Crete
SESSION 1: DATA CULTURE, DATA STEWARDSHIP, PRACTICAL AND POLICY TOOLS

INPUT PAPER

European Open Science Cloud (EOSC) Summit
12 June 2017 | Charlemagne Building | Brussels

RECOGNISING the challenges of data driven research in pursuing excellent science;
GRANTING that the vision of European Open Science Cloud is of a research data commons, widely inclusive of all disciplines and Member States, sustainable in the long-term, the EC;
PROPOSES that the stakeholders present at this summit commit to share the following intents and will actively support their implementation in the respective capacities:

- **[Common culture of data stewardship]**: In recognition of the long-term reuse value for science and for innovation of much data created by research activities, European science must be grounded in a common culture of data stewardship, so that research data is recognised as a significant output of research and is appropriately curated throughout the research lifecycle.

- **[Open access by-default]**: Open access is the default setting for all results of publically funded research in Europe; this means that publicly funded research data should be as open as possible...
[Common culture of data stewardship]: In recognition of the long-term reuse value for science and for innovation of much data created by research activities, European science must be grounded in a common culture of data stewardship, so that research data is recognised as a significant output of research and is appropriately curated throughout the research lifecycle.

[Open access by-default]: Open access is the default setting for all results of publicly funded research in Europe: this means that publicly-funded research data should be as open as possible and as closed as necessary (allowing for proportionate limitations only in duly justified cases of personal data protection, IPR concerns, national security or similar).

[FAIR data]: In order to have the greatest utility for science and innovation, data produced by European publicly-funded research should be FAIR: Findable, Accessible, Interoperable and Reusable. These attributes will guide the stewardship of European research data.

[Reward and citation system]: Researchers that make research data open and FAIR should be rewarded for this; FAIR data management should become a criterion both in the career assessment of researchers and for the evaluation of projects (initial funding and review of performance and impact). A data citation system should be put in place that rewards the provision of excellent open data. This will assist both the assessment of researchers and their projects, and help implement the findability, accessibility, interoperability and reusability ("FAIRness") of data.

[User needs]: Researchers need to have effective support from their host institutions, national and European infrastructures (human and technical) for the tasks of data stewardship for Open and FAIR data. Institutions in particular, should be encouraged and incentivised to ensure that systems and processes are in place to support lifecycle data stewardship in European research projects.

[Data Management Plan]: A key element of good data management is a Data Management Plan (DMP); the use of DMPs should become obligatory in all research projects generating or collecting data, based on online tools conforming to common methodologies.

[Minimal standards]: All repositories for research data should conform to minimal standards of interoperability (technical, semantic, legal and organisational) and stewardship (e.g. Data Seal of Approval).
GO FAIR – Ready for Take-Off

At the Competitiveness Council at the end of May, Germany and the Netherlands made their Joint Position Paper on the European Open Science Cloud (EOSC) public and presented their approach to boosting the EOSC through the GO FAIR initiative.

Leibniz Information Centre for Economics
The EOSCpilot measures will contribute to establishing the overall governance framework for the EOSC, whereas GO FAIR will focus on early implementation needs of existing networks and consortia.

GO FAIR – Ready for Take-Off

At the Competitiveness Council at the end of May, Germany and the Netherlands made their Joint Position Paper on the European Open Science Cloud (EOSC) public and presented their approach to boosting the EOSC through the GO FAIR initiative.

Leibniz Information Centre for Economics
Germany and the Netherlands call for rapid action on the European Open Science Cloud

News item | 30-05-2017 | 19:32

At the occasion of today’s Competitiveness Council, Germany and the Netherlands made clear that it is important to boost the development of the European Open Science Cloud (EOSC) and capitalise on the momentum of the digital era. ‘Time for action is now,’ say State Secretaries Georg Schütte (Germany) and Sander Dekker (the Netherlands) in their position paper on the EOSC that was presented during today’s Council meeting in Brussels.

Research data should not be stored away on personal computers or USB-sticks, nor in research infrastructures only researchers themselves know how to use. Making data easily accessible for other researchers will enhance scientific progress. Making them accessible for a broader audience such as citizens and entrepreneurs will greatly boost the impact and utilisation of science. The sooner we are able to make this happen, the better.

GO FAIR

Schütte and Dekker propose to support the GO FAIR initiative, as a promising approach towards establishing the EOSC. GO FAIR is completely open-to-all and can contribute to a broad involvement of the European science community as a whole. They called on other Member States to join the movement and urged the European Commission to strengthen its efforts through proposing appropriate governance and funding frameworks for the realisation of the EOSC. Finally, they called on the research and e-infrastructure communities throughout Europe to join GO FAIR and participate in shaping its future.
An Implementation initiative towards the Internet of FAIR Data and Services

https://www.dtls.nl/go-fair/
The Technology

Jointly Designing a Data FAIRPORT

Workshop: 13 - 16 January 2014, Leiden, the Netherlands

Scientific Organizers
- Scott Lusher, NLeSC 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.

The Lorentz Center is an international center in the sciences. Its aim is to organize workshops for scientists in an atmosphere that fosters collaborative work, discussions, and interactions. For registration see: www.lorentzcenter.nl

www.lorentzcenter.nl
The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier […] Barend Mons

Abstract

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.
The Technology
EUDAT Summer School, 3-7 July 2017, Crete
EUDAT Summer School, 3-7 July 2017, Crete
EUDAT Summer School, 3-7 July 2017, Crete

FDP
- title
- language
- identifier
- label
- description
- license
- version
- publisher
- content
- API version

Catalog
- label
- title
- identifier
- language
- version

Dataset
- title
- label
- identifier
- language
- version
- description
- keywords
- landingpage
- publisher
- theme
- creator

Distribution
- title
- label
- identifier
- version
- license
- access
- url
- mediatype
<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>FDP of lorentz.fair-dtls.surf-hosted.nl</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metadata ID</strong></td>
<td>3e77134d-9338-482a-a4f7-8e6933b47469</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>FDP of lorentz.fair-dtls.surf-hosted.nl</td>
</tr>
<tr>
<td><strong>Issued</strong></td>
<td>2017-05-12T12:11:36.343Z</td>
</tr>
<tr>
<td><strong>Modified</strong></td>
<td>2017-05-12T12:11:36.355Z</td>
</tr>
<tr>
<td><strong>License</strong></td>
<td>license</td>
</tr>
<tr>
<td><strong>Catalogs</strong></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/wur">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/wur</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/DvBC1">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/DvBC1</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/PKD_2_0">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/PKD_2_0</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/AGIS">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/AGIS</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/PKD_1_0">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/PKD_1_0</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/test_example_number_2_2">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/test_example_number_2_2</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/post_example_2.0">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/post_example_2.0</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/cds-v1">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/cds-v1</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/post-test_1.0">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/post-test_1.0</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/testcatalog_1.0">https://lorentz.fair-dtls.surf-hosted.nl/fdp/catalog/testcatalog_1.0</a></td>
</tr>
</tbody>
</table>
inhibitors  p65
ReLA/NFkB p65 Inhibitors: Novus Biologicals
https://www.novusbio.com/inhibitors/rela-nfkbp65
ReLA/NFkB p65 Inhibitors available through Novus Biologicals. Browse our ReLA/ NFkB p65 Inhibitor catalog backed by our Guarantee+.

Inhibiting NF-κB Activation by Small Molecules As a Therapeutic ...
by SC Gupta - 2010 - Cited by 345 - Related articles
May 21, 2010 - pathway (Fig 3). Table 1. A list of small molecules as inhibitors of NF-κB pathway. Blocking NF-κB activation by inhibitors of p65 acetylation.

Suppression of p65 phosphorylation coincides with inhibition of ... - NCBI
by J Hu - 2005 - Cited by 26 - Related articles

Identification of a p65 peptide that selectively inhibits NF-kappa B ...
by Y Takada - 2004 - Cited by 157 - Related articles
Jan 7, 2004 - Identification of a p65 peptide that selectively inhibits NF-kappa B activation induced by various inflammatory stimuli and its role in ...

NF-κB - Wikipedia
https://en.wikipedia.org/wiki/NF-%CE%9B
NF-κB is a protein complex that controls transcription of DNA, cytokine production and cell .... Concerning known protein inhibitors of NF-κB activity, one of them is IFRD1, which represses the activity of NF-κB p65 by enhancing the HDAC-mediated deacetylation of the p65 subunit at lysine 310, by favoring the recruitment of ...
Find me all known low molecular weight inhibitors of the Human p65 Protein. Separate the list based on those that were found in curated databases, from those that were found in self-deposited data archives. Keep track of the license and citation information for each one. If data is relevant, but not public, please provide the contact information for the person I need so I can request the data.
Awareness

- Problems & Responsibilities
- What is data stewardship?
- What is FAIR?
- Target: domain experts

Skills

- BYOD
- FAIR Hackathons
- Metricthons
- Target: certified professionals

500k
Training

Awareness
- Problems & Responsibilities
- What is data stewardship?
- What is FAIR?
- Target: domain experts

Skills
- BYOD
- FAIR Hackathons
- Metricthons
- Target: certified professionals
Awareness

- Problems & Responsibilities
- What is data stewardship?
- What is FAIR?
- Target: domain experts

Skills

- BYOD
- FAIR Hackathons
- Metricthons
- Target: certified professionals

- half-day (seminars)
- 1-day (workshops)
- 2-day
- 3-day (hackathons)
- 5-day (summer school)
Jointly Designing a Data FAIRPORT

Workshop: 13 – 16 January 2014, Leiden, the Netherlands

Scientific Organizers
- Scott Lusher, NLeSC 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.

Applications
### How well does tranSMART match the FDP specification?

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Re-usable</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1. (meta)data are assigned a globally unique and persistent identifier</td>
<td>A1. (meta)data are retrievable by their identifier using a standardized communications protocol</td>
<td>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</td>
<td>R1. meta(data) are richly described with a plurality of accurate and relevant attributes</td>
</tr>
<tr>
<td>F2. data are described with rich metadata (defined by R1 below)</td>
<td>A1.1 the protocol is open, free, and universally implementable</td>
<td>I2. (meta)data use vocabularies that follow FAIR principles</td>
<td>R1.1. (meta)data are released with a clear and accessible data usage license</td>
</tr>
<tr>
<td>F3. metadata clearly and explicitly include the identifier of the data it describes</td>
<td>A1.2 the protocol allows for an authentication and authorization procedure, where necessary</td>
<td>I3. (meta)data include qualified references to other (meta)data</td>
<td>R1.2. (meta)data are associated with detailed provenance</td>
</tr>
<tr>
<td>F4. (meta)data are registered or indexed in a searchable resource</td>
<td>A2. metadata are accessible, even when the data are no longer available</td>
<td></td>
<td>R1.3. (meta)data meet domain-relevant community standards</td>
</tr>
</tbody>
</table>

How well does tranSMART match the FDP specification?
### tranSMART FAIR data point connector.

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Re-usable</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1. (meta)data are assigned a globally unique and persistent identifier</td>
<td>A1. (meta)data are retrievable by their identifier using a standardized communications protocol</td>
<td>11. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation</td>
<td>R1. meta(data) are richly described with a plurality of accurate and relevant attributes</td>
</tr>
<tr>
<td>F2. data are described with rich metadata (defined by R1 below)</td>
<td>A1.1 the protocol is open, free, and universally implementable</td>
<td>12. (meta)data use vocabularies that follow FAIR principles</td>
<td>R1.1. (meta)data are released with a clear and accessible data usage license</td>
</tr>
<tr>
<td>F3. metadata clearly and explicitly include the identifier of the data it describes</td>
<td>A1.2 the protocol allows for an authentication and authorization procedure, where necessary</td>
<td>13. (meta)data include qualified references to other (meta)data</td>
<td>R1.2. (meta)data are associated with detailed provenance</td>
</tr>
<tr>
<td>F4. (meta)data are registered or indexed in a searchable resource</td>
<td>A2. metadata are accessible, even when the data are no longer available</td>
<td></td>
<td>R1.3. (meta)data meet domain-relevant community standards</td>
</tr>
</tbody>
</table>
GO FAIR Metrics Group

How FAIR is your data?

The GO FAIR Metrics Group is collaborating with a broad set of stakeholders to define metrics enabling both qualitative and quantitative assessment of the degree to which online resources comply with the 15 Principles of FAIR Data as they were originally stated in the "The FAIR Guiding Principles for scientific data management and stewardship".

Founding Members

- Mark Wilkinson, Universidad Politécnica de Madrid
- Susanna Sansone, University of Oxford
- Michel Dumontier, Maastricht University
- Peter Doorn, DANS
- Luiz Olavo Bonino, VU/DTL
- Erik Schultes, DTL

Next meeting
Leiden, The Netherlands, June 3-4, 2017
FAIR Metrics

Findable
- F1: (meta)data are assigned globally unique and persistent identifiers
- 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 communication protocol.
- A1.1: The protocol is open, free and universally implementable
- A1.2: The protocol allows for an authentication and authorization when required
- A2: Metadata should be accessible even when the data is 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 the 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
Data Stewardship plans

Data cycle step 1:

1/30. (1.12.1) What will the IP be existing data

1/16. (1.10) Do you need to

1/3. (1.3) Will you use reference data from literature?

1/53. (del) What is the volume of the level changed?

1/15. (del) What format is it in?

1/13. (del) Will any usage of it? Or use it via the network?

1/7. Which version will you use?

64. (1.15) Are there any data format changes?

57. (del) Which database will you use?

42. (del) Does Result information go reversed?

37. Will you Define ways to detect Reversible?

25. Are all metadata that is in the file/sample swaps, e.g. by different database?

29. Do you need frequent access to disc? Online or in Fire/disaster proof?

24. Will you be running a bulk/routine analysis?

24. Is it more than 10TB?

23. How will you work with your computer time) when a file system of Cartesius crashed.

22. Use shared infrastructure?

21. Will project partners use the external (super)computer system you use?

20. Use persistent identifiers for e.g. batch e. g. from EUDAT.

19. Summer School, 3-7 July 2017, Crete

18. Will all metadata that is in the lab or external hard drives need to glue workflows across the lab?

17. Will all data be collected in the proper metadata?

16. Will your data be changing over time if the temporary data does not survived-it accidentally delete a file before it

15. Ease of use?

14. Can "workflow decay" be used.

13. Can workflows be run remotely?

12. Can workflows be part of a tool used by Biologists?

11. Can "modeling" be part of a tool used by Chipster?

10. Chipster - will be used by Biologists?

9. Will data formats be upgraded if needed to new place after the project has ends you need?

8. Will you publish also if the data they need on their own is completely open?

7. Will you be adding your data to a knowledge base?

6. Take pictures of the equipment actually have been filled in?

5. Can workflows be part of a tool used by Biologists?

4. Will you be building pathways?

3. Is special care needed to get a lightpath.

2. Will you do Systems biology from the measurement? For the data they need on their own are developers able to take the data they need on their own?

1. Who will do the measurements?
Data Stewardship plans

Design of experiments data (including Other People’s about samples be stored? Biobank 1/30. (1.12.1) What will the IP be 20. (1.11) What/how/who will needs to be made computer samples be added to an existing collection?

(1.11.1) Will you need to add (del) Is “harmonizing” enough, or do LS1/46. How will you check whether What is the procedure followed?

1/4 (1.4) Do you know where the reference data is available, represented in relevant Catalogues to the FAIR principles?

37. (del) Does the biobank adhere version or export updates?

1/10. (del) No? Will you need to downloading?

51. (del) Select the proper time changed?

77. (1.16.3.1) Is the collected data cient for storage?

71. (del) Do images need to be Downloading?

Data Design and planning

2/2 (2.2) Will you be using new your data format?

2/9 Which data type registries will possible

(1.18.4.2) Shared ownership?

2/45 Will the data be downloading?

2/44 How long does it need to be kept?

Will your data be changing over time?

2/65 How available must the Do your people need training for your data?

2/5 Is the Risk of information loss / your ethical committee?

2/6 Compute Capacity Planning

2/81 Project members have been actually have been filled in?

Data cycle step 6: Project members have been actually have been filled in?

2/24 Keep relations between data Can a workflow be edited different modality?

2/47 How will you work with your can be restored?

2/8 Project members have been actually have been filled in?

2/48 How long will you keep your data openable?

2/32 Will it be stored on Tape or Fire/disaster proof?

2/33 are archives stored in a remote different modality?

2/41 How: Keep sample list under different modality?

2/40 Permanent identifiers for e.g. Data stored in their own labs reason?

2/53 Is the Risk of information loss / data?

2/47 How will you work with your can be restored?

2/81 Project members have been actually have been filled in?

Data Processing and Curation

51. (del) Select the proper time changed?

2/45 Will the data be downloading?

2/44 How long does it need to be kept?

Will your data be changing over time?

2/65 How available must the Data stored in their own labs reason?

2/5 Is the Risk of information loss / your ethical committee?

2/6 Compute Capacity Planning

2/81 Project members have been actually have been filled in?

Data cycle step 7:

Significance

2/65 How available must the Data stored in their own labs reason?

2/5 Is the Risk of information loss / your ethical committee?

2/6 Compute Capacity Planning

2/81 Project members have been actually have been filled in?

Data cycle step 5:

Connect to text mining data?

6/1 Will this step need significant

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?

5/2 Will you be using common or

1/36 Do a search for data that have run (G M4)

6/2 Will you be doing structure

6/36 Consideration of the versions of the tools

6/25 Will the data cycle step 6: Data cycle step 7: Connect to text mining data?
FAIR DATA KNOWLEDGE & EXPERTISE

FAIR Data
- Linked Data 101
- Webinar on FAIR Data stewardship and BYODS
- Interview with Barend Mons about FAIR Data stewardship
- FAIRification process – VCF/OpenRefine Tutorial
- Data set voting

Technology
- Create: FAIRifier
  - Walkthrough
- Meta data editor
  - Walkthrough
- Publish: FAIR Data point
  - Example FAIR Data points
  - API
  - GUI
  - Walkthrough
- Find: FAIR Search Engine
  - Walkthrough
- Annotate: ORKA (Open RDF Knowledge annotator)
  - Demo
- Data FAIR port

Data management
- Wizard: FAIR Data stewardship tool (test version)
  - GitHub Repository
- DMP Online
- Data4biosciences generic guidelines document HANDS

Under development
- Developers Kit: Super Library
- Model Registry (Open Data Models)
- Registry of Data Stewardship Plans
- FAIR principles tutorial
- FAIR check list / metrics

Expertise
- FAIR Data: Luiz Bonino
- BYODs and hackathons: Mascha Jansen
- FAIR Data training: Erik Schultes and Celia van Gelder
- DSP Web-Wizard: Robert Pergl and Rob Hooft

Fini

www.eudat.eu