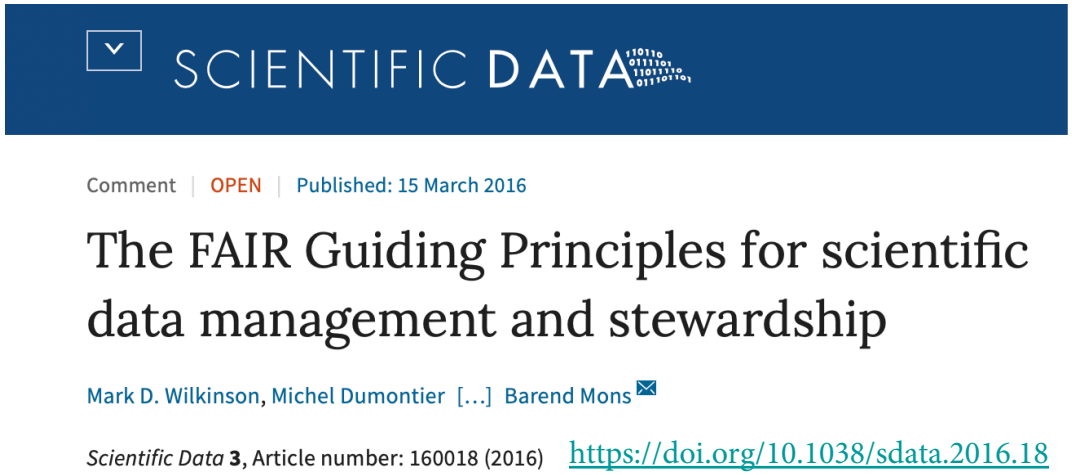
Making your data FAIR

Kristina Hettne 08/05/2020

Universiteit Leiden
The Netherlands

Findable, Accessible, Interoperable, Reusable (FAIR)



SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

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

Scientific Data **3**, Article number: 160018 (2016) https://doi.org/10.1038/sdata.2016.18

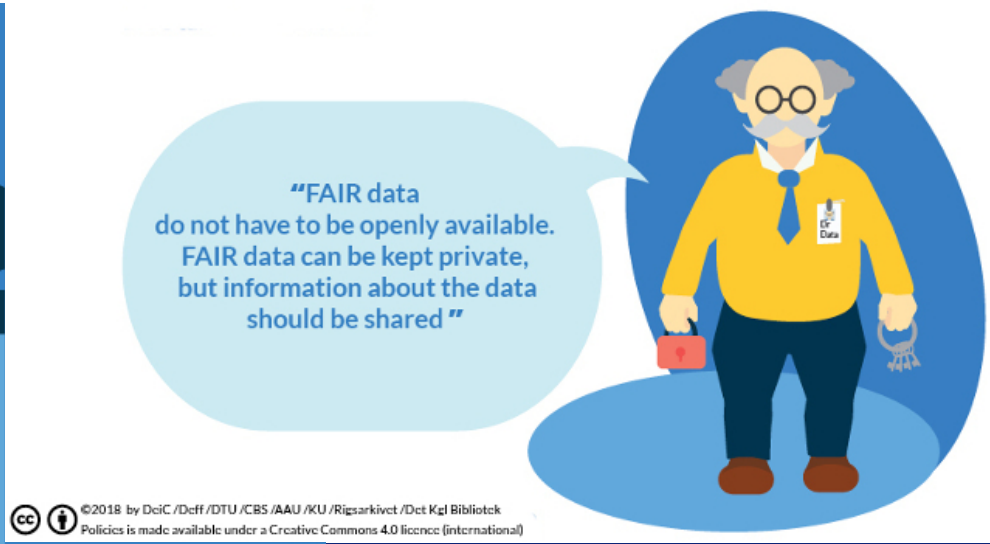
Research data needs to:

- Be accessible under clear conditions and licenses
- With clear references
- With rich metadata

Privacy-sensitive data can meet the FAIR principles

By making your data FAIR you…

- Create opportunities for sharing and reuse
- Enlarge your exposure
- Enhance your impact
- Show your future employer what you have done
- Avoid issues about verification
- Comply with requirements from funders



It's all about 15 principles

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

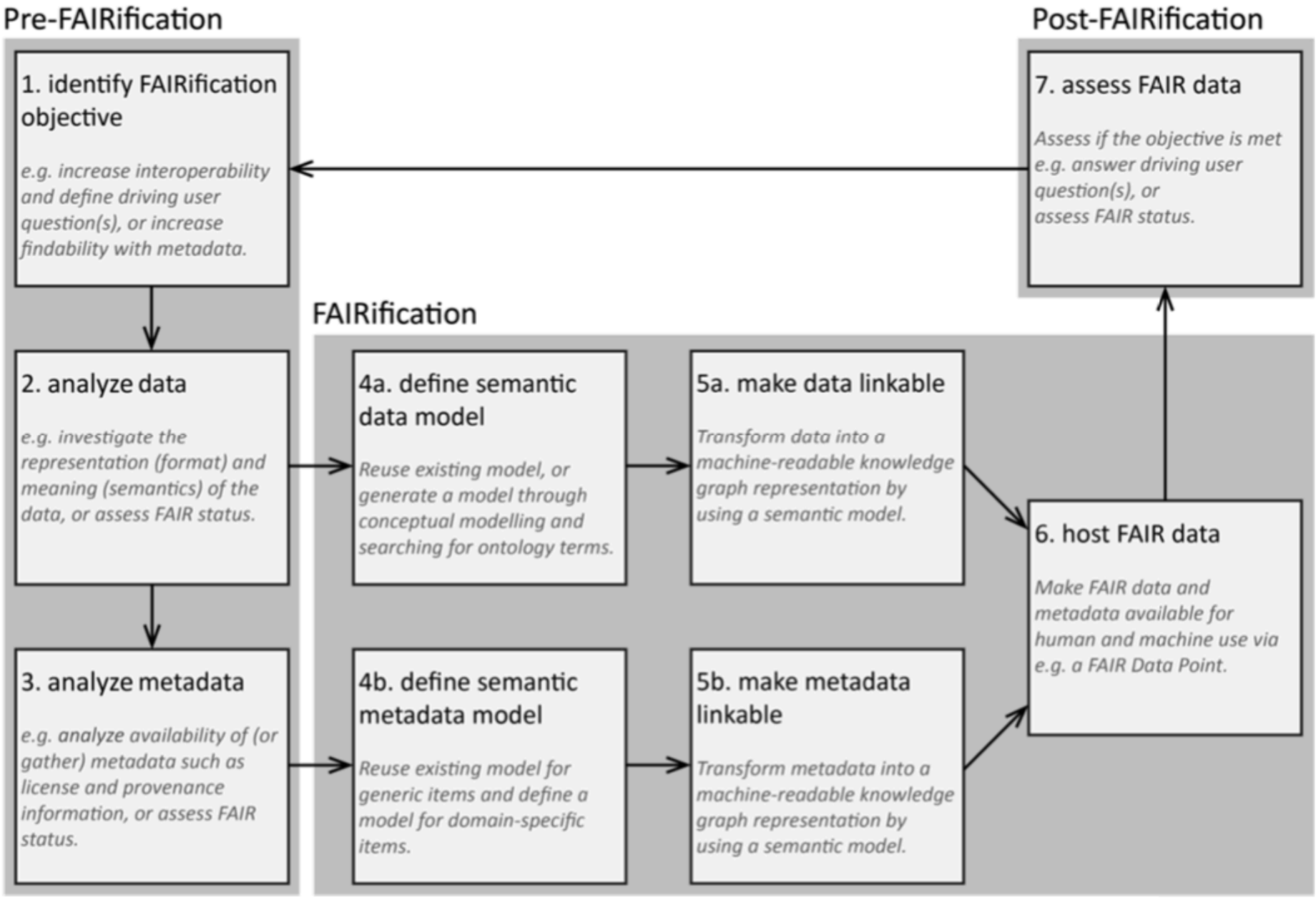
To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be 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

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

http://www.nature.com/articles/sdata201618 https://www.go-fair.org/fair-principles/

The FAIRification process



https://www.mitpressjournals.org/doi/full/10.1162/dint_a_00028

FAIR-ness: what is in the hands of the researcher?

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

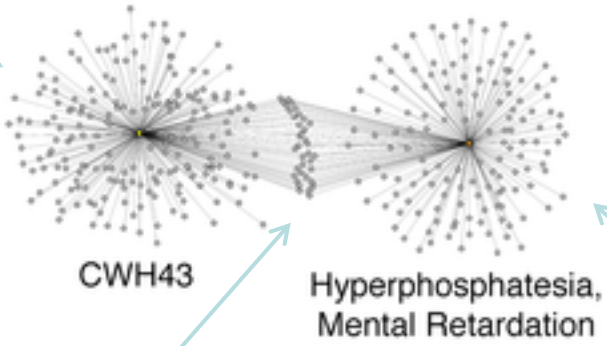
To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be 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

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

Example 1: finding new disease genes mined from scientific literature

All information in the literature about a gene



All information in the literature about a disease

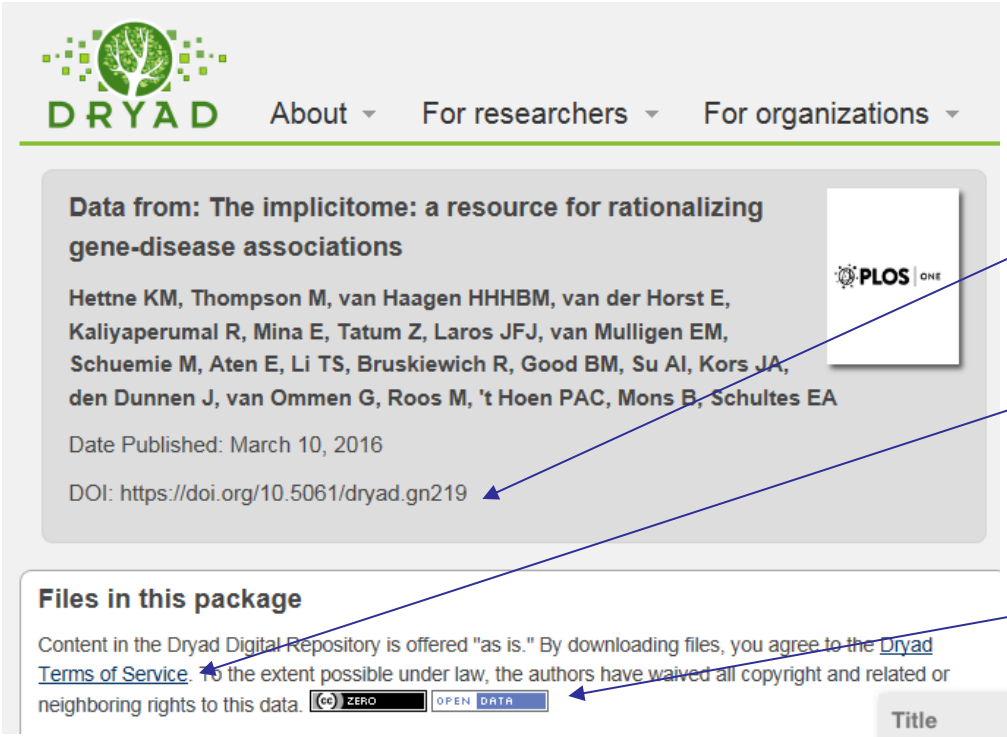
Overlapping information

=

New evidence for associating a gene with a disease

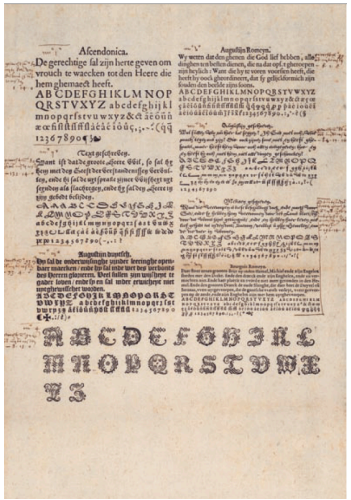
Thanks to data reuse: ~204.000.000 novel gene-disease associations

Hettne KM, Thompson M, van Haagen HHHBM, van der Horst E, Kaliyaperumal R, et al. (2016) The Implicitome: A Resource for Rationalizing Gene-Disease Associations. PLOS ONE 11(2): e0149621. https://doi.org/10.1371/journal.pone.0149621
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0149621


DRYAD
About
For researchers
For organizations
Data from: The implicitome: a resource for rationalizing gene-disease associations
PLOS ONE
Hettne KM, Thompson M, van Haagen HHHBM, van der Horst E, Kaliyaperumal R, Mina E, Tatum Z, Laros JFJ, van Mulligen EM, Schuemie M, Aten E, Li TS, Bruskiewich R, Good BM, Su AI, Kors JA, den Dunnen J, van Ommen G, Roos M, 't Hoen PAC, Mons B, Schultes EA
Date Published: March 10, 2016
DOI: https://doi.org/10.5061/dryad.gn219
Files in this package
Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the Dryad Terms of Service. To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.
CC ZERO
OPEN DATA
F
A
I
R
Title
All associations as nanopublications
Downloaded
49 times
Description
The complete set of all ~204 million associations (explicit and implicit) as nanopublications. Each nanopublication asserts an association between a gene and a disease concept and the percentile rank of the match score.
Download
gda-np.nq.gz (29.22 Gb)
Details
View File Details
Title
All associations as CSV
Downloaded
50 times
Description
All ~204 million associations (explicit and implicit) listed as a CSV text file. Note that concept pairs are specified by concept ID and its first label according to our thesaurus.
Download
README.txt (228 bytes)
Download
matchscores.release.csv.gz (7.009 Gb)
Details
View File Details

Example 2: Book trade history in Leiden

- Scholarly archive of book historian Prof. dr. Paul Hoftijzer
- Archive contains descriptions of people, organizations and events relevant to Leiden book trade in the early modern period
- 1. Why FAIRify?
  - Teach Digital Humanities techniques
  - Open up for possible reuse



Peter Verhaar
Digital Scholarship Librarian

Leiden University Fund (LUF) teaching grant

Discover the world at Leiden University

Raw data (Microsoft Word)

Aa, Cornelis van der (* 1749?; † ?; w. 1767-?)
Boekverkoper.
Geen lid van de grote boekverkopersfamilie Van der Aa.
Gilde: Pre-1765 L bij Johannes Lemair; 5-8-1765 bij Jacobus van der Spijck voor 4 jaar. 2-10-1767 Vrijmeester (AB 83a, f. 34v).
1796 te Haarlem, 1816 te Amsterdam.
GAL, prentverzameling 46601 portret Cornelis van der Aa, boekverkoper, geb. Leiden 1749, gegraveerd door Reinier Vinkeles naar C. van Geulen.
Veilingen: 11-11-1783 Veiling, met Vincent van der Vinne, van de collecties van Cornelius Asconius van Sypesteyn en C. en G. Schertzer, waaronder ook kunstvoorwerpen.
Lit.: Ledeboer.

Aa, Hillebrand van der (* 1661 (doop 22-3); † ± 1721; w. 1697-17?)
Plaatsnijder en beeldhouwer (bij zijn eerste huwelijk). Luthers. Zoon van Boudewijn Pietersz van der Aa en Annetje Poortemuller, broer van Pieter van der Aa. Getuigen bij zijn doop Roocks Immerseel, Tönjes Lockers en Elsche Heinrichsen.
Adres: 1683 Nieuwsteeg; 1685 Salomonssteeg (ouderlijk huis); 1696-99 Rapenburg?; 1701 Kloksteeg.
Huwelijk: 1. SH 28-4-1683 Maria Badde uit Haarlem (getuige zijn broer Pieter; † 29-1-1684, begr. PK); 2. SH 23-6-1684 Catharina Oesinger (Pieter van der Aa in de Nieuwsteeg zijn getuige; † 29-11-1749; zij werd begraven buiten Leiden; haar adres is Haarlemmerstraat bij de Turfmarkt); kinderen: Maria en Balduinus, de laatste werd predikant in de Leidse Lutherse gemeente.

Discover the world at Leiden University

Data cleaning and structuring

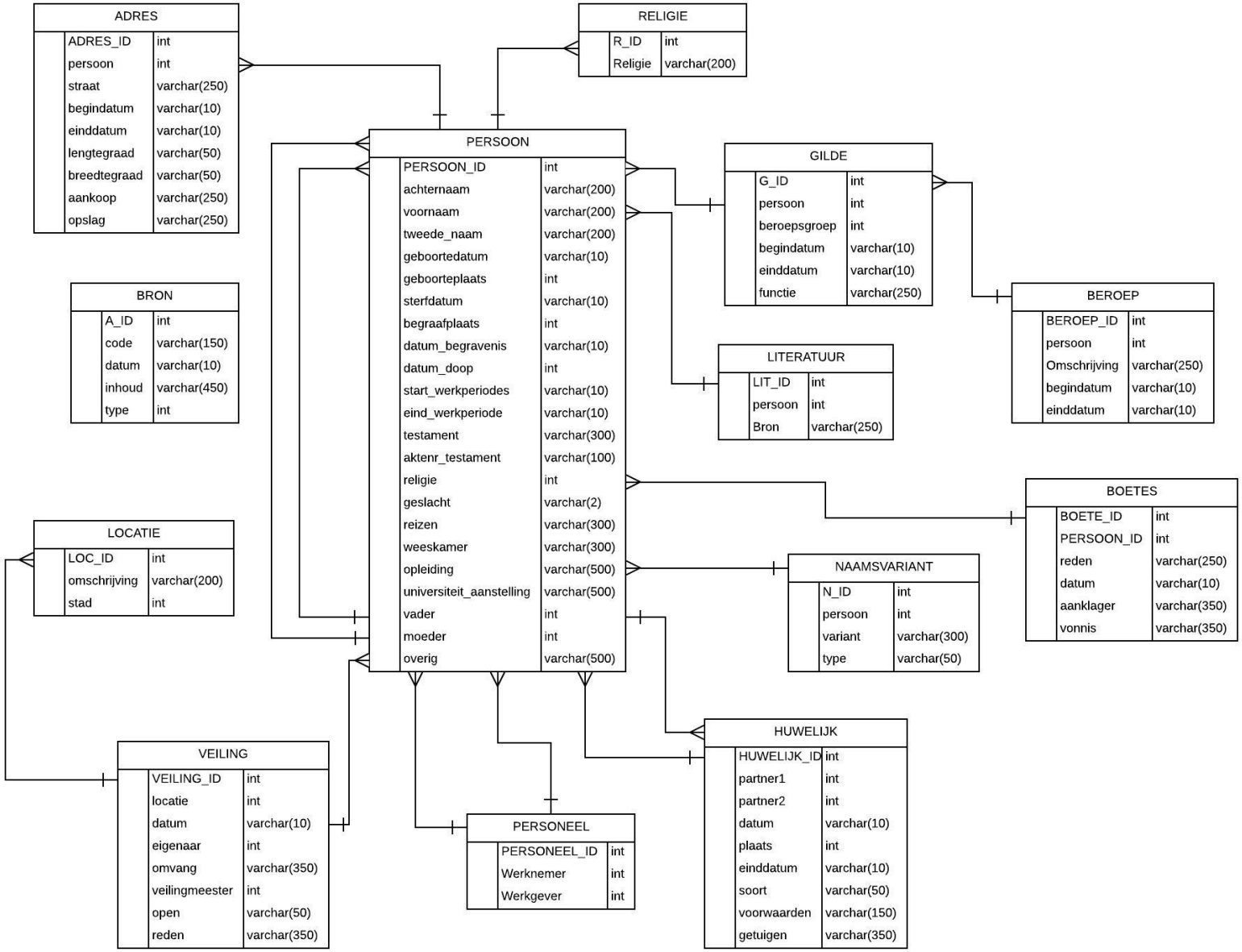
- Organize/Structure data before it is rendered machine-readable

Aa, Pieter Jansz van der (* Leiden 1697; † 2-8-1751 [begr. PK 31-7/7-8-1751]; w. 1719-36)

firstName	lastName	secondName	dateOfBirth	dateOfDeath	placeOfBirth
Pieter	Aa	Jansz van der	1697	1751-08-02	leiden
Boudewijn	Aa	Jansz van der	1692	NULL	leiden
Cornelis	Aa	van der	1749	NULL	NULL
Hillebrand	Aa	van der	1661	1721	NULL

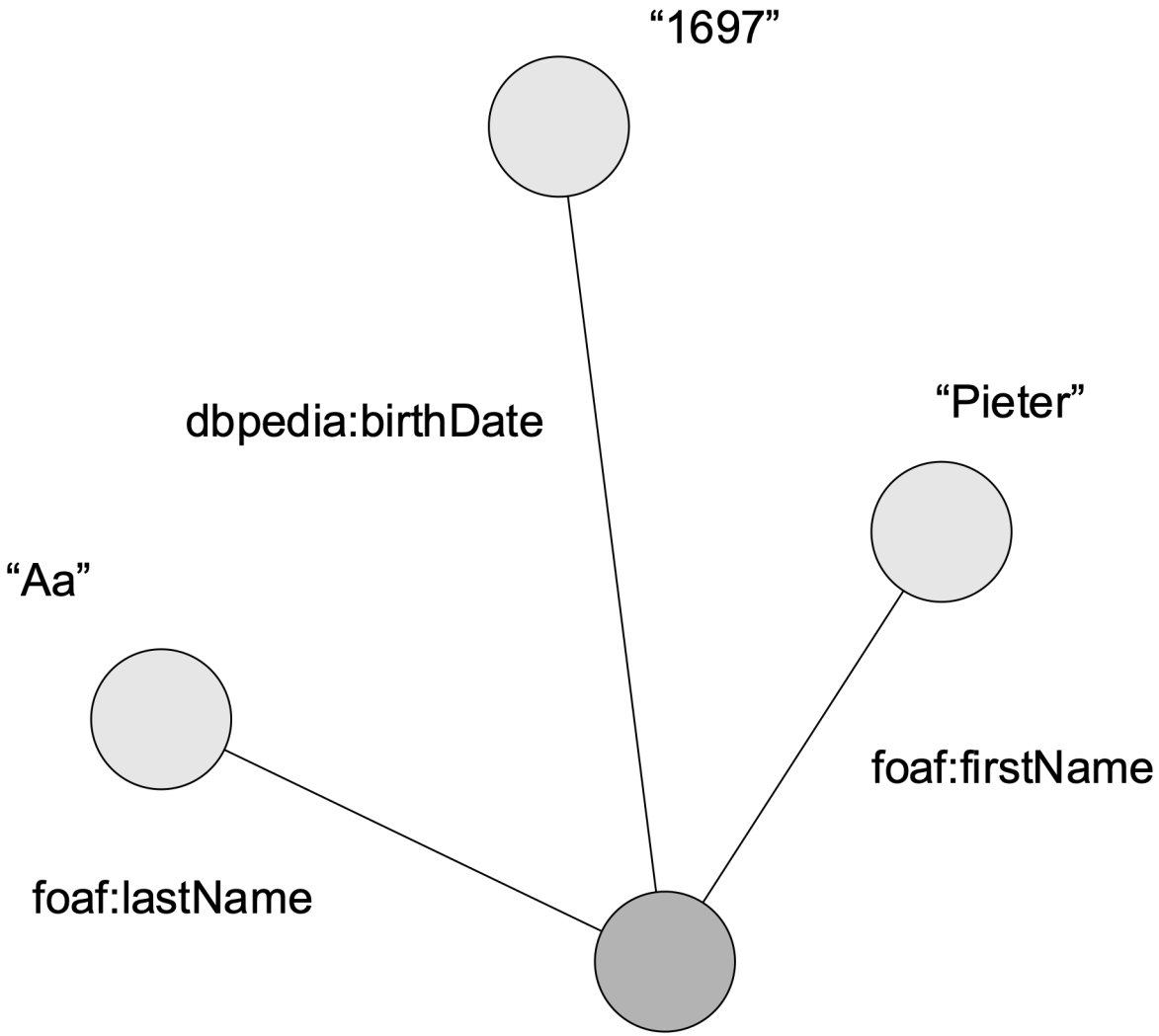
Discover the world at Leiden University

Define data model



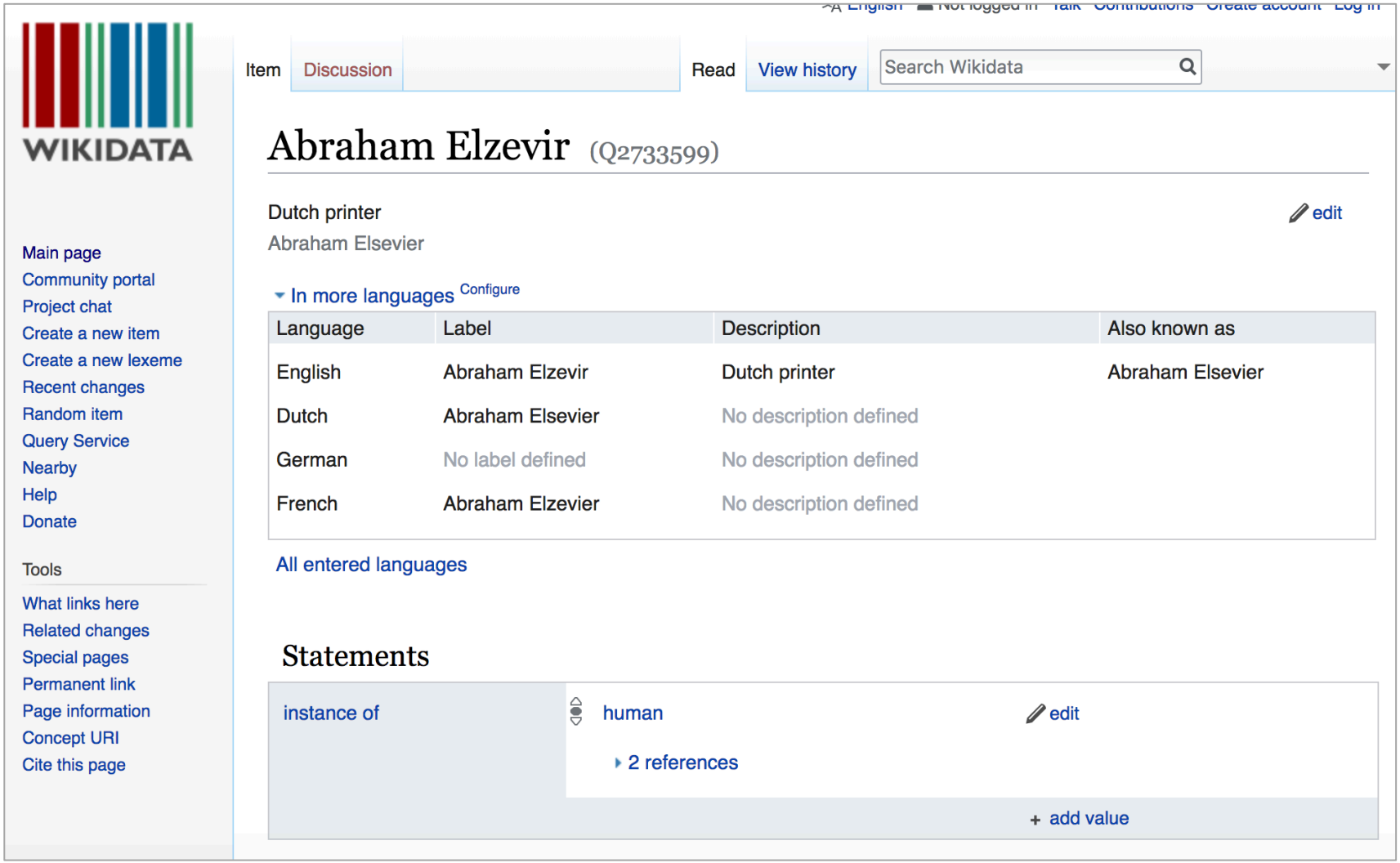
Discover the world at Leiden University

Define the semantic data model: a person

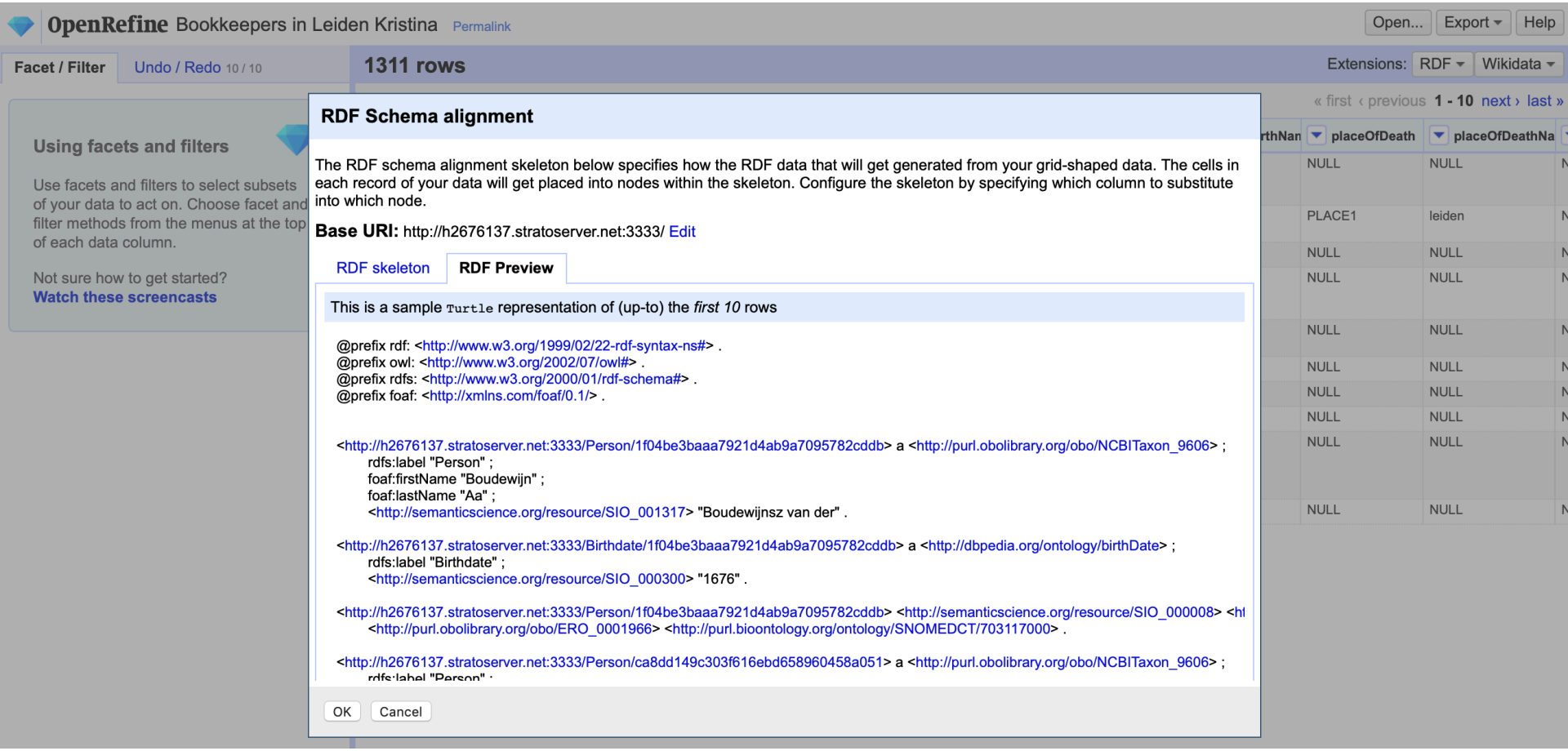


<https://nonsolus.leidenuniv.nl/person/ABB1>

Discover the world at Leiden University


WIKIDATA
Item
Discussion
Read
View history
Search Wikidata
Abraham Elzevir (Q2733599)
Dutch printer
Abraham Elsevier
edit
Main page
Community portal
Project chat
Create a new item
Create a new lexeme
Recent changes
Random item
Query Service
Nearby
Help
Donate
Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page
In more languages Configure
Language
Label
Description
Also known as
English
Abraham Elzevir
Dutch printer
Abraham Elsevier
Dutch
Abraham Elsevier
No description defined
German
No label defined
No description defined
French
Abraham Elzevier
No description defined
All entered languages
Statements
instance of
human
edit
2 references
add value

Make data linkable: transform to RDF



This was done with the OpenRefine RDF plugin, but there are many other ways…

Discover the world at Leiden University

Host FAIR data

Universiteit Leiden

Onderzoek | Onderwijs | Wetenschappers | Over ons | Faculteiten | Campus Den Haag | Alumni | Bibliotheek

Database over het Leidse Boek

Kopieer om de database te bevragen de genoemde zoekvraag in het tekstveld en klik op "Zoek!"!

Alle personen die geboren zijn buiten Leiden:

```
SELECT achternaam , voornaam, tweede_naam , geboortedatum , s.Naam
FROM PERSOON p , STAD s
WHERE p.geboorteplaats = s.S_ID
AND s.Naam != 'leiden'
ORDER BY p.geboortedatum
```

Zoek!

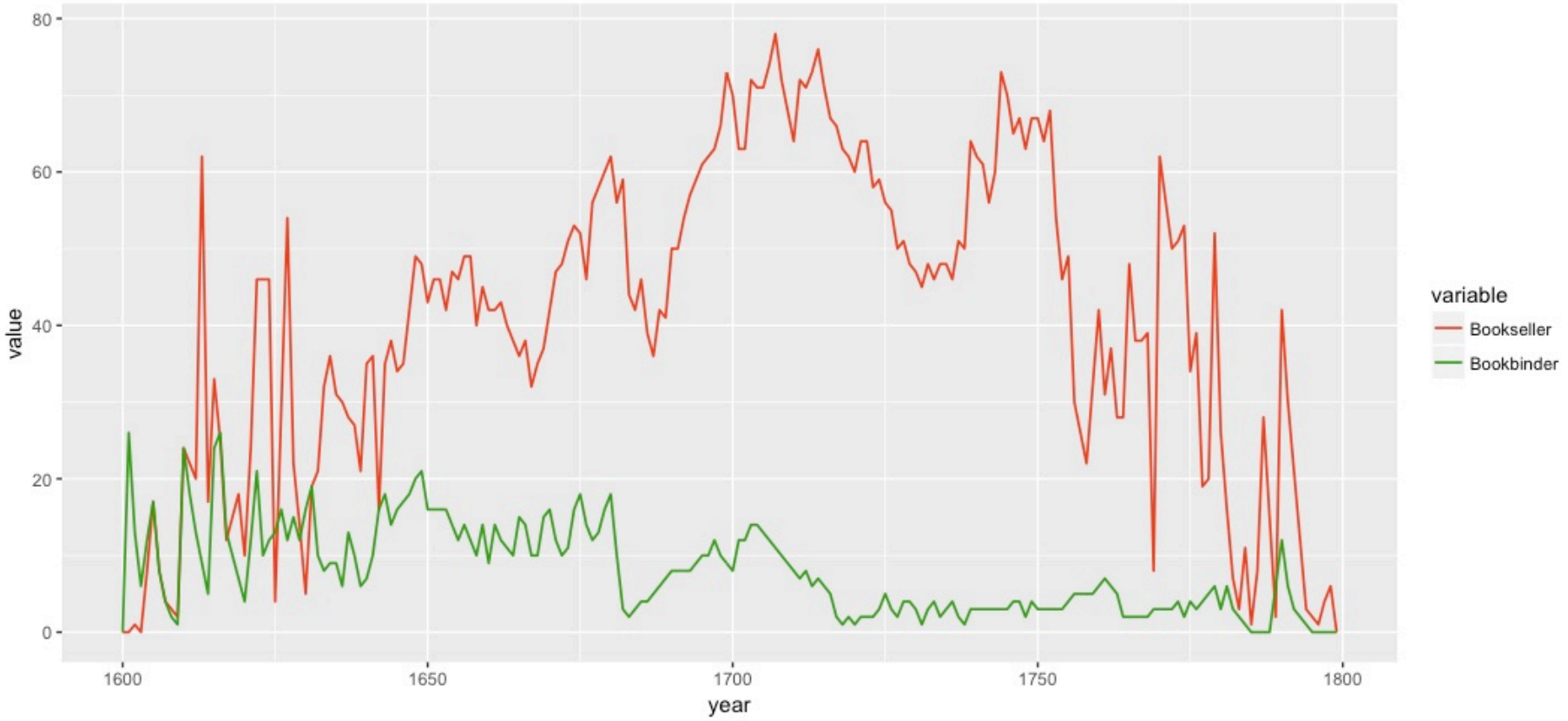
https://bookandbyte.universiteitleiden.nl/Boekgeschiedenis/

Discover the world at Leiden University

BA student projects "Boekgeschiedenis in de praktijk" spring 2019

- The database has been used to study the following topics:
  - The history of the guild of booksellers in Leiden
  - The network of family relations
  - The influx of foreigners in the Leiden book trade
  - The occurrences of banned books on the catalogues of book actions organised in Leiden
- Insights from these projects led to improvements of the database
- Ongoing project

Discover the world at Leiden University



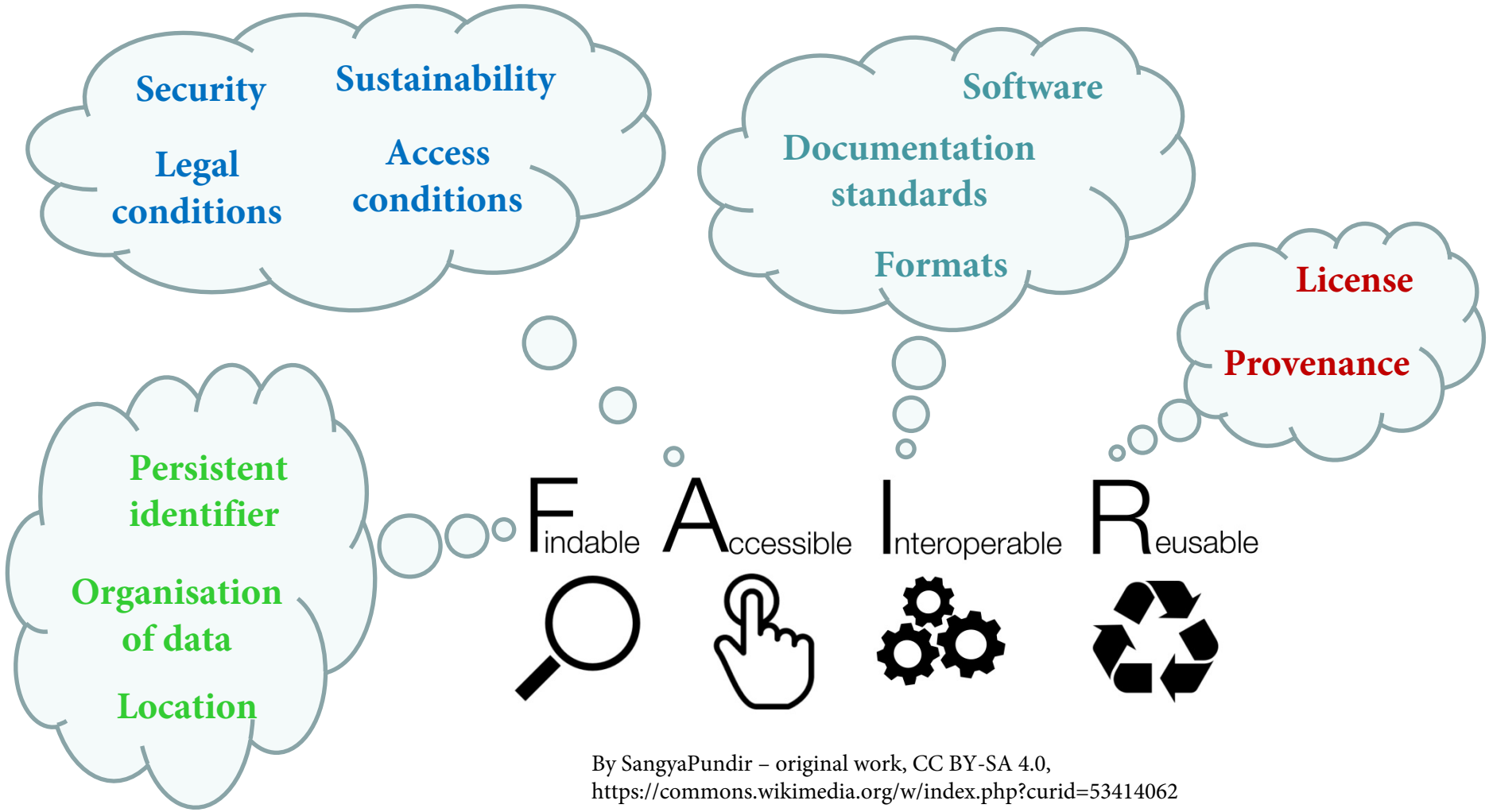
Number of booksellers and bookbinders, 1600-1800

Discover the world at Leiden University

Locations of booksellers in 1700

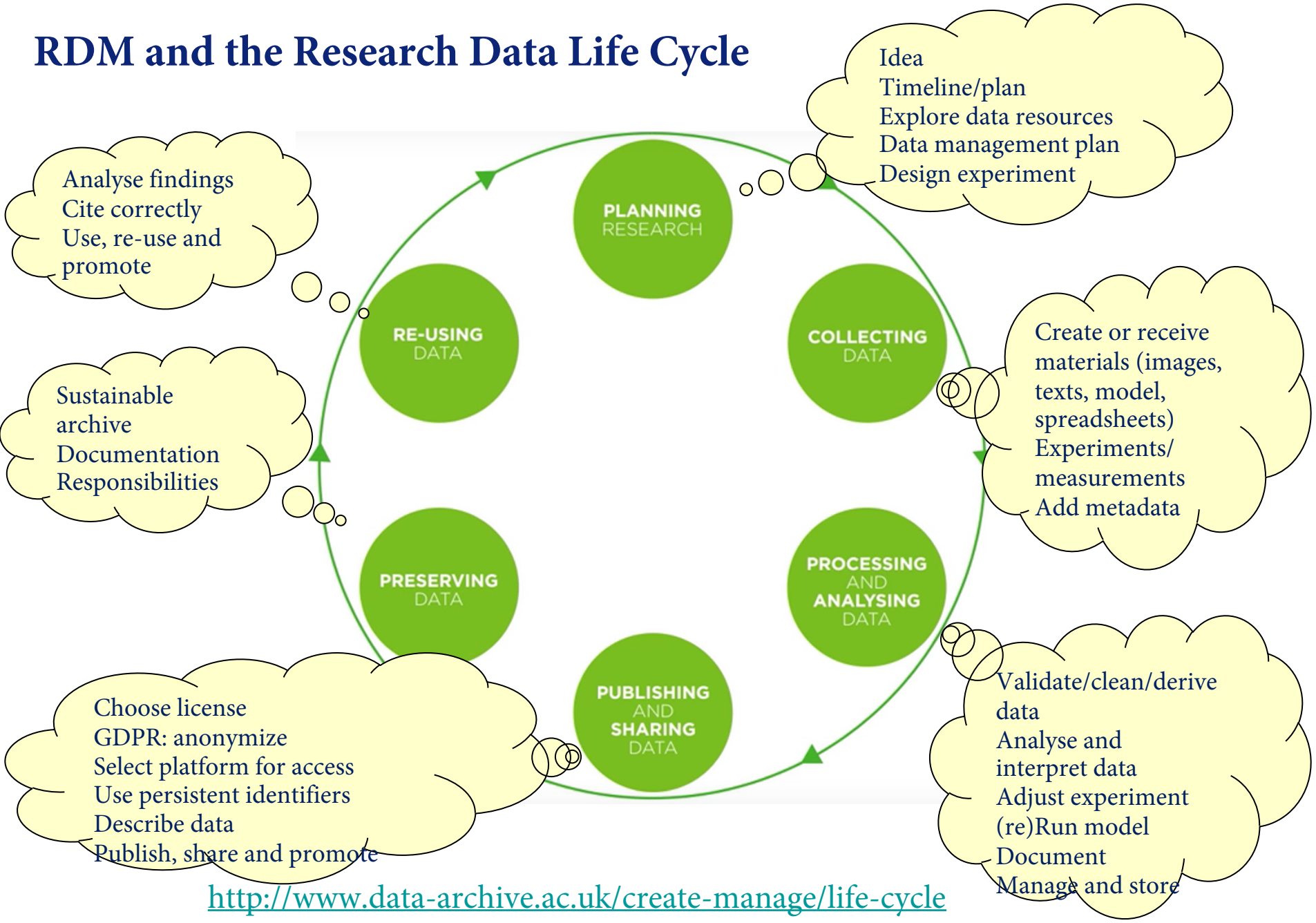
Discover the world at Leiden University

FAIR data management planning



By SangyaPundir – original work, CC BY-SA 4.0,
https://commons.wikimedia.org/w/index.php?curid=53414062

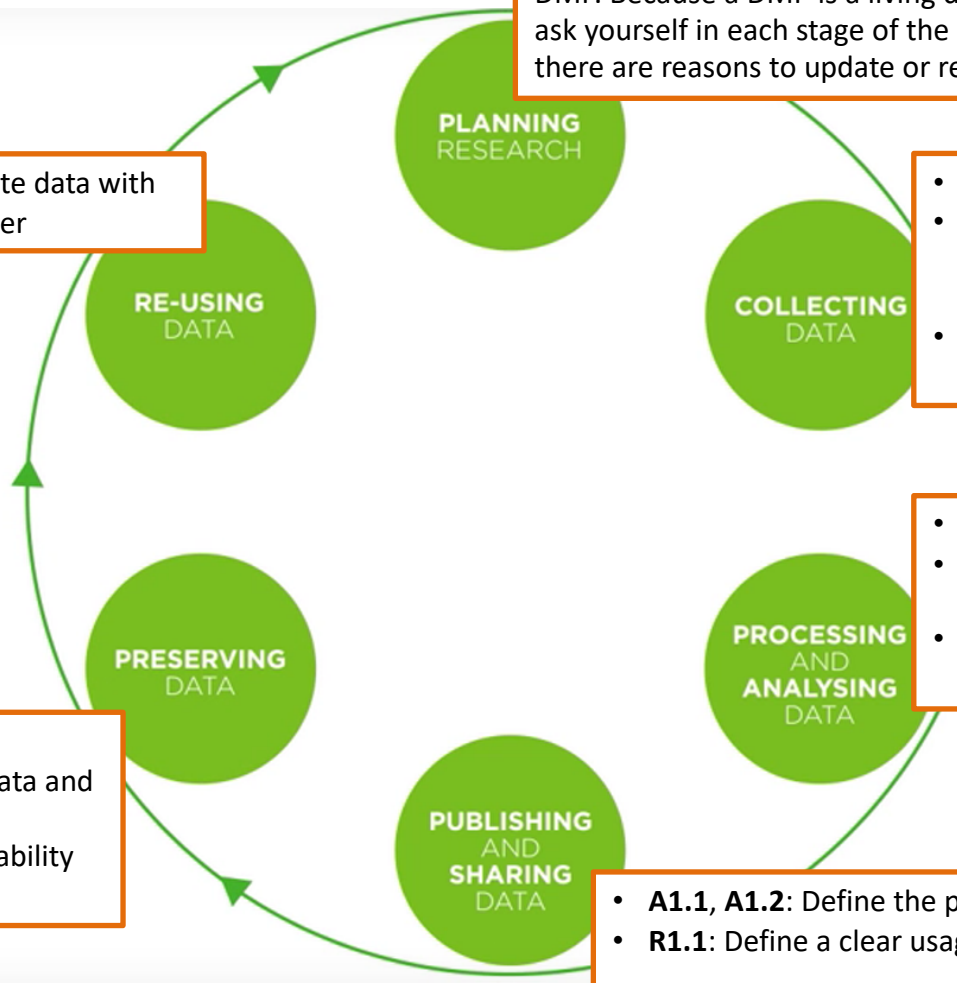
RDM and the Research Data Life Cycle



http://www.data-archive.ac.uk/create-manage/life-cycle

FAIR and the Research Data Life Cycle

Plan for FAIR and well-managed data with a DMP. Because a DMP is a living document, ask yourself in each stage of the lifecycle if there are reasons to update or refine it.



- **F1**: Make sure you cite data with its persistent identifier

- **R1.2**: Document what you do
- **I1**: Implement procedure for knowledge representation (data model)
- **F2**, **I2**, **R1.3**: Implement choice of metadata and data fields

- **R1.2**: Document what you do
- **I1**: Follow procedure for knowledge representation (data model)
- **F2**, **I2**, **R1.3**: Follow choice of metadata and data fields

- **F1**, **F3**, **F4**, **A1**, **A2**, **I3**, **R1**: Deposit data with metadata and documentation for interoperability and reusability with repository

- **A1.1**, **A1.2**: Define the protocols for access
- **R1.1**: Define a clear usage license

- The above choices will impact the choice of repository

http://www.data-archive.ac.uk/create-manage/life-cycle

4 steps to make your data more FAIR

- F: Put your data in a repository
- A: Make sure there is a data access protocol
- I: Describe your data using the metadata scheme offered by the repository
- R: Choose a license

Dataset FAIR-ness

- As a researcher you do a good job if you:
  - Fill out as much as possible of the repository fields when you submit your data
- As a researcher you do a great job if you:
  - Use metadata standards to record metadata elements (this is hard, ask colleagues or a data steward for help!)
- As a researcher you do an amazing job if you:
  - Use standard vocabularies to record data elements (this is hard, ask a data steward for help!)
  - Save your data in an FAIR interoperable format such as XML or RDF (this is hard, ask a data steward for help!)

http://www.nature.com/articles/sdata201618 https://www.go-fair.org/fair-principles/

Want to learn more?

FEATURED ARTICLES

MORE +

The FAIR Principles: First Generation Implementation Choices and Challenges

Author : Barend Mons; Erik Schultes; Fenghong Liu; Annika Jacobsen

Institution : Leiden University Medical Center, Leiden 2333 ZA, The Netherlands; GO FAIR Inter…

Doi : 10.1162/dint_e_00023

Abstract (views 93) | Full Text PDF | Full Text HTML

FAIR Principles: Interpretations and Implementation Considerations

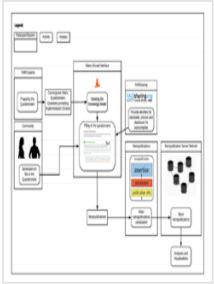
Author : Annika Jacobsen; Ricardo de Miranda Azevedo; Nick Juty; Dominique Batista; Simon C…

Institution : Leiden University Medical Center, Leiden 2333 ZA, The Netherlands; Institute of Dat…

Keywords : FAIR guiding principles; FAIR implementation; FAIR convergence; FAIR communitie…

Doi : 10.1162/dint_r_00024

Abstract (views 105) | Full Text PDF | Full Text HTML



FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources

Author : Hana Pergl Sustkova; Kristina Maria Hettne; Peter Wittenburg; Annika Jacobsen; Tobia…

Institution : GO FAIR International Support and Coordination Office, Leiden, The Netherlands; C…

Keywords : FAIR Implementation Choices and Challenges； Convergence； FAIR Communities

Doi : 10.1162/dint_a_00038

Abstract (views 114) | Full Text PDF | Full Text HTML

Data Intelligence

Host: National Science Library, Chinese Academy of Sciences

Publisher: National Science Library, Chinese Academy of Sciences

Co-Editors-in-Chief:

James Hendler , Huizhou Liu , Ying Ding

Executive Editors-in-Chief:

Guilin Qi , Yan Zhao

29 original papers
Official issue out in 2020

http://www.data-intelligence-journal.org

Discover the world at Leiden University

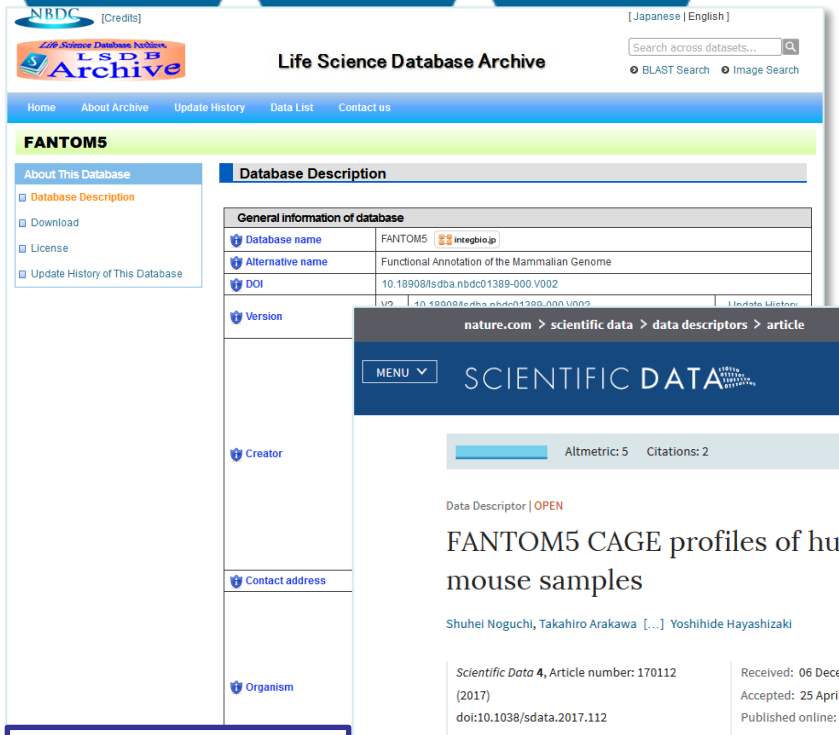
Maximum Exposure for yourself, your publications and your data

Total citations

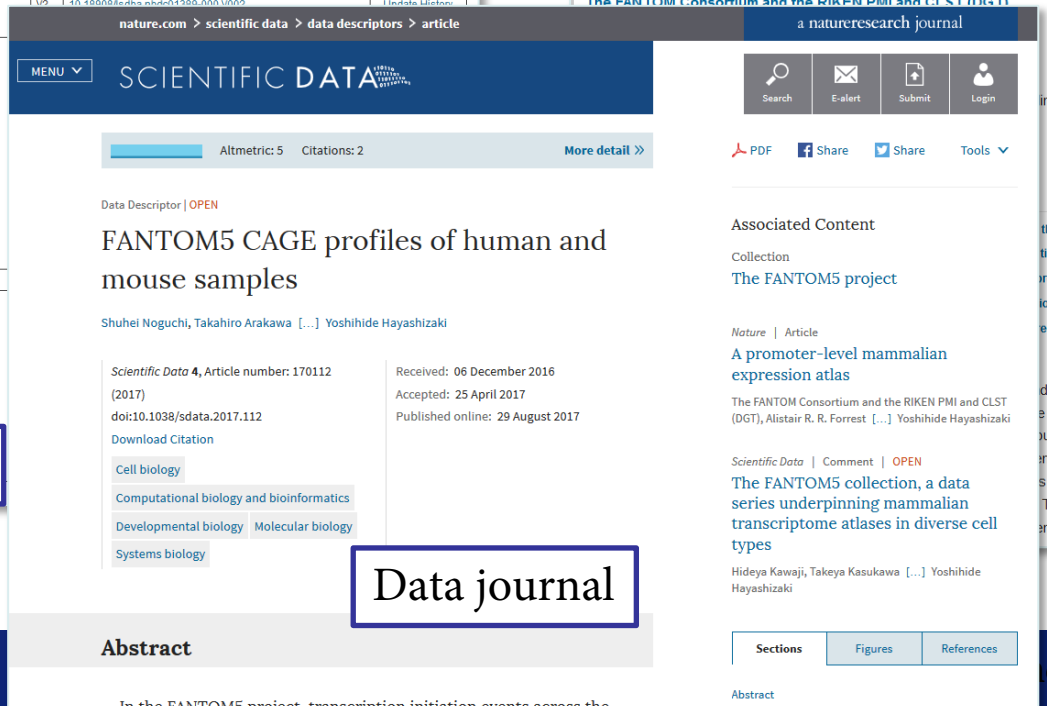
Online attention

Altmetric score (what's this?)

Tweeted by 177
On 7 Facebook pages
Mentioned in 2 Google+ posts



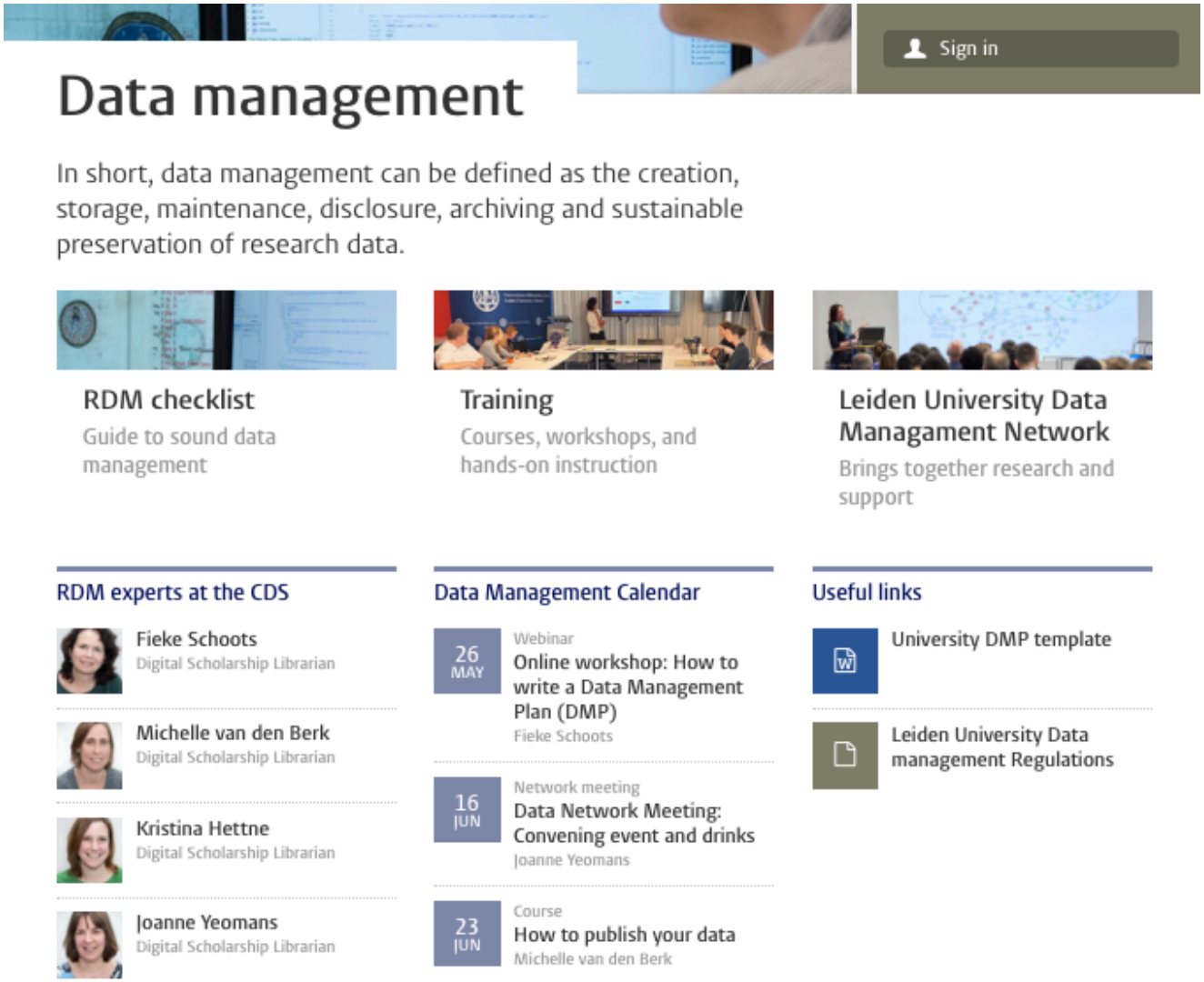
Data repository



Data journal

Journal article

Centre for Digital Scholarship



https://www.library.universiteitleiden.nl/researchers/data-management