FAIR Data Steward
Bruna Dos Santos Vieira

Radboud UMC

‘FAIR Data Stewardship’

EuRREB
European Registries for Rare
Endocrine and Bone conditions

FAIR Data Stewardship

Bruna dos Santos Vieira

Radboudumc

EuRREB
European Registries for Rare
Endocrine and Bone conditions

Funded by
the European Union

FAIR Principles

1

Data Stewardship

2

FAIR Data Stewardship Rare Diseases

3

FAIR Principles

80%

of data collected in scientific research cannot be reused

Based on: Mons, Barend, et al. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." Information services & use 37.1 (2017): 49-56.

80%

of data collected in scientific research cannot be reused

Data management focuses on using data for a single project

Data are unstructured, or not well documented

Based on: Mons, Barend, et al. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." Information services & use 37.1 (2017): 49-56.

The FAIR Principles

15 guiding principles

Scientific data management and stewardship

Published in 2016

Covering data, metadata, and infrastructure

www.nature.com/scientificdata

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 measureable 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.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

Source: Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3.1 (2016): 1-9.

Metadata

Data describing your data / biobank / registry / …

Make sure that others understand:

- **What your research project is about**
- **What (type of) data your collect**
- **If/how they could access that data**

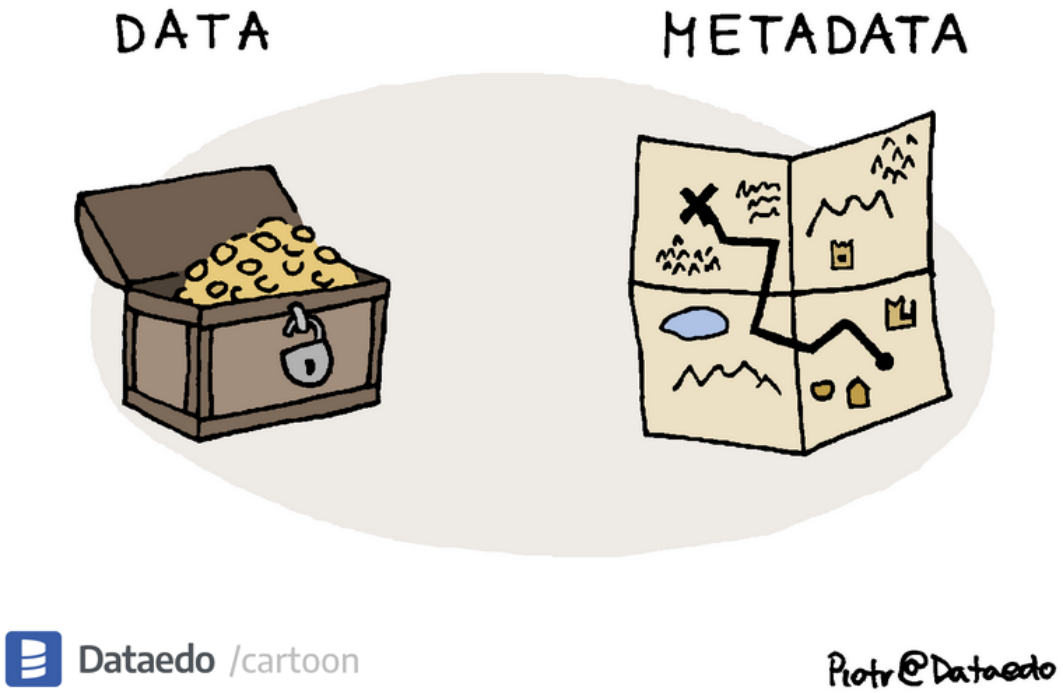


Image: Dataedo | Data is treasure, but how do you know where it is burried?
https://dataedo.com/cartoon/data-vs-metadata-2

The FAIR Principles

Describe <u>what</u> you should do with your (meta)data

Do <u>not</u> describe <u>how</u> you should do that

www.nature.com/scientificdata

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 measureable 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.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

Source: Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3.1 (2016): 1-9.

FAIR: an acronym for humans & machines

Findable

Described, identified and registered or indexed in a clear and unambiguous manner

Accessible

Accessible through a clearly defined access procedure & metadata should always remain accessible

Interoperable

Conceptualised, expressed and structured using common, published standards

Reusable

Characteristics and provenance are described in detail, with clear conditions for use

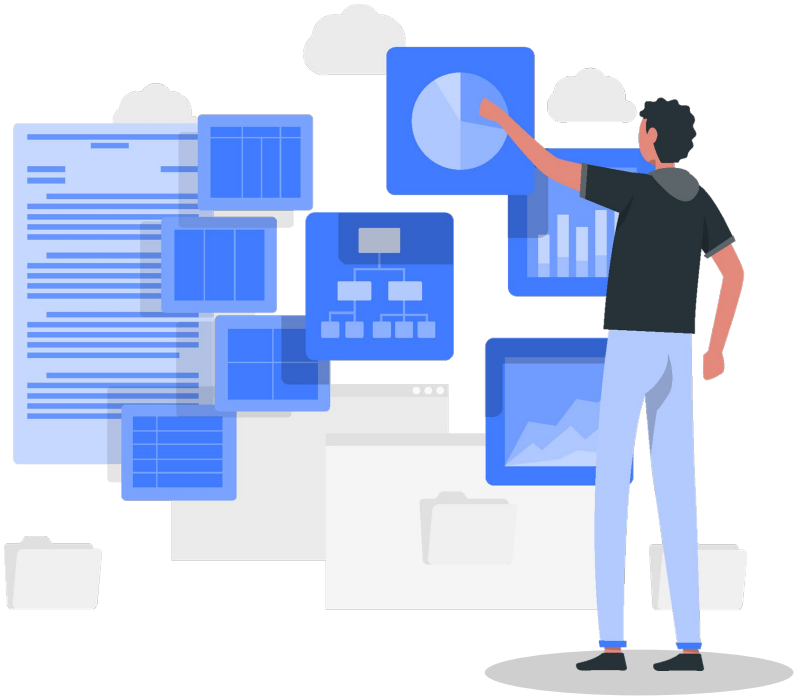
Source: Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3.1 (2016): 1-9.

The need for machine-readable, FAIR (meta)data

The body of knowledge and amount of available data is growing exponentially

Maximize the knowledge gained from the efforts and sacrifices of participants

We are unable to operate at the scope, scale, and speed required for this



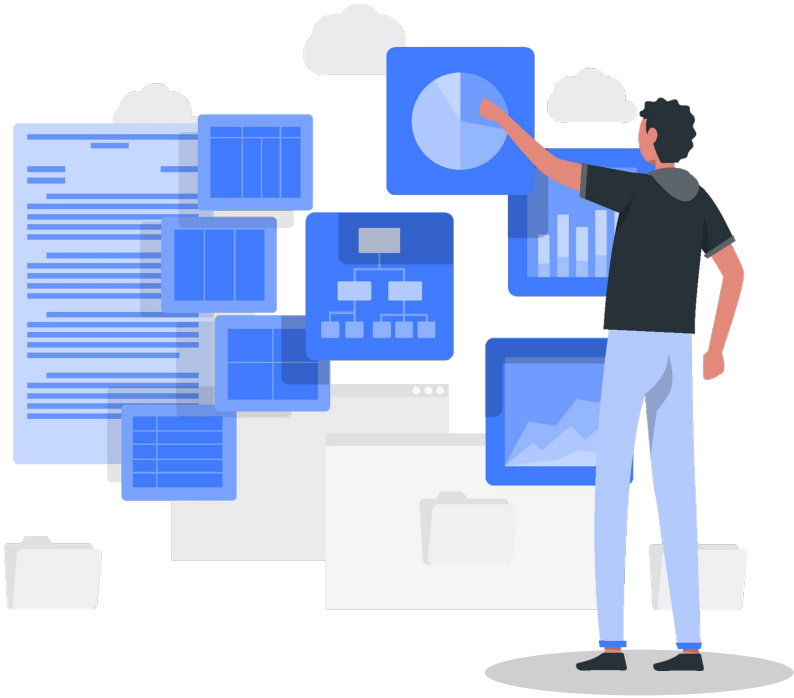
The need for machine-readable, FAIR (meta)data

The body of knowledge and amount of available data is growing exponentially

Maximize the knowledge gained from the efforts and sacrifices of participants

We are unable to operate at the scope, scale, and speed required for this

We need machines to find, access, interoperate, and reuse data → machine-readable (meta)data



Misconceptions about the principles

FAIR is not a standard

FAIR is not equal to Open

FAIR is not just about humans being able to find, access, reformat and finally reuse data

Information Services & Use 37 (2017) 49–56
DOI 10.3233/ISU-170824
IOS Press

Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud

Barend Mons [a,b,c,*], Cameron Neylon [d], Jan Velterop [e], Michel Dumontier [f], Luiz Olavo Bonino da Silva Santos [b,g] and Mark D. Wilkinson [h]

[a] *Leiden University Medical Centre, Leiden, The Netherlands*
E-mail: b.mons@lumc.nl
[b] *Dutch Techcentre for Life Sciences, Utrecht, The Netherlands*
[c] *Netherlands eScience Centre, Amsterdam, The Netherlands*
[d] *Centre for Culture and Technology, Curtin University, Perth, Western Australia*
[e] *Independent Open Access Publishing Consultant, Guildford, United Kingdom*
[f] *Institute for Data Science, Maastricht University, Maastricht, The Netherlands*
[g] *Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
[h] *Centre for Plant Biotechnology and Genomics U.P.M. – I.N.I.A., Madrid, Spain*

Abstract. The FAIR Data Principles propose that all scholarly output should be Findable, Accessible, Interoperable, and Reusable. As a set of guiding principles, expressing only the kinds of behaviours that researchers should expect from contemporary data resources, how the FAIR principles should manifest in reality was largely open to interpretation. As support for the Principles has spread, so has the breadth of these interpretations. In observing this creeping spread of interpretation, several of the original authors felt it was now appropriate to revisit the Principles, to clarify both what FAIRness is, and is not.

Keywords: FAIR Data, Open Science, interoperability, data integration, standards

1. Growing awareness of FAIRness

Open Science is a growing movement. The European Council adopted Open Science and the reusability of research data as a priority, as did the G7 at their summit in Japan [9]. This provided fertile ground for the rapid uptake of the FAIR Data Principles [25] since their recent publication [3]. The DG RTD (the Directorate General for Research and Innovation) of the European Commission took the lead [6], but in close collaboration with other directorates and the USA-based Big Data to Knowledge (BD2K) of the NIH (National Institutes of Health) [15]. Science Europe has adopted FAIR principles as the basis for sharing administrative data on funding [7]. The G20 went further in the 2016 Hangzhou summit by endorsing the FAIR Principles by name [8]. The Principles have also resonated in many discussions beyond their original scope of research data sharing, in domains as diverse as Archaeology [22], and

Source: Mons, Barend, et al. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." Information services & use 37.1 (2017): 49-56.

Misconceptions about the principles

FAIR is not a standard

FAIR is not equal to Open

FAIR is not just about humans being able to find, access, reformat and finally reuse data

FAIR at its core:
machine-readable (meta)data

Information Services & Use 37 (2017) 49–56
DOI 10.3233/ISU-170824
IOS Press

Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud

Barend Mons [a,b,c,*], Cameron Neylon [d], Jan Velterop [e], Michel Dumontier [f], Luiz Olavo Bonino da Silva Santos [b,g] and Mark D. Wilkinson [h]

[a] *Leiden University Medical Centre, Leiden, The Netherlands*
E-mail: b.mons@lumc.nl
[b] *Dutch Techcentre for Life Sciences, Utrecht, The Netherlands*
[c] *Netherlands eScience Centre, Amsterdam, The Netherlands*
[d] *Centre for Culture and Technology, Curtin University, Perth, Western Australia*
[e] *Independent Open Access Publishing Consultant, Guildford, United Kingdom*
[f] *Institute for Data Science, Maastricht University, Maastricht, The Netherlands*
[g] *Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
[h] *Centre for Plant Biotechnology and Genomics U.P.M. – I.N.I.A., Madrid, Spain*

Abstract. The FAIR Data Principles propose that all scholarly output should be Findable, Accessible, Interoperable, and Reusable. As a set of guiding principles, expressing only the kinds of behaviours that researchers should expect from contemporary data resources, how the FAIR principles should manifest in reality was largely open to interpretation. As support for the Principles has spread, so has the breadth of these interpretations. In observing this creeping spread of interpretation, several of the original authors felt it was now appropriate to revisit the Principles, to clarify both what FAIRness is, and is not.

Keywords: FAIR Data, Open Science, interoperability, data integration, standards

1. Growing awareness of FAIRness

Open Science is a growing movement. The European Council adopted Open Science and the reusability of research data as a priority, as did the G7 at their summit in Japan [9]. This provided fertile ground for the rapid uptake of the FAIR Data Principles [25] since their recent publication [3]. The DG RTD (the Directorate General for Research and Innovation) of the European Commission took the lead [6], but in close collaboration with other directorates and the USA-based Big Data to Knowledge (BD2K) of the NIH (National Institutes of Health) [15]. Science Europe has adopted FAIR principles as the basis for sharing administrative data on funding [7]. The G20 went further in the 2016 Hangzhou summit by endorsing the FAIR Principles by name [8]. The Principles have also resonated in many discussions beyond their original scope of research data sharing, in domains as diverse as Archaeology [22], and

Source: Mons, Barend, et al. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." Information services & use 37.1 (2017): 49-56.

The benefits of machine-readable, FAIR (meta)data

- **Allow for unambiguous interpretation and analysis**
- **Allow for performing distributed analyses**
- **Allow for automated processing in a more efficient and scalable manner**

The benefits of machine-readable, FAIR (meta)data

Allow for unambiguous interpretation and analysis

Allow for performing distributed analyses

Allow for automated processing in a more efficient and scalable manner

Boost the performance of data re-users for the benefit of people living with a rare disease

FAIR: an acronym for humans & machines

Findable

Described, identified and registered or indexed in a clear and unambiguous manner

Accessible

Accessible through a clearly defined access procedure & metadata should always remain accessible

Interoperable

Conceptualised, expressed and structured using common, published standards

Reusable

Characteristics and provenance are described in detail, with clear conditions for use

Source: Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3.1 (2016): 1-9.

Findable

Datasets should be described, identified and registered or indexed in a clear and unambiguous manner.

Data and metadata must be assigned globally unique and persistent identifiers (F1)

Metadata should be extensive (F2)**, allowing for a dataset to be located using attributes in its metadata** (F2) **or its identifier** (F3)

Data and metadata must be included in a searchable resource (F4)

Based on: Jacobsen, Annika, et al. “FAIR Principles: Interpretations and Implementation Considerations”. Data Intelligence 2020; 2 (1-2): 10–29.

Fin

Dat nd

una

Dat
ass
per e

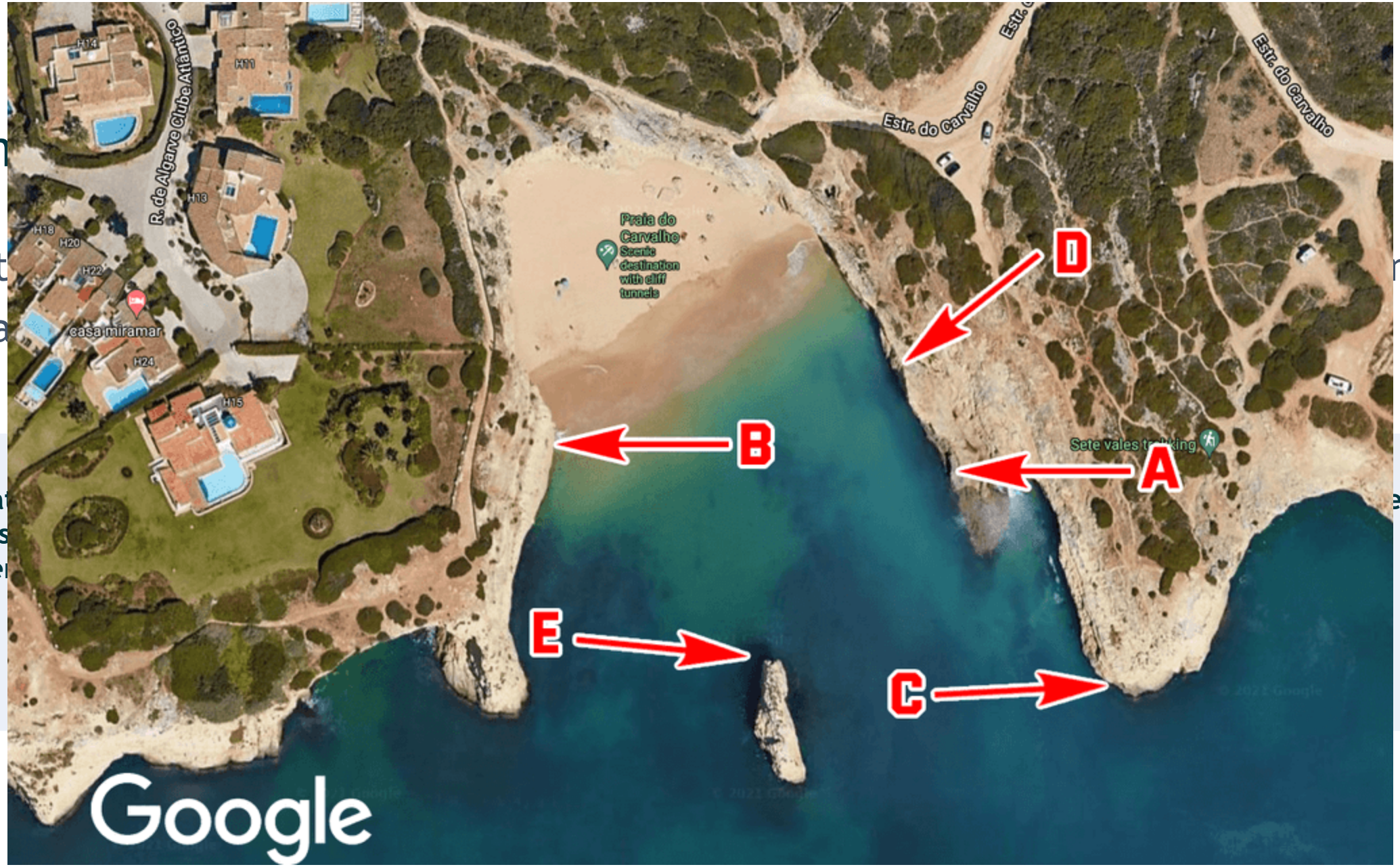


Image source: A guide to cliff diving in Benagil | pablo.gg

Accessible

Datasets should be accessible through a clearly defined access procedure, ideally using automated means. Metadata should always remain accessible.

Data and metadata can be retrieved using a standardized protocol (A1) **that is open, free, universally implementable** (A1.1)

The protocol allows for authentication and authorization procedures where necessary (A1.2)

Regardless of whether the data is still available, metadata should always be accessible (A2)

Based on: Jacobsen, Annika, et al. “FAIR Principles: Interpretations and Implementation Considerations”. Data Intelligence 2020; 2 (1-2): 10–29.

Accessibility

CAN I jump off this cliff?

Can I approach the edge of the cliff?

Image source: [Wheelchair-bound athlete honoured for climbing up mountain | Reuters]
[The Top 5 Coasteering Destinations in Spain | Manawa]

Interoperable

Data and metadata are conceptualised, expressed and structured using common, published standards.

Knowledge in data and metadata is represented using a formal, accessible, shared, and broadly applicable language (I1)

Data and metadata should use vocabularies that follow FAIR principles (I2)

Data and metadata should include qualified references to other data and metadata (I3)

Based on: Jacobsen, Annika, et al. “FAIR Principles: Interpretations and Implementation Considerations”. Data Intelligence 2020; 2 (1-2): 10–29.

Interoperable

Data
com

Kno
met
usin
shar
app

es
:a

Image source: https://cruxrange.com/wp-content/uploads/2019/12/rope-labels-for-rope-type.jpg.webp

Reusable

Characteristics of data and their provenance are described in detail according to domain-relevant standards, with clear and accessible conditions for use.

Data and metadata should be richly described with attributes to assess appropriateness for reuse (R1)

Data and metadata should include a clear and accessible data usage license (R1.1) **and detailed provenance** (R1.2)

Data and metadata should meet domain-relevant community standards (R1.3)

Based on: Jacobsen, Annika, et al. “FAIR Principles: Interpretations and Implementation Considerations”. Data Intelligence 2020; 2 (1-2): 10–29.

Reusability

- Do you have a paraglider? Or your swimming shoes/clothes? Can you swim? **(license)**
- Do you have your go-pro on and running and did you check if is the correct time of the year to capture the video you want? Did you check the weather and the water conditions? **(provenance)**
- Have you paid the paragliding company to have a safe jump? **(citation)**

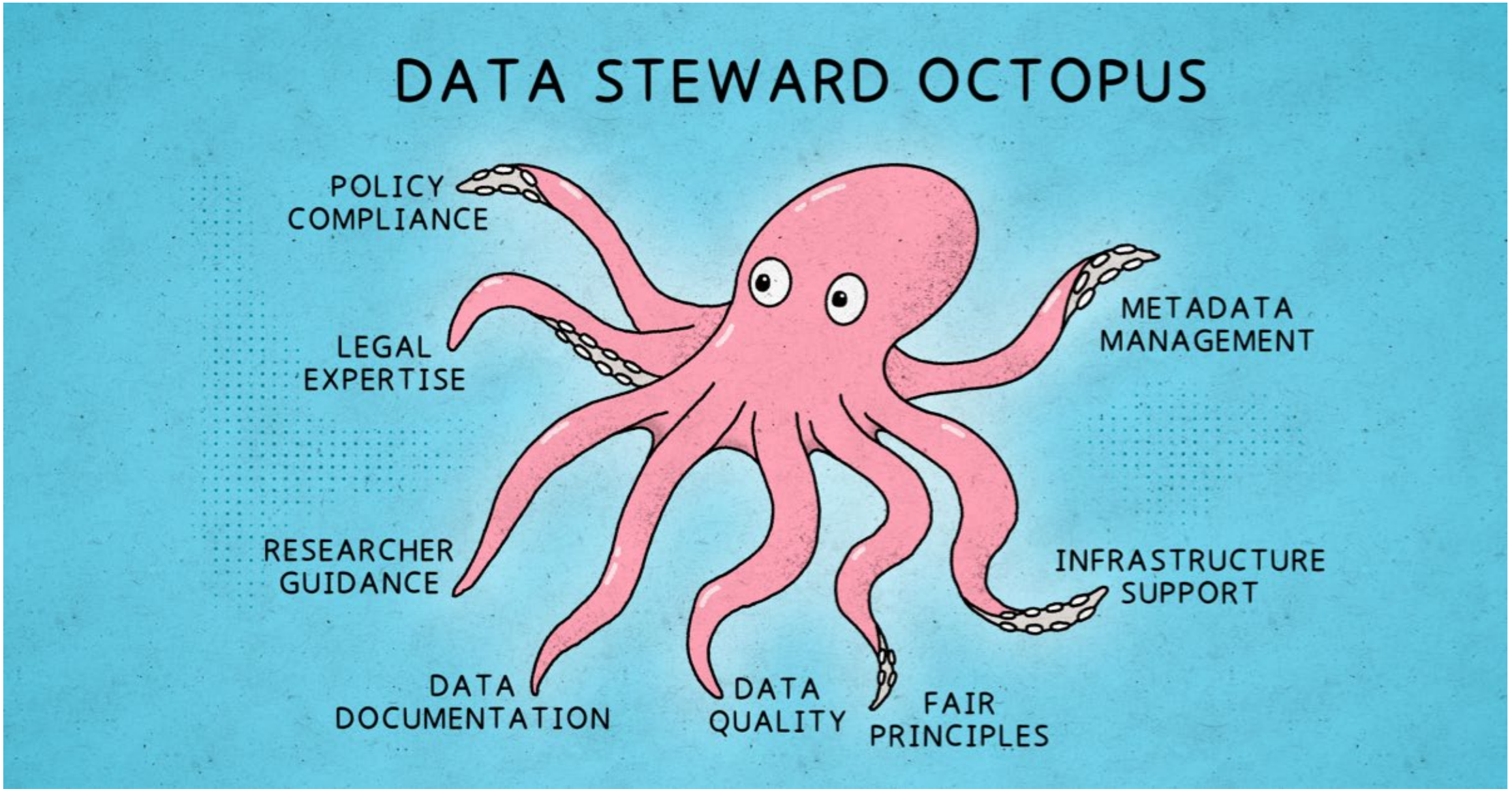
Reusability

Then sure!
Jump off the cliff!
Enjoy!!!

Data Stewardship

Data Stewardship

- Data Stewardship caters for making data FAIR in the research life cycle
- Data stewards, are professionals trained to deal with data through the whole data life cycle from data management to archiving, including FAIRification.
- Data stewardship takes care of the data in a way that the domain expert, the clinician, the researcher or the IT personnel cannot, or do not have the time and resources to do so.



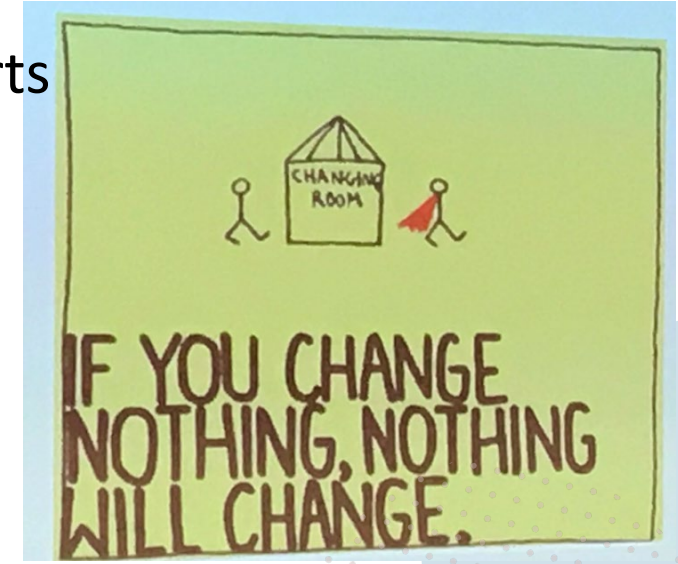
Source: [Who Is a Data Steward | FAIR Wizard]

FAIR Data Stewardship in Rare Diseases

- VASCA Registry (VASCERN)
  - The de novo FAIRification where we applied the FAIR principles during the creation of the registry, and during the implementation of the registry in the electronic data capture system
  - Now evolved to decentralized registry with many institutions involved
- FAIRification – EJP RD – European Joint Programme on Rare Diseases
  - EU-wide FAIRification services where we identified problems and solutions for FAIRification in large scale
  - Data stewardship services, consultancy, training, technical help
  - Onboarding to Rare Diseases Virtual Platform
- Data Hub - ERDERA
  - Co-Creation: aim at co-creating Rare Disease Resources in a FAIR manner and connect them to the EU RD ecosystem
  - FAIR Training Program 2025 - 2027 Part 1 opens registration – ERDERA

Take Home Messages

- FAIR is about Reuse of data, and how humans can make data ready for machines to help other humans find, access and connect data for reuse – it is sustainability in science
- FAIR is not: about open data and it is not a standard. Is a set of guiding principles, that require that you use certain standards that fit what is suggested in the principles
- FAIR is not a goal, is a mean to a goal
- “10% FAIR” is great, if that is enough for your goal
- FAIR is an iterative process
- Data stewardship is here to connect domain experts and data experts
- Data stewards cater for healthy and sustainable research life cycle
- Data stewards are here to support researchers
  to make their data reusable or to reuse other data



Acknowledgements

Slides credits:

- **Mark Wilkinson's and Marco Roos' minds**
- **Martijn Kersloot**
  - https://doi.org/10.5281/zenodo.14234494
  - https://doi.org/10.5281/zenodo.15461467

Team:

- **Inês Henriques**
- **Sergi Castillo**
- **Daniela Albring**
- **Cesar Bernabé**
- **Alberto Camara**
- **Nirupama Benis**
- **Qasim Khalid**

EuRREB
European Registries for Rare Endocrine and Bone conditions

Endo-ERN
European Reference Network on Rare Endocrine Conditions

ERN BOND
EUROPEAN REFERENCE NETWORK ON RARE BONE DISEASES

Thank you

Where to find me:

https://orcid.org/0000-0001-7893-0505

bruna.dossantosvieira@radboudumc.nl

/svbruna

EuRREB
European Registries for Rare Endocrine and Bone conditions

Endo-ERN
European Reference Network on Rare Endocrine Conditions

ERN BOND
EUROPEAN REFERENCE NETWORK ON RARE BONE DISEASES