FAIR

Stop data sharing, ….visit

AgBioData

Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

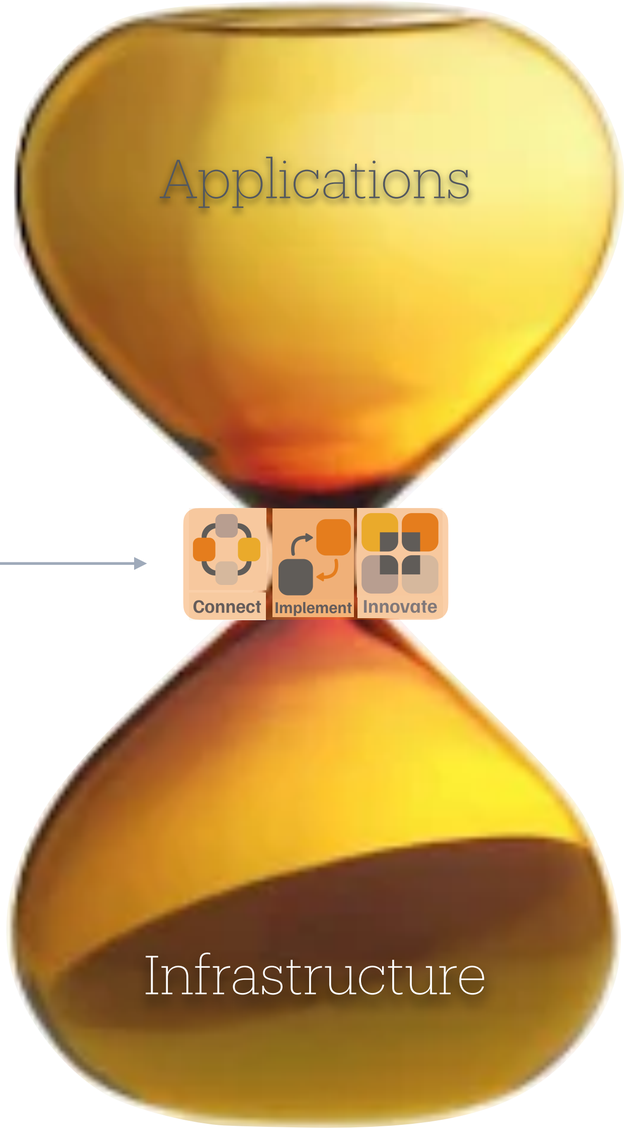
Barend Mons AgBioData, 2025

The end of data sharing

The Dawn of data Visiting

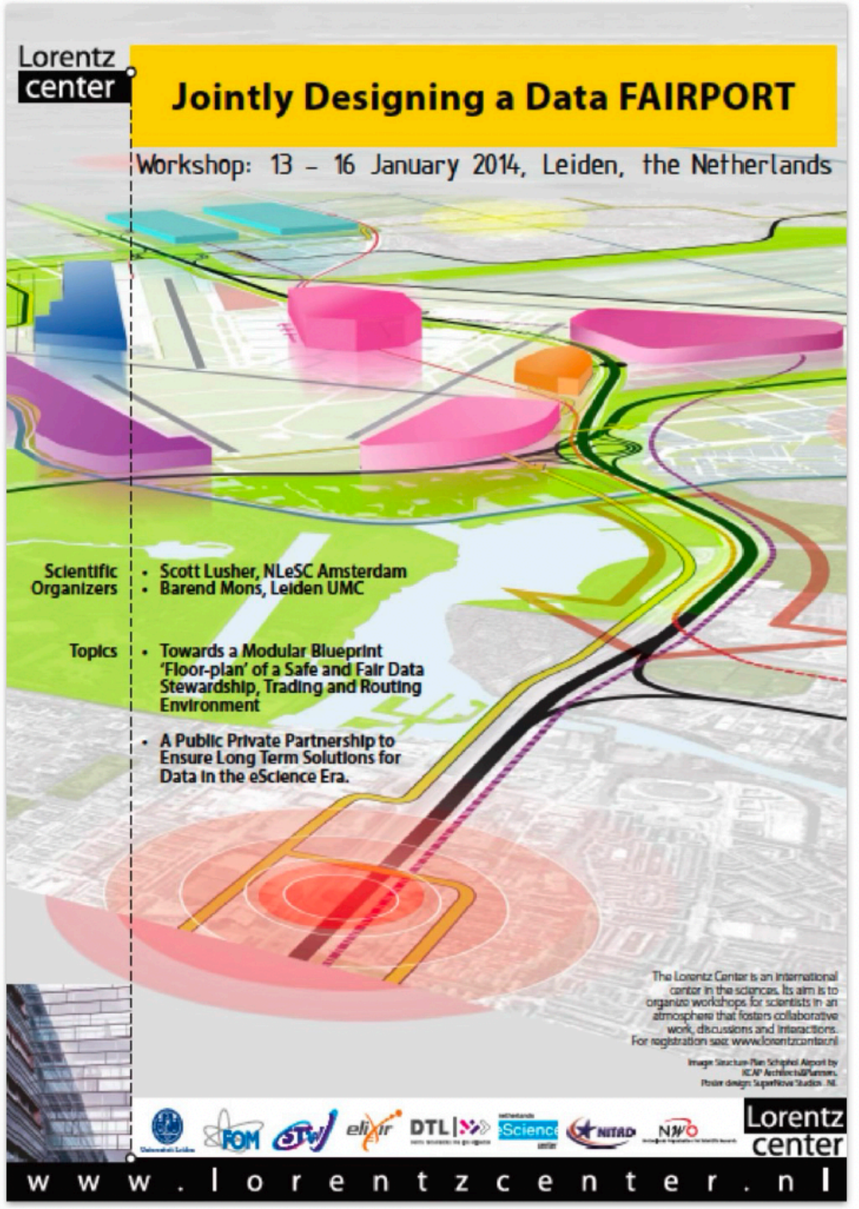
LIFE S

The **Leiden Initiative** for **FAIR** & **Equitable Science**



FAIR Data

- **2014-2016**
  - A set of 15 principles that **enhance the ability of machines** to do **F**, **A**, **I** and **R**
  - Reduce **ambiguity** and increase the **context** of data
  - **Socio-technical** framework (humans helping machines to help humans)
  - Rapidly and broadly accepted (skewed toward higher-level stakeholders)



Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

scientific data 2016

Explore content · About the journal · Publish with us

nature › scientific data › comment › article

Comment | Open access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, ... Barend Mons + Show authors

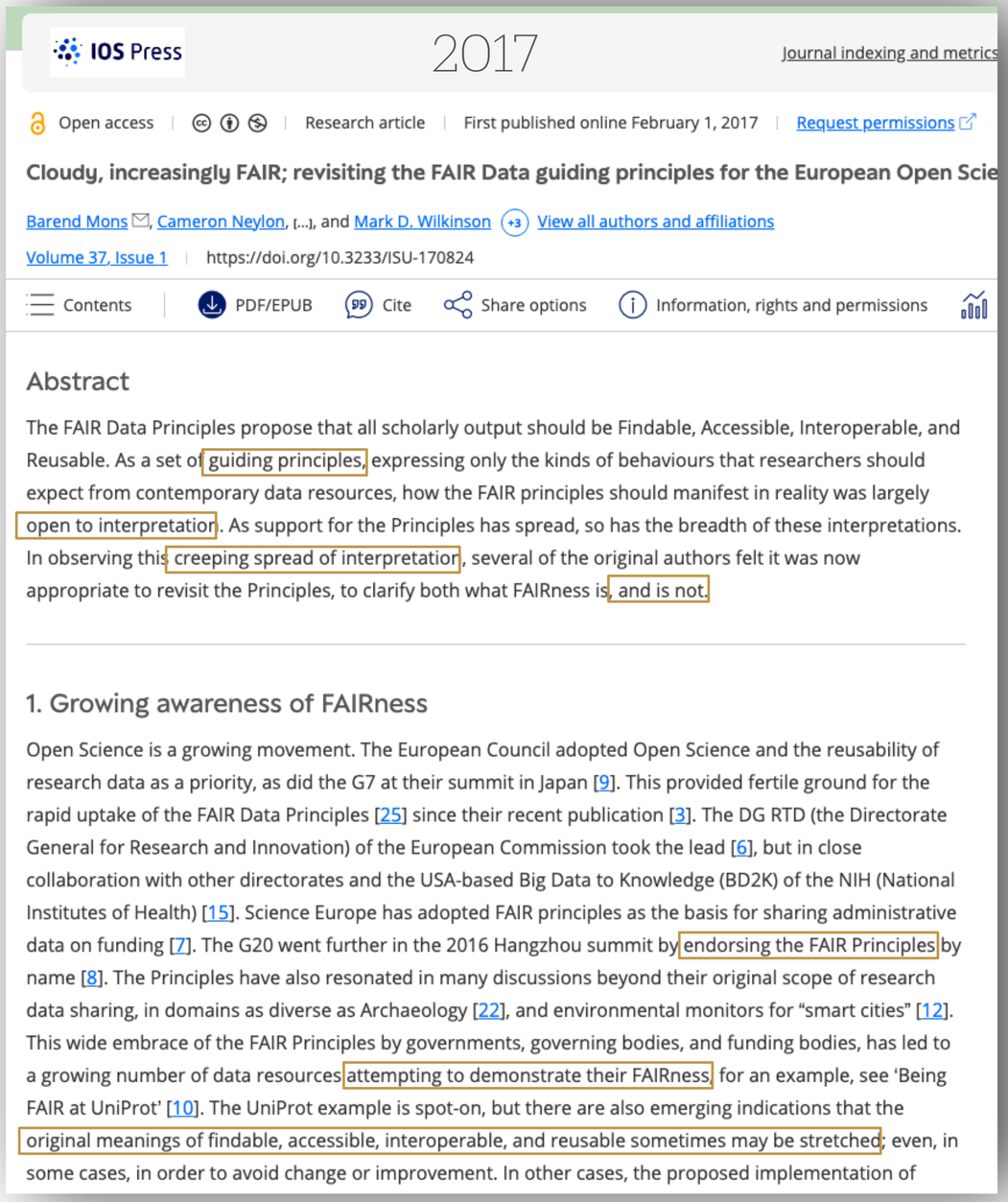
Scientific Data **3**, Article number: 160018 (2016) | Cite this article

1.02m Accesses | **13k** Citations | **2351** Altmetric | Metrics

An Addendum to this article was published on 19 March 2019

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 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.



IOS Press 2017 Journal indexing and metrics

Open access | Research article | First published online February 1, 2017 | Request permissions

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

Barend Mons, Cameron Neylon, [...], and Mark D. Wilkinson +3 View all authors and affiliations

Volume 37, Issue 1 | https://doi.org/10.3233/ISU-170824

Contents | PDF/EPUB | Cite | Share options | Information, rights and permissions

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.

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 environmental monitors for "smart cities" [12]. This wide embrace of the FAIR Principles by governments, governing bodies, and funding bodies, has led to a growing number of data resources attempting to demonstrate their FAIRness, for an example, see 'Being FAIR at UniProt' [10]. The UniProt example is spot-on, but there are also emerging indications that the original meanings of findable, accessible, interoperable, and reusable sometimes may be stretched; even, in some cases, in order to avoid change or improvement. In other cases, the proposed implementation of

The Ogden Triangle – Concepts versus words

Unique ID
Concept

Http://GOFAIR/UUID.??

Token or word or icon:

"cancer"
Malignant Neoplasms
Krebskrankheit
C0-265
Etc.

object, entity, defined meaning

The relations between the corners:

1. **Object** evokes **Concept** (in writer's or speaker's mind)
2. Writer/speaker uses **Token** to refer to **Object**
3. **Token** evokes **Concept** (in reader's or listener's mind)
4. Reader/listener refers **Token** back to **Object**

Data Intelligence 2020

Volume 2, Issue 1-2
Winter-Spring 2020

‹ Previous Article Next Article ›

Article Contents

Abstract
1. INTRODUCTION
2. FROM INTERPRETATION TO IMPLEMENTATION
3. INTERPRETATIONS AND IMPLEMENTATION CONSIDERATIONS PER FAIR GUIDING PRINCIPLE
4. DISCUSSION
ACKNOWLEDGEMENTS
Notes
REFERENCES

January 01 2020

FAIR Principles: Interpretations and Implementation Considerations

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffery, Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Barbara Magagna, Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reisen, Philippe Rocca-Serra, Robert Pergl, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Schneider, George Strawn, Mark Thompson, Andra Waagmeester, Tobias Weigel, Mark D. Wilkinson, Egon L. Willighagen, Peter Wittenburg, Marco Roos, Barend Mons ✉, Erik Schultes

Data Intelligence (2020) 2 (1-2): 10–29.

https://doi.org/10.1162/dint_r_00024

Abstract

The FAIR principles have been widely cited, endorsed and adopted by a broad range of stakeholders since their publication in 2016. By intention, the 15 FAIR guiding principles do not dictate specific technological implementations, but provide *guidance* for improving Findability, Accessibility, Interoperability and Reusability of digital resources. This has likely contributed to the broad adoption of the FAIR principles, because individual stakeholder communities can implement their own FAIR solutions. However, it has also resulted in inconsistent interpretations that carry the risk of leading to incompatible implementations. Thus, while the FAIR principles are formulated on a high level and may be interpreted and implemented in different ways, for true interoperability we need to support convergence in implementation choices that are widely accessible and (re)-usable. We introduce the concept of *FAIR implementation considerations* to assist accelerated global participation and convergence towards accessible, robust, widespread and consistent FAIR implementations. Any self-identified stakeholder community may either *choose* to reuse solutions from existing implementations, or when they spot a gap, accept the *challenge* to create the needed solution, which, ideally, can be used again by other communities in the future. Here, we provide interpretations and implementation considerations (choices and challenges) for each FAIR principle.

Keywords: FAIR guiding principles, FAIR implementation, FAIR convergence, FAIR communities, choices and challenges

1. INTRODUCTION

The notion of good data stewardship (i.e., maximizing the opportunities for the efficient discovery and reuse of research outputs) has been around for decades and many implementation choices have already been made by pioneering communities to extend stewardship with the notion of machine-actionability. The FAIR principles can be seen as a consolidation of these earlier efforts and emerged

Interpretations

The FAIR Guiding Principles [1] provide guidance when improving Findability, Accessibility, Interoperability and Reusability of digital resources. But they do not dictate specific technological implementations. The GO FAIR Foundation believes that what ever FAIR implementation choices are made, they should always ensure, as much as possible, interoperability, machine-actionability, global participation and convergence towards accessible, robust, widespread and consistent FAIR implementations. Towards this end, the GO FAIR Foundation has consolidated from the community of FAIR experts, explicit interpretations of the FAIR Principles and implementation considerations. The aim is to provide a reference for continuing coherent dialogue on "what FAIR is" and a target, with minimal guarantees on machine-actionability, to which the community can confidently build towards. Following closely Jacobsen et al [2], the GO FAIR Foundation's interpretations are provided here as referenceable webpages:

Interpretation of R1.3

Where community standards or best practices for data archiving and sharing exist, they should be followed. Several disciplinary communities have defined Minimal Information Standards describing most often the minimal set of metadata items required to assess the quality of the data acquisition and processing and to facilitate reproducibility. Such standards are a good start, noting that true (interdisciplinary) reusability will generally require richer metadata. For a list of such standards, consult for instance FAIRsharing. The required richness of the provenance metadata will be strongly dependent on the norms generated and agreed upon in the most related research communities.

Interpretation of F1

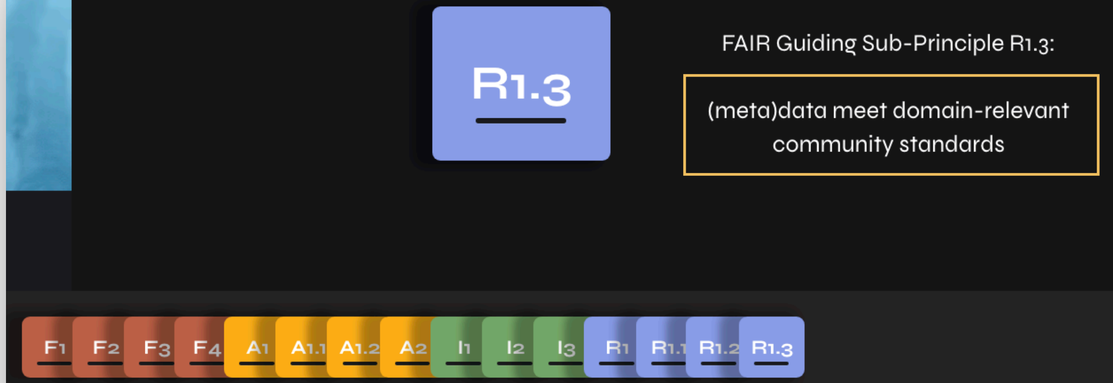
Principle F1 states that digital resources, i.e., data and metadata, must be assigned a globally unique and persistent dentifier which serves as a permanent machine interpretable reference. The GO FAIR Foundation emphasises the need for persistence and global uniqueness, as well the property of resolvability of the identifiers (see also A1). Globally unique means that the identifier is guaranteed to unambiguously refer to the intended resources (where 'global' is intended to mean 'universal' as there are described digital assets outside the 'world'). Therefore, it is insufficient for it to be unique only locally (e.g. unique within a single, local database). Persistence refers to the

Interpretation of A1.2

This principle clearly demonstrates that following the FAIR guiding principles is not equal to making all data 'open'. Some digital resources, such as data that have access restrictions based on ethical, legal or contractual constraints, require additional conditions/steps to be accessed. This often pertains to assuring that the access requester is indeed that requester (authentication), that the requester's profile and credentials match the access conditions of the resource (authorization), and that the intended use matches permitted use cases (e.g. for a particular purpose only) (see also R1.1, where there are requirements to provide explicit documentation about who may use the data, and for what purposes). At the level of technical implementation, an additional authentication and authorization procedure must be specified, if it is not already defined by the protocol (see A1.1). A requester can be a human or a machine agent. In the latter case it is probably a proxy for a human or an organization to which the authentication

Interpretation of I2

In Principle I2 we referred to "vocabularies" as the methods that unambiguously represent concepts that exist in a given domain. The use of shared, and formally structured (principle I1), sets of terms is an essential part of FAIR. Terminology systems, including flat "vocabularies", hierarchical "thesauri" and more granular specifications of knowledge such as data models and consistently structured ontologies, play an important role in community standards. However, the vocabularies used for metadata or data also need to be findable, accessible, interoperable, and reusable in their own right so that users (including machines) can fully understand the meaning of the terms used in the metadata. This principle has been criticized as "circular" but as has been made clear earlier in the Digital Intelligence article, the simple use of a "label" (e.g. "temperature") is insufficient to enable a machine to understand both the intent of that label (Body temperature? Melting temperature?) and the contexts



Interpretation of R1.3

Where community standards or best practices for data archiving and sharing exist, they should be followed. Several disciplinary communities have defined Minimal Information Standards describing most often the minimal set of metadata items required to assess the quality of the data acquisition and processing and to facilitate reproducibility. Such standards are a good start, noting that true (interdisciplinary) reusability will generally require richer metadata. For a list of such standards, consult for instance FAIRsharing. The required richness of the provenance metadata will be strongly dependent on the norms generated and agreed upon in the most related research communities.

FAIR Implementation Profiles

Tell the world

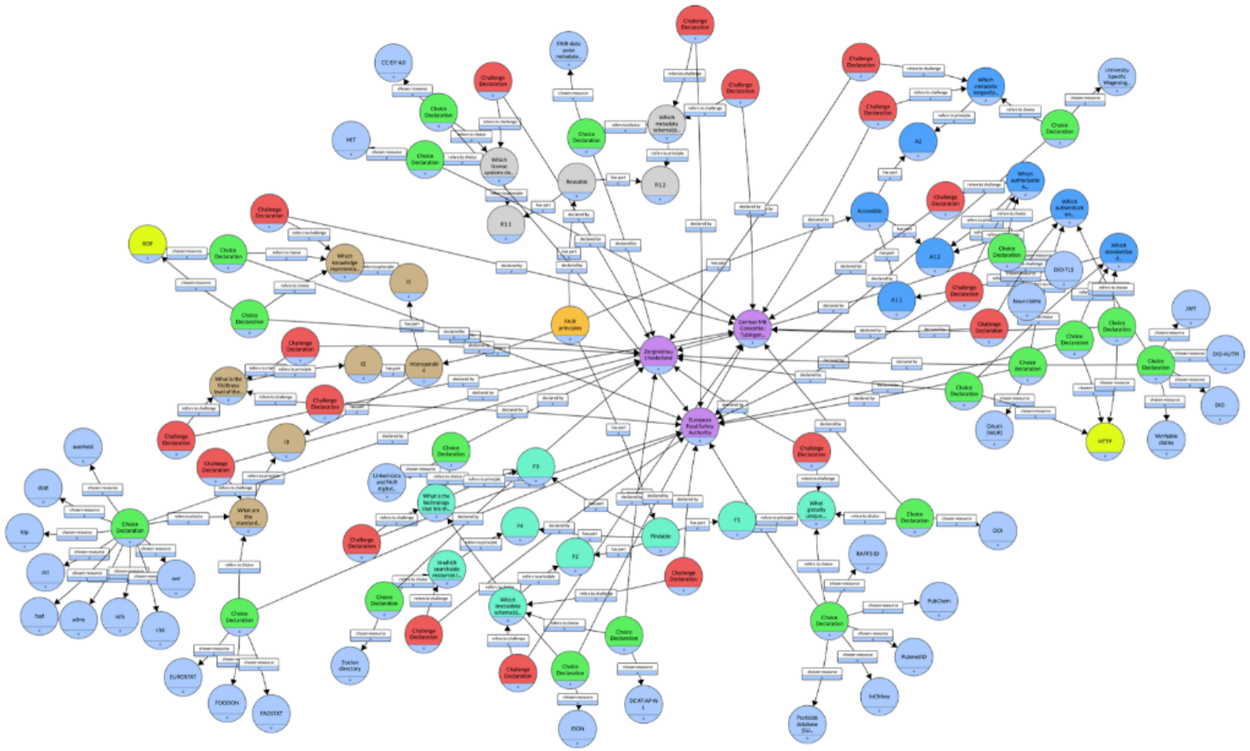
Which choices you made !

download the questionnaire in **PDF**.

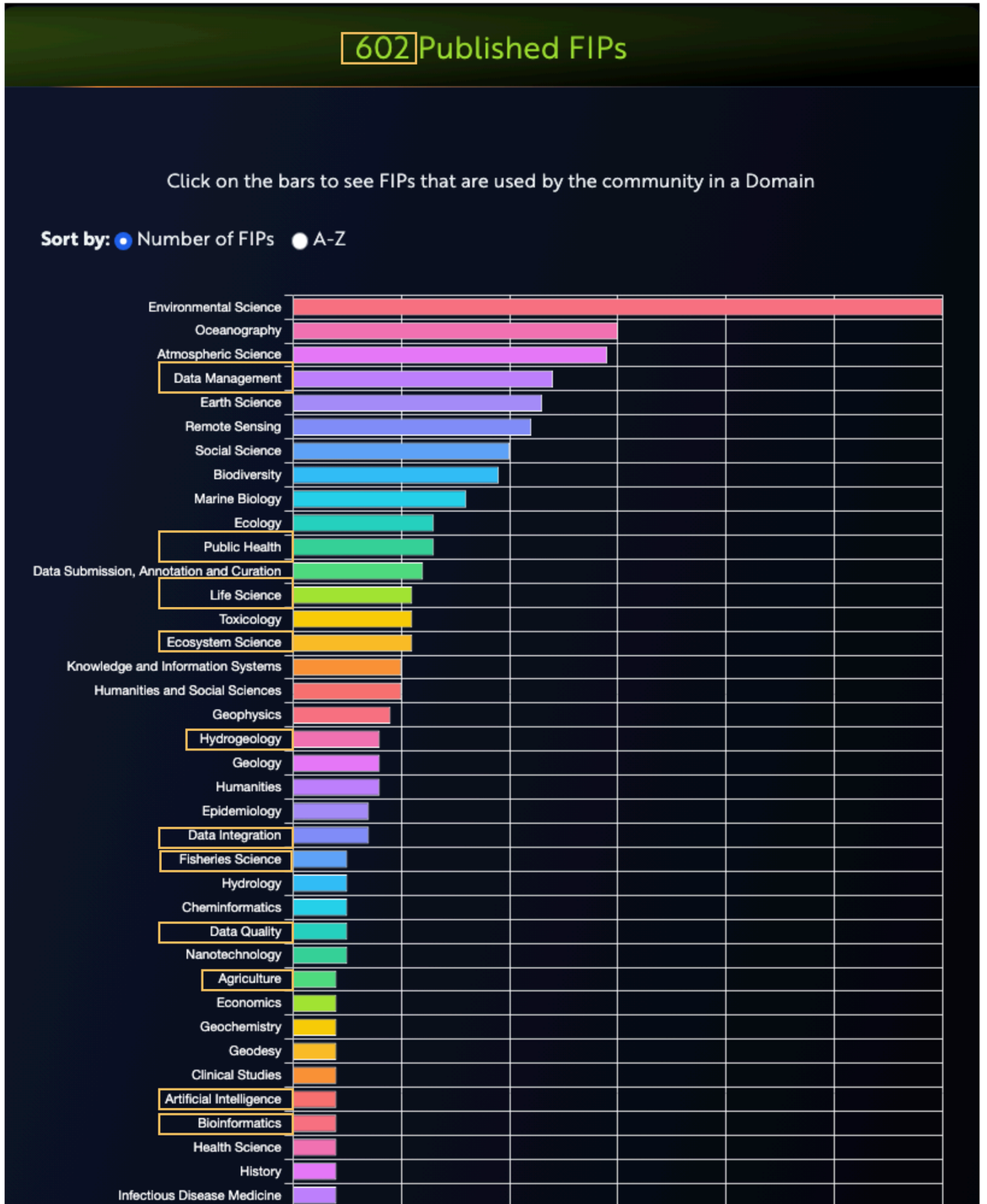
FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

Reading material

Magagna, B, et al. 2020. **Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence.**
OSF Preprints: **https://doi.org/10.31219/osf.io/2p85g**
Publication: **https://link.springer.com/chapter/10.1007/978-3-030-65847-2_13**



And Published !



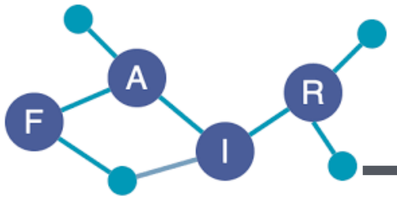
GO FAIR

- **2018**
  - Rigorous adherence to the **FAIR Principles**
  - As **open** as possible
  - **Decentralized** architectures as a default
  - **Hourglass** architectures
  - Aiming to **mitigate vendor lock-in**

FAIR & Equitable Science

- **2024**
  - **Machine-actionability** of data
  - **Equitability** of data
  - Data that are **Fully AI Ready**

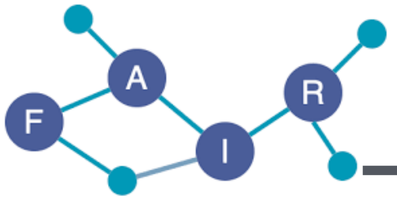
LLM is not AI!



1
Conceptual
Model

It's all hallucination, but we only **call it that when we notice it's wrong**.
The problem is, large language models are so good at what they do that what they make up looks right most of the time. And that makes trusting them hard.

LLM is not AI



1
Conceptual
Model

Hybrid Intelligence support is what scientists are waiting for !

BETA VERSION

Structure-Based Prediction of SARS-CoV-2 Variant Properties Using Machine Learning on Mutational Neighborhoods

Max van den Boom, Erik Schultes, Thomas Hankemeier

More info

0 / 5 0 reviews

Abstract | Data Explorer | Data Package | Funding | Certification

Keywords:

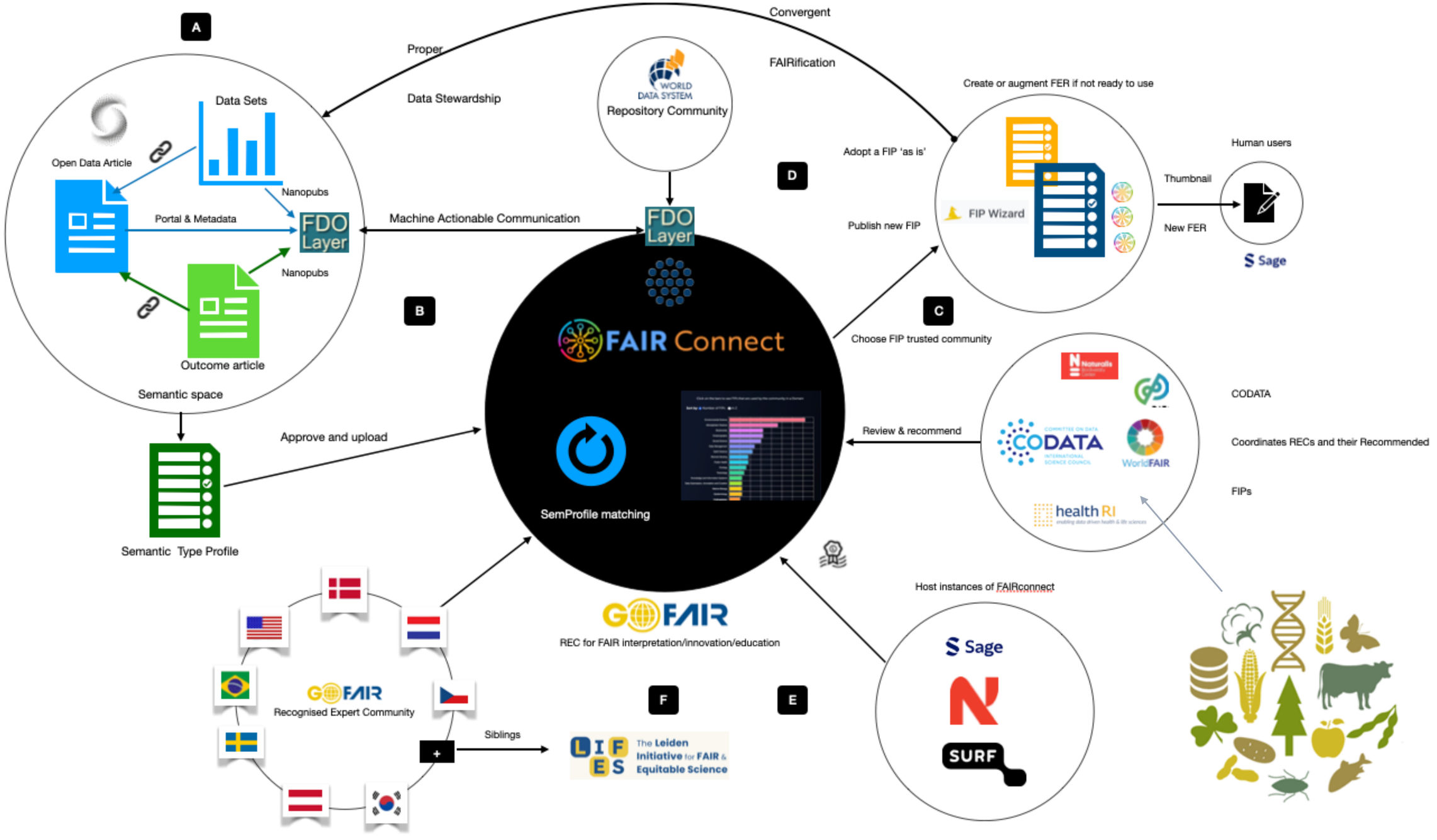
SARS-CoV-2, Spike protein, Receptor-binding domain (RBD), Protein structure prediction, AlphaFold2, ESMFold, Deep mutational scanning, Variant of Concern (VoC)

Abstract

This dataset presents a structure-enriched resource of theoretical and empirical SARS-CoV-2 spike receptor-binding domain (RBD) variants, developed under the STAYAHEAD project for pandemic preparedness. It integrates large-scale in silico structure predictions with empirical biophysical measurements. The dataset includes 3,705 single-point Wuhan-Hu-1 RBD variants and 100 higher-order Omicron BA.1/BA.2 variants, annotated with AlphaFold2 and ESMFold metrics and Bio2Byte sequence-based predictors. Structural descriptors—RMSD, TM-score, plDDT, solvent accessibility, hydrophobicity, aggregation propensity—are linked to ACE2 binding and expression data from deep mutational scanning. Provided as a FAIR² Data Package, it supports structure–function analysis, variant modeling, and responsible reuse in virology, structural biology, and computational protein science.

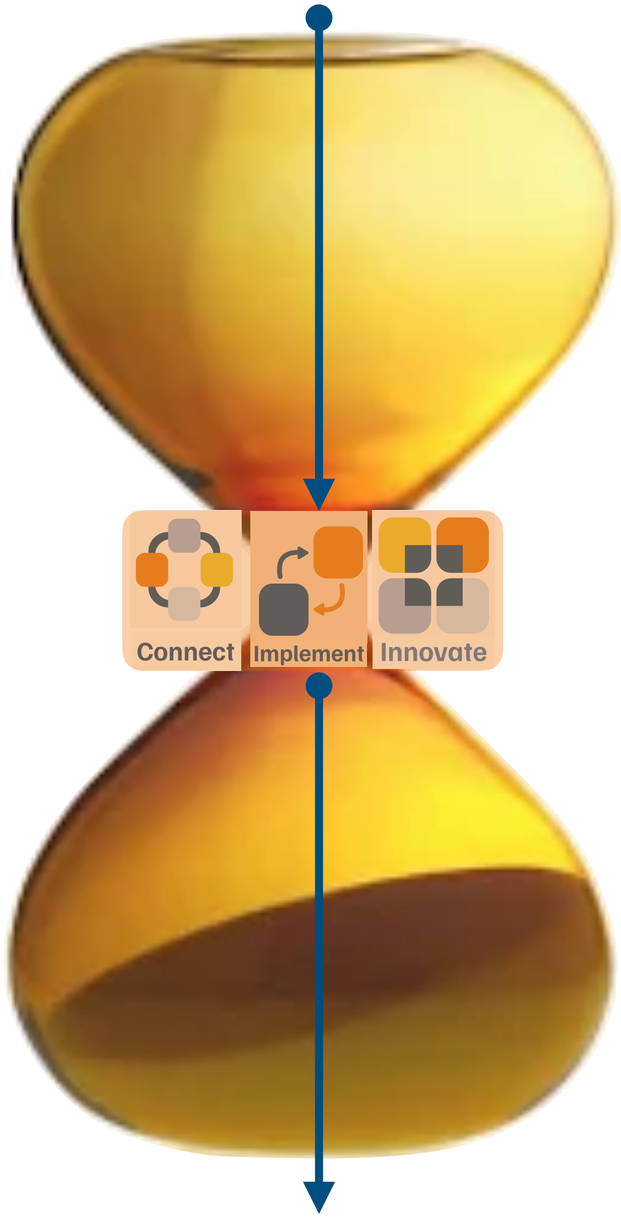
Data Explorer

Reset selection


A
Data Sets
Open Data Article
Nanopubs
Portal & Metadata
FDO Layer
Nanopubs
Outcome article
Semantic space
Semantic Type Profile
Approve and upload
Proper
Data Stewardship
Convergent
FAIRification
WORLD DATA SYSTEM
Repository Community
Machine Actionable Communication
FDO Layer
B
FAIR Connect
SemProfile matching
D
Adopt a FIP 'as is'
Publish new FIP
Create or augment FER if not ready to use
FIP Wizard
Thumbnail
New FER
Human users
Sage
C
Choose FIP trusted community
Review & recommend
CODATA
Coordinates RECs and their Recommended
FIPs
health RI
GO FAIR
REC for FAIR interpretation/innovation/education
Host instances of FAIRconnect
Sage
SURF
E
F
GO FAIR
Recognised Expert Community
+
Siblings
The Leiden Initiative for FAIR & Equitable Science

Core activity LIFES: **FAIR Implementation Guidance (FIG)** by and for its members

The **Leiden Initiative** for **FAIR** & **Equitable Science**



LIFES is a global association that unites public and private data owners, users, ASPs and communities to jointly evolve the ecosystem for FAIR, equitable and sustainable data reuse for science & innovation

Core goal: **Harmonisation** and **Defragmentation** for its members and beyond

LIFES conference 2025

We look back on a great LIFES conference which focused on **jointly evolving the ecosystem for FAIR & equitable data reuse.** Below a short impression video and an **interview with Barend Mons**, founding board member of LIFES and **Robbert Dijkgraaf** the opening speaker at the conference.

Impressions

Interviews

www.lifes.institute

Conference Report

Pesthuislaan 7

Leiden

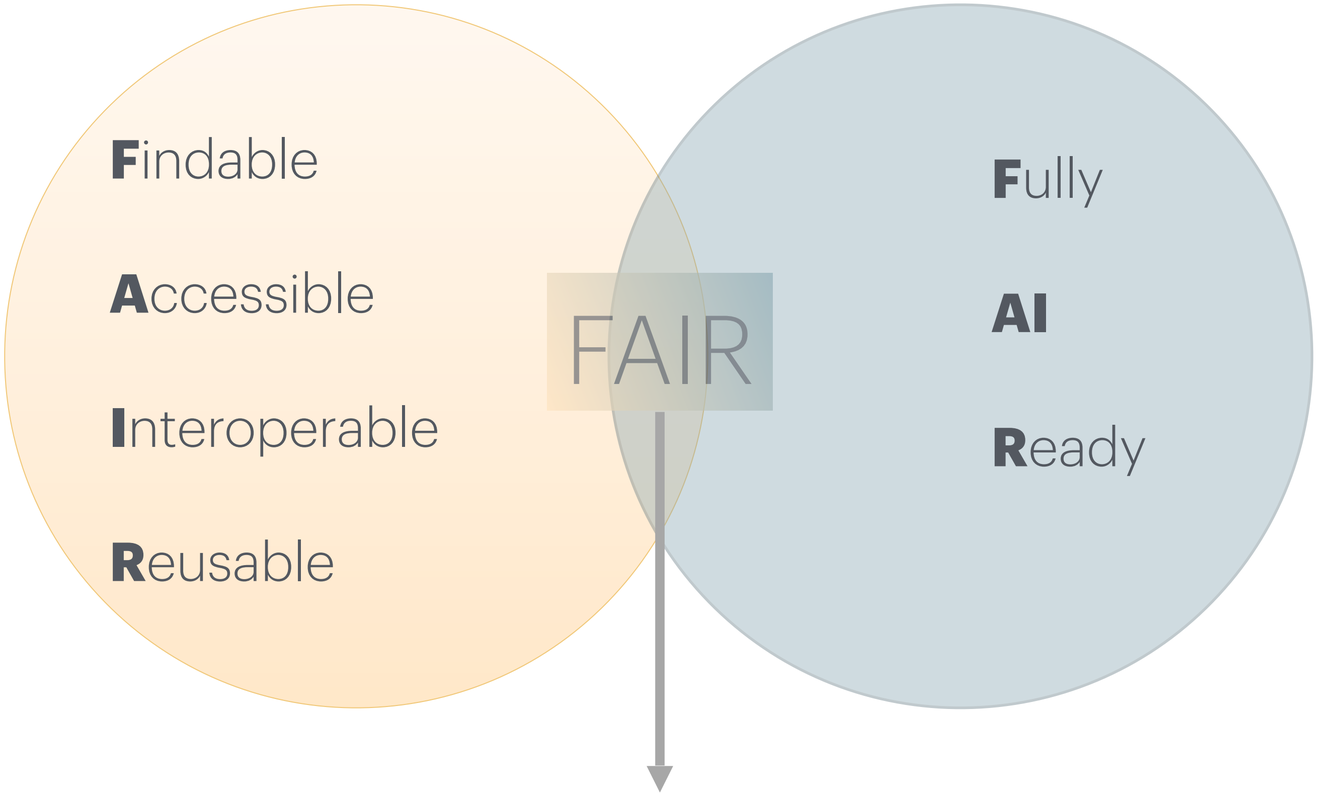
Mailadress

Data remains
at the source

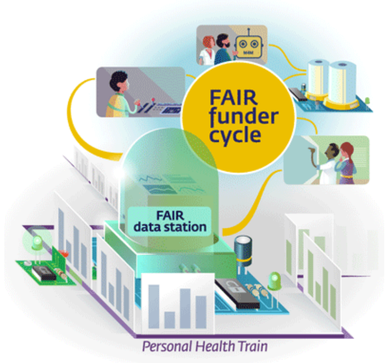
- **Experimental Data**
- **Real-World Data**
- **Established Knowledge**

The end of data sharing

The Dawn of data Visiting



The Machine Knows what we mean



References

https://link.springer.com/article/10.1186/1471-2105-6-142

https://link.springer.com/article/10.1186/gb-2008-9-s2-s7

https://www.nature.com/articles/ng0411-281

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007894

https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.10392

https://academic.oup.com/bioinformatics/article/21/9/2049/409019