Universitat Pompeu Fabra Barcelona

Historic England

Towards FAIR phytolith data

First steps down a long and winding road

Emma Karoune, Carla Lancelotti, Javier Ruiz-Pérez, Juanjo García-Granero, Marco Madella

DOI:10.5281/zenodo.5336872

EOSC-Life

What is FAIR?

The **FAIR principles**:

Findable

Accessible

Interoperable

Reusable

DOI:10.5281/zenodo.5336872

FAIR data analogy

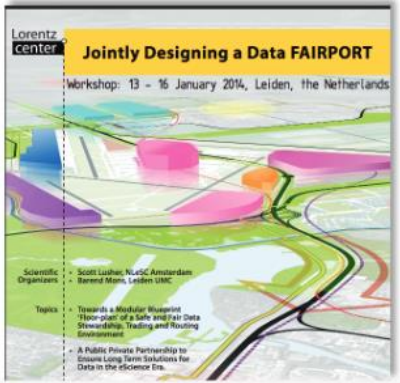
Annotation makes it easier to find important things

You would not buy food with no labels!

Labels make different foods easier to find and access in stores, combine with other foods (interoperable) and use in different ways.

Adapted from talk by Philippe Rocca-Serra (2020)

DOI:10.5281/zenodo.5336872



2014

2016

SCIENTIFIC DATA

SPRINGER NATURE

OPEN

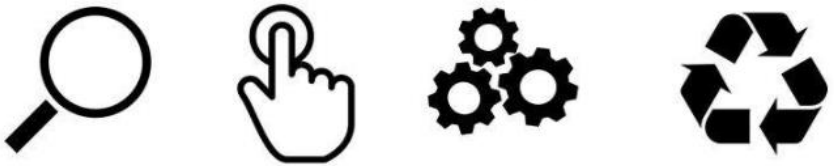
SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: 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, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

Findable **A**ccessible **I**nteroperable **R**eusable



A set of principles to enhance the value of all digital resources

Developed and **endorsed** by *researchers, service providers, publishers, funding agencies and industry partners*

FAIR principles from **Wilkinson *et al*. (2016)**
DOI: 10.1038/sdata.2016.18

FAIR does not mean Open

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

Taken from Wilkinson et al. 2016.

DOI:10.5281/zenodo.5336872

“As open as possible, as closed as necessary”

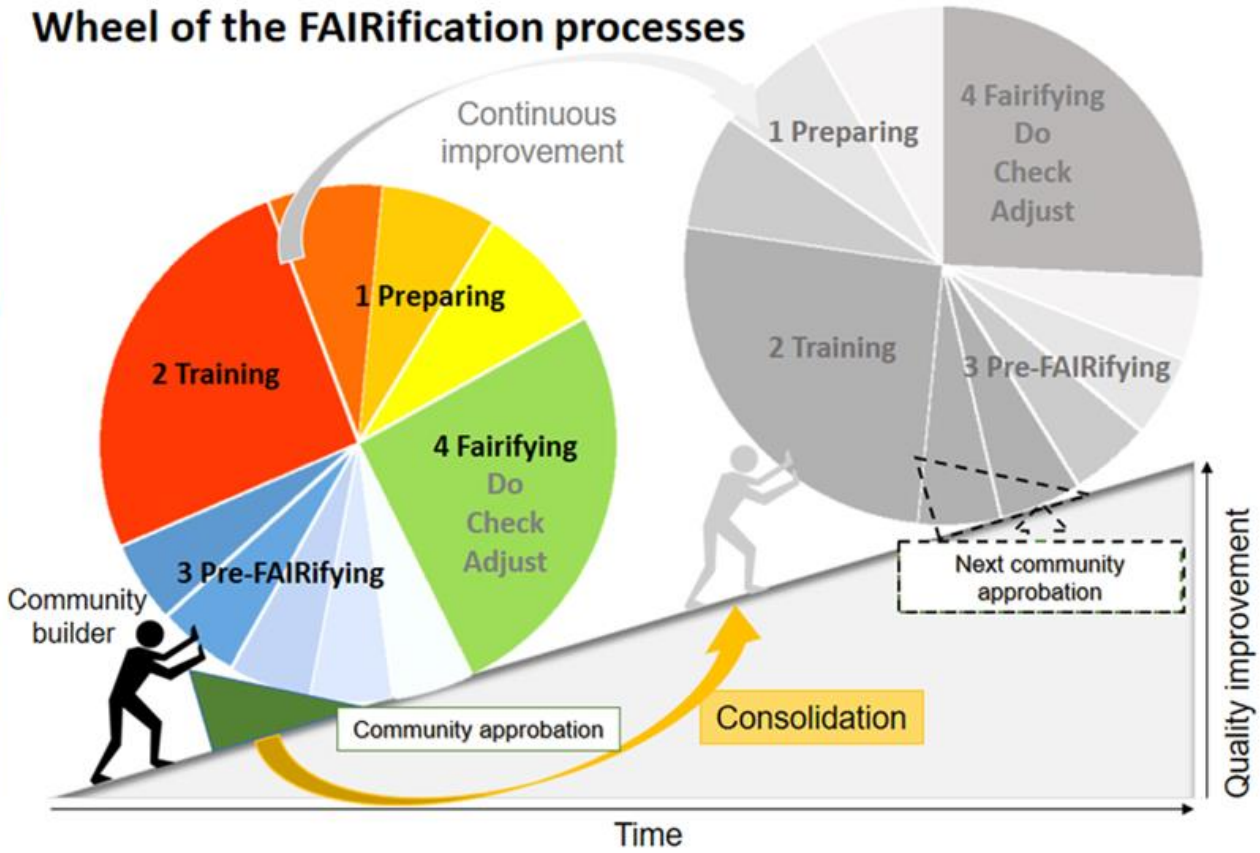
C - Collective benefit
A - Authority to control
R - Responsibility
E - Ethics

F - Findable
A - Accessible
I - Interoperable
R - Reusable

Carroll *et al.* 2021, https://doi.org/10.1038/s41597-021-00892-0

DOI:10.5281/zenodo.5336872

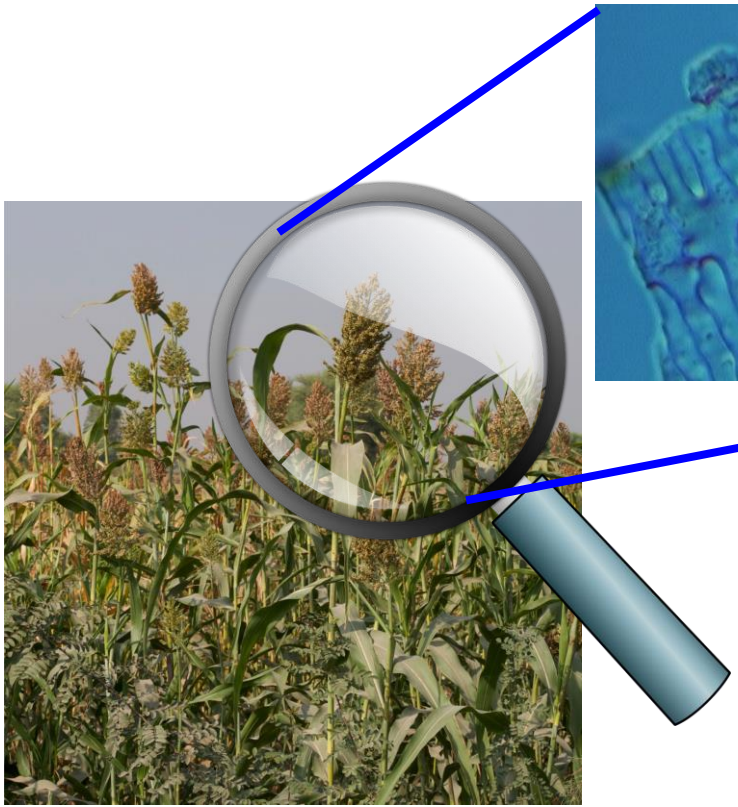
Process	Steps
Preparing FAIRification	Explain FAIRification
	Define constraints
	Define advantages
Training	Increase FAIR literacy Convince partners
Pre-FAIRifying	Building shared strategy
	Define community
	Define objects and variables
	Select items to be identified
	Analyse common denominators
FAIRifying	**Do:** Downward levelling
	Check: first interoperations
	Adjust: Identifying gaps and new expectation

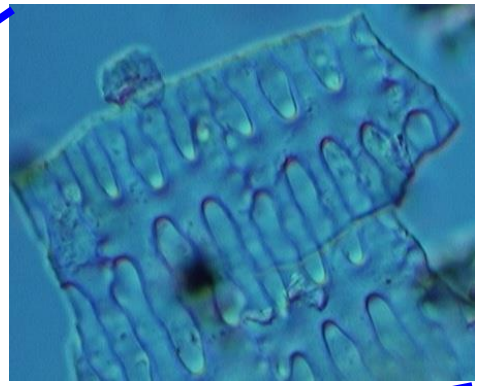


Taken from David *et al.* 2020 - FAIRness literacy

DOI:10.5281/zenodo.5336872

Phytoliths… why do we need **FAIRification?**





International Code for Phytolith Nomenclature (ICPN) 2.0 FREE

International Committee for Phytolith Taxonomy (ICPT)

Annals of Botany, Volume 124, Issue 2, 24 July 2019, Pages 189–199,
https://doi.org/10.1093/aob/mcz064

Published: 24 September 2019 **Article history** ▾

International Phytolith Society

DOI:10.5281/zenodo.5336872

Phytoliths... why do we need FAIRification?

Karoune 2020 - Review of Open Science Practices in Phytolith Research

- **341 articles** with primary phytolith data.

Collected data on:

- data format,
- reusability of data,
- photo inclusion,
- fully described method,
- use of standard nomenclature (ICPN 1.0),
- open access.

Results:

- Data sharing in any form = 53%
- Reusable raw data = 4%

Data paper - http://doi.org/10.5334/joad.67 **Pre-print of research articles** - https://osf.io/fa7q3/

Growing evidence that open practices have benefits

1. Research quality - Palaeogenetics
2. Citation benefits of linking data and open access

RESEARCH ARTICLE

PLOS ONE

When Data Sharing Gets Close to 100%: What Human Paleogenetics Can Teach the Open Science Movement

Paolo Anagnostou[1,2]*, Marco Capocasa[2,3], Nicola Milia[4], Emanuele Sanna[4], Cinzia Battaggia[1], Daniela Luzi[5], Giovanni Destro Bisol[1,2]*

97% data sharing in 162 article.

DOI:10.5281/zenodo.5336872

Growing evidence that open practices have benefits

1. Research quality - Palaeogenetics
2. Citation benefits of linking data and open access

RESEARCH ARTICLE

PLOS ONE

The citation advantage of linking publications to research data

Giovanni Colavizza[1,2], Iain Hrynaszkiewicz[3,4], Isla Staden[1,5], Kirstie Whitaker[1,6], Barbara McGillivray[1,6]*

25% higher citation impact for papers that include a link to data in a repository

DOI:10.5281/zenodo.5336872

What are the advantages of making our data FAIR and open?

For our wider community

1. Increased quality of research
2. Increased collaboration in our community
3. More sustainable research for the next generation
4. More diverse and inclusive research community

For you

1. Increased citations of your published research articles.
2. Greater discoverability and enhanced visibility of your work.
3. Credit for all your work.

DOI:10.5281/zenodo.5336872

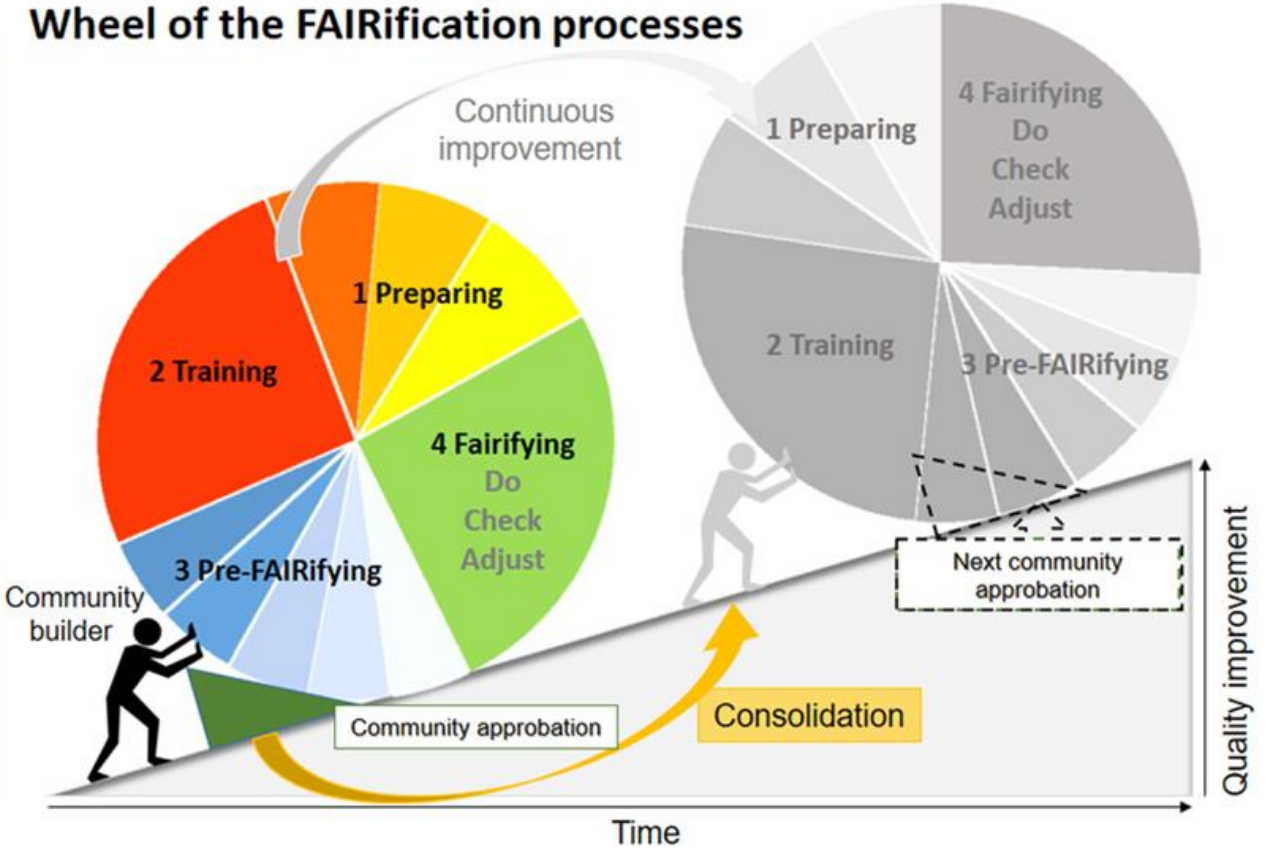
If you want to start now with FAIR and open data

In your next article:

1. Make sure it is open access - gold or green.
2. Include a fully described method (protocol).
3. Write a data availability statement - describe clearly how to access the full raw data.
4. Put your data in an open repository and give it a license.

DOI:10.5281/zenodo.5336872

Process	Steps
Preparing FAIRification	Explain FAIRification
	Define constraints
	Define advantages
Training	Increase FAIR literacy Convince partners
Pre-FAIRifying	Building shared strategy
	Define community
	Define objects and variables
	Select items to be identified
	Analyse common denominators
FAIRifying	**Do:** Downward levelling
	Check: first interoperations
	Adjust: Identifying gaps and new expectation



Taken from David *et al.* 2020 - FAIRness literacy

DOI:10.5281/zenodo.5336872

What has this project done so far?

1. Created a **community survey**
2. Opened a project **repository** on GitHub
3. Designed a **webpage**, using GitHub
4. Planning **FAIR assessment** of two large datasets

DOI:10.5281/zenodo.5336872

Moving forward

Autumn 2021

- Start FAIR assessment
- Close survey and analyses results
- Start reaching out to interested researchers

Winter 2021

- Offer training - onboarding into the project - Github and FAIR assessment.
- Finalise FAIR assessment datasets.

DOI:10.5281/zenodo.5336872

Moving forward

Spring 2022

- Analysing results of FAIR assessment
- Drawing up first draft of guidelines
- Consulting wider community on guidelines

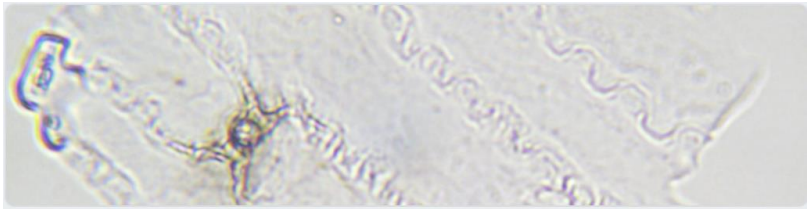
Summer 2022

- Writing data paper
- Writing community FAIR guidelines

DOI:10.5281/zenodo.5336872

How you can get involved

1. **Complete our survey**
2. Get involved:
   a. **Training** in Github and FAIR
   b. **Help with our work** - translation, review of FAIR guidelines, open and FAIR publications.



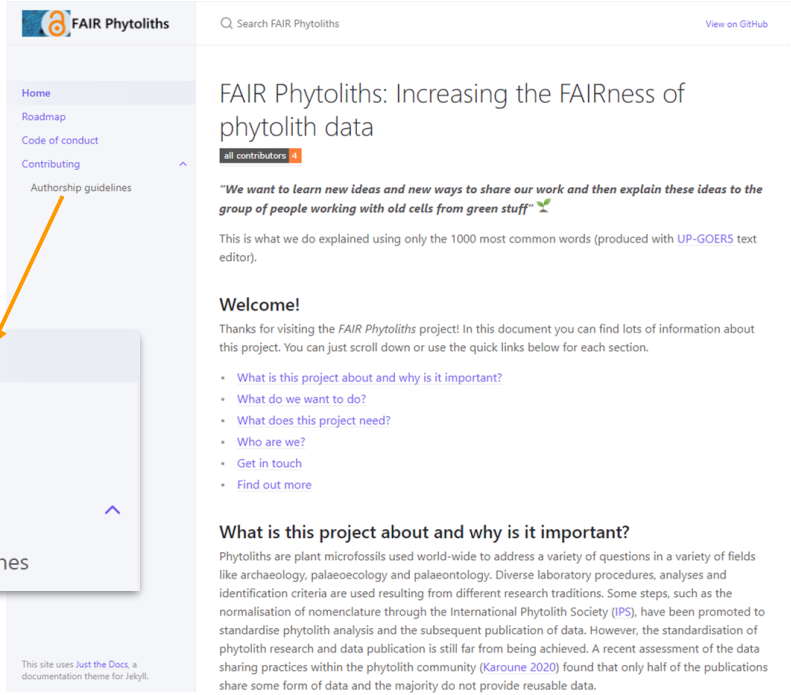
COMMUNITY SURVEY - FAIR Phytoliths: Increasing the FAIRness of phytolith data

Hello! This short questionnaire is aimed at understanding current practices concerning open science within the phytolith community as part of the project FAIR Phytoliths: Increasing the FAIRness of phytolith data. The project strives to make phytolith data more FAIR (Findable, Accessible, Interoperable and Reusable). This initiative is supported by the International

- Survey Part 1
- Survey Part 2

DOI:10.5281/zenodo.5336872

Webpage and Twitter:
Look out for more news from our project



@open_phytoliths

https://open-phytoliths.github.io/FAIR-phytoliths/

DOI:10.5281/zenodo.5336872

Thanks!

From the core team - Emma Karoune, Carla Lancelotti, Javier Ruiz-Pérez, Juanjo Garcia Granero, Marco Madella

Our new core team member: Celine Kerfant.

Our website: https://open-phytoliths.github.io/FAIR-phytoliths/

Twitter: @open_phytoliths

EOSC-Life

upf. Universitat Pompeu Fabra Barcelona

CaSEs Culture and Socio-Ecological Dynamics research group

Historic England

Open Life Science

International Phytolith Society

DOI:10.5281/zenodo.5336872

References used

- Anagnostou P, Capocasa M, Milia N, Sanna E, Battaggia C, Luzi D, et al. (2015) When Data Sharing Gets Close to 100%: What Human Paleogenetics Can Teach the Open Science Movement. PLoS ONE 10(3): e0121409. doi:10.1371/journal.pone.0121409
- Carroll, S.R., Herczog, E., Hudson, M. *et al.* Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Sci Data* 8, 108 (2021). https://doi.org/10.1038/s41597-021-00892-0
- Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. PLOS ONE 15(4): e0230416. https://doi.org/10.1371/journal.pone.0230416
- David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur, E., Cambon-Thomsen, A. and Alliance – SHAring Reward and Credit (SHARC) Interest Group, T.R.D., 2020. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *Data Science Journal*, 19(1), p.32. DOI: http://doi.org/10.5334/dsj-2020-032
- Karoune, E. (2020, August 16). Pre-print of Assessing Open Science Practices in Phytolith Research. https://doi.org/10.31219/osf.io/fa7q3
- Karoune, E., 2020. Data from "Assessing Open Science Practices in Phytolith Research". *Journal of Open Archaeology Data*, 8(1), p.6. DOI: http://doi.org/10.5334/joad.67
- 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

DOI:10.5281/zenodo.5336872