

Implementing FAIR and Responsible Data Management: The Importance of a Data & Software Management Plan for Your Research Project

Damien Belvèze

2024-10-14

CC-by:4.0 Damien Belvèze



ARDoISE

Atelier rennais
de la donnée

- ARDoISE : Rennes data hub for researchers
- this [presentation files](#) are available online
- speakers notes may be displayed by pressing S key

3 pillars of open science

1. three pillars of Open Science
2. FAIR applied to data
3. FAIR applied to software (as much as possible)

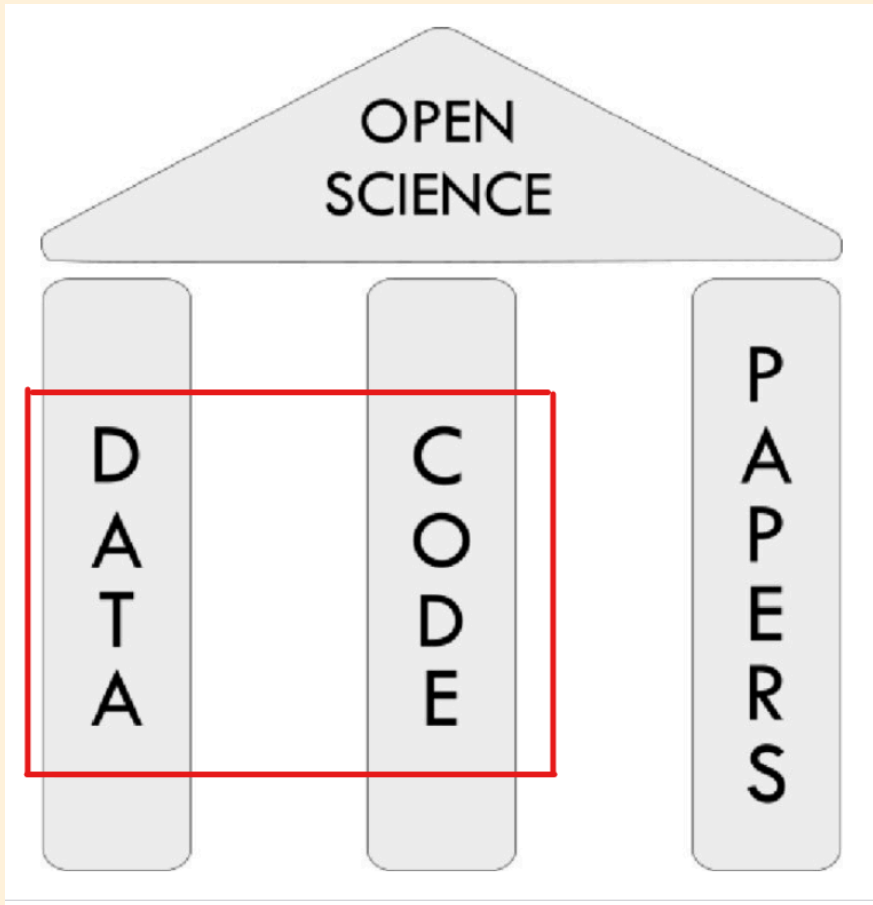


Figure 1



ARDoISE
Atelier rennais
de la donnée

data and software management plans

- information required by funding bodies (ANR, OpenAire)
- first draft must be sent to the funder no later than 6 months after contract is signed
- template available on [DMP OPIDOR](#)
- Contact ARDoISE data hub (through DMP OPIDOR or direct mail to guichet-ardoise@groupe.renater.fr)

what is a paper published without access to the underlying data



remember the [Lancetgate](#).

See also [here](#)

no reproducibility = no public trust in Science

first step: your paper must provide a DAS (Data availability statement)

FAIR principles



1. How to make your data findable
2. How to make your data accessible
3. How to make your data interoperable
4. How to make your data reusable

Findability

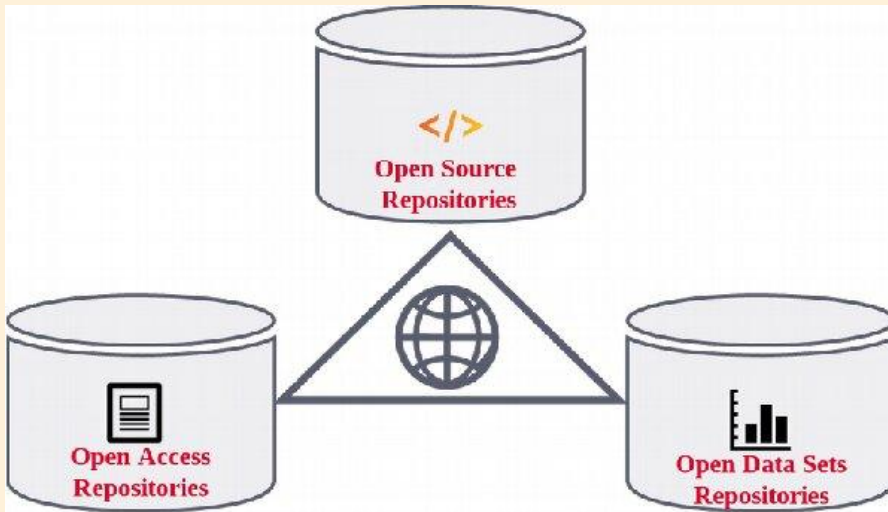


Figure 2

Findability through unique and permanent identifiers

publication	dataset	software
doi, HAL-ID	DOI, Ark	SWHID

Let's have a look to [some of them](#)

Do not trust data repositories that do not provide identifiers to deposited datasets

Papers with a DAS leading to open dataset **get a 25% citation benefit** compared with others  ²

Findability in FAIR principles

FAIR principles  3



Figure 3

Speaking about repositories

- avoid commercial ones
- check licence and policy
- curated by someone ?
- located in Europe (GDPR) ?

COSO guide

recherche data gouv

FAIR principles (2016)



1. How to make your data findable

2. How to make your data accessible

3. How to make your data interoperable

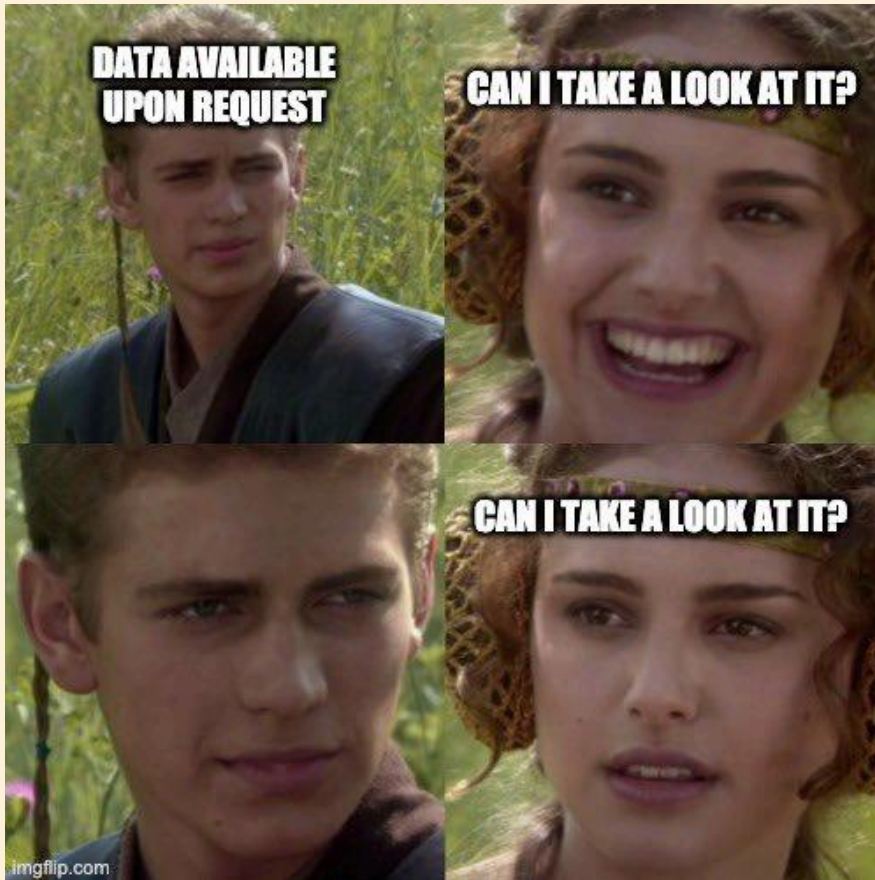
4. How to make your data reusable

As open as possible as closed as necessary

Openness may be limited by:

- GDPR (personal data involved)
- consortium agreement:
 - industrial interest
 - intellectual property (id data are of intellectual nature :
photographs, texts)

Access to data: play by the rules



- A DAS cannot contain: “Data not available” 😡
- if Available Upon Request (AUR): somebody has to answer the request
- n=3416 : no answer or negative answer without any justification 📖 (4)

Figure 4

Interoperability



1. How to make your data findable
2. How to make your data accessible
- 3. How to make your data interoperable**
4. How to make your data reusable

use open formats



- cf. RDM scary tale ([costly birthdays](#))
- choose non proprietary software (forget Microsoft)
- use transparent tools (R is a good option)

reusability



1. How to make your data findable
2. How to make your data accessible
3. How to make your data interoperable
- 4. How to make your data reusable**

provide an open and standard licence

- FR: 2 options are recommended = ODBL / Etalab
- you have the **authorship** but not the **ownership** of the data
- you must be credited (CC:by 4.0 is also possible)
- But, access limitations are regulated (it's not up to you)

what about source code



Figure 5

- scientific integrity debates focus on data sharing practices, not enough on source code replicability
- while data sharing practices are becoming standards, not enough incentive to share source code are provided  ⁷
- only a few publishers ask reviewers to replicate source code (NeurIPS)

Archive and make visible

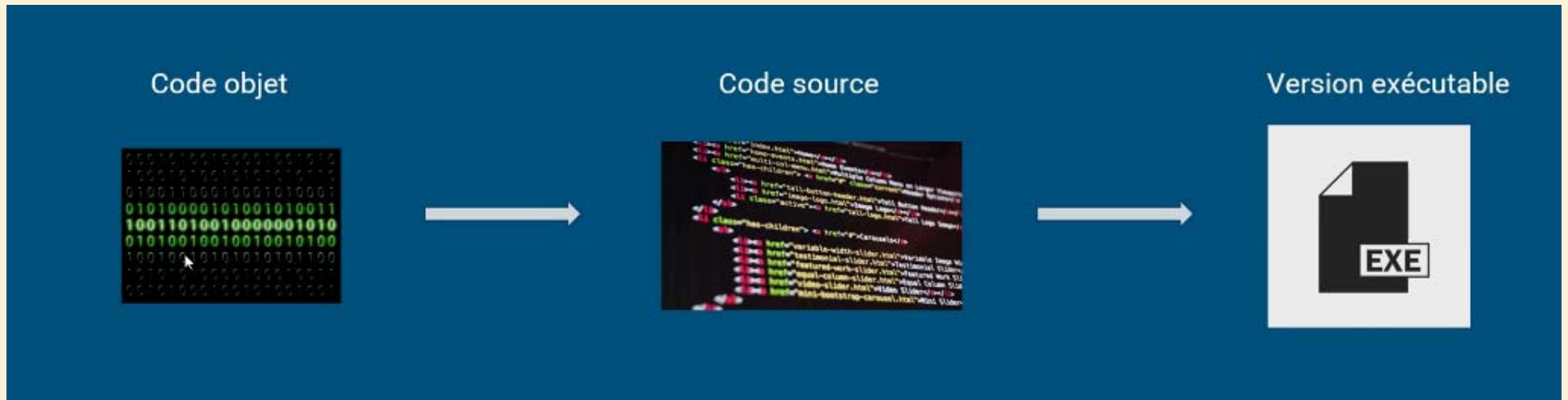


Figure 6

- **source code** (not algorithm or binaries/executable)
- A forge is not an archive (**forges disappear sometimes** - i.e Bitbucket in 2020)

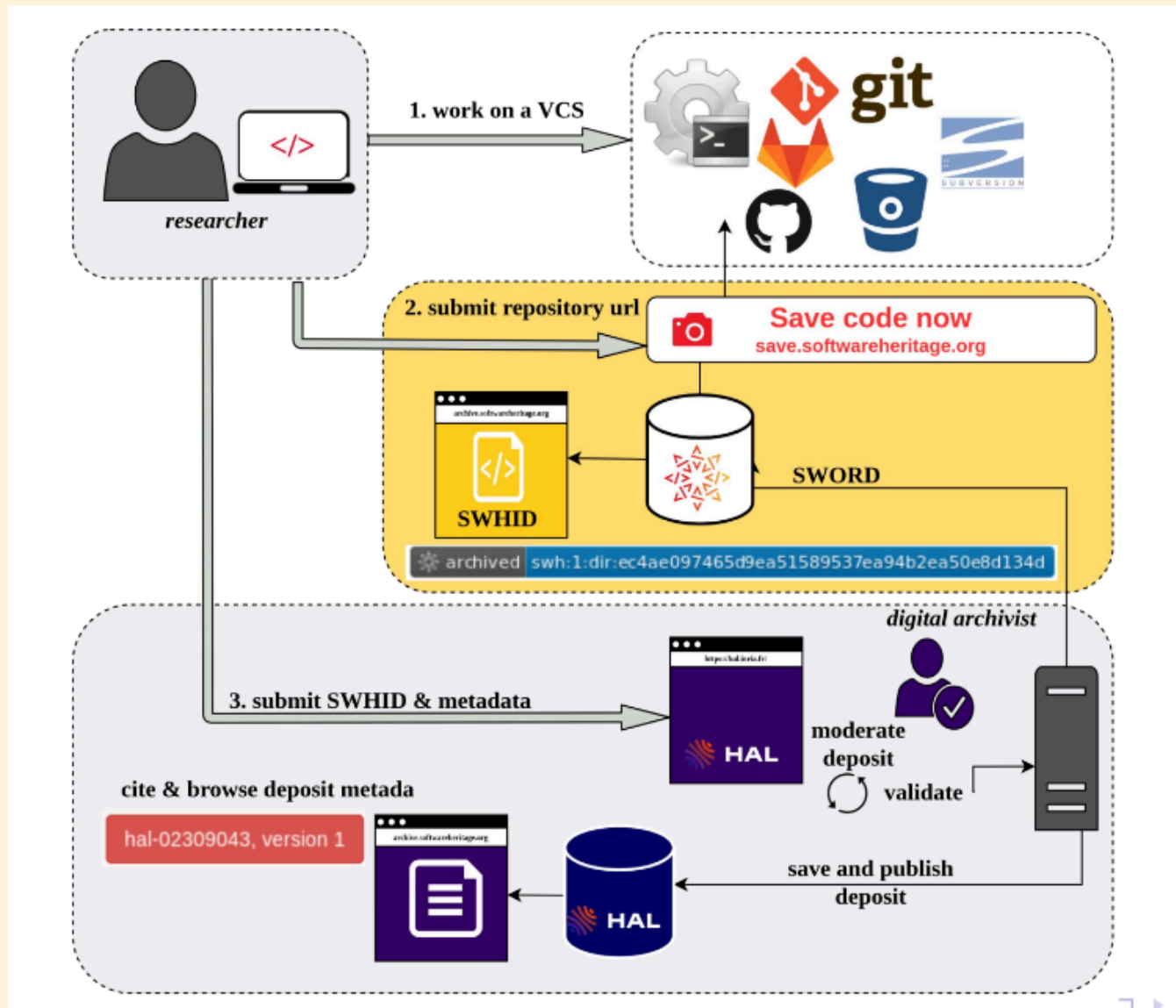


Figure 7

code replicability

save not only the source code files but also:

the software stack within virtual environments (venv/Renv)

the runtime and other Operating system settings through
(Docker) images  8

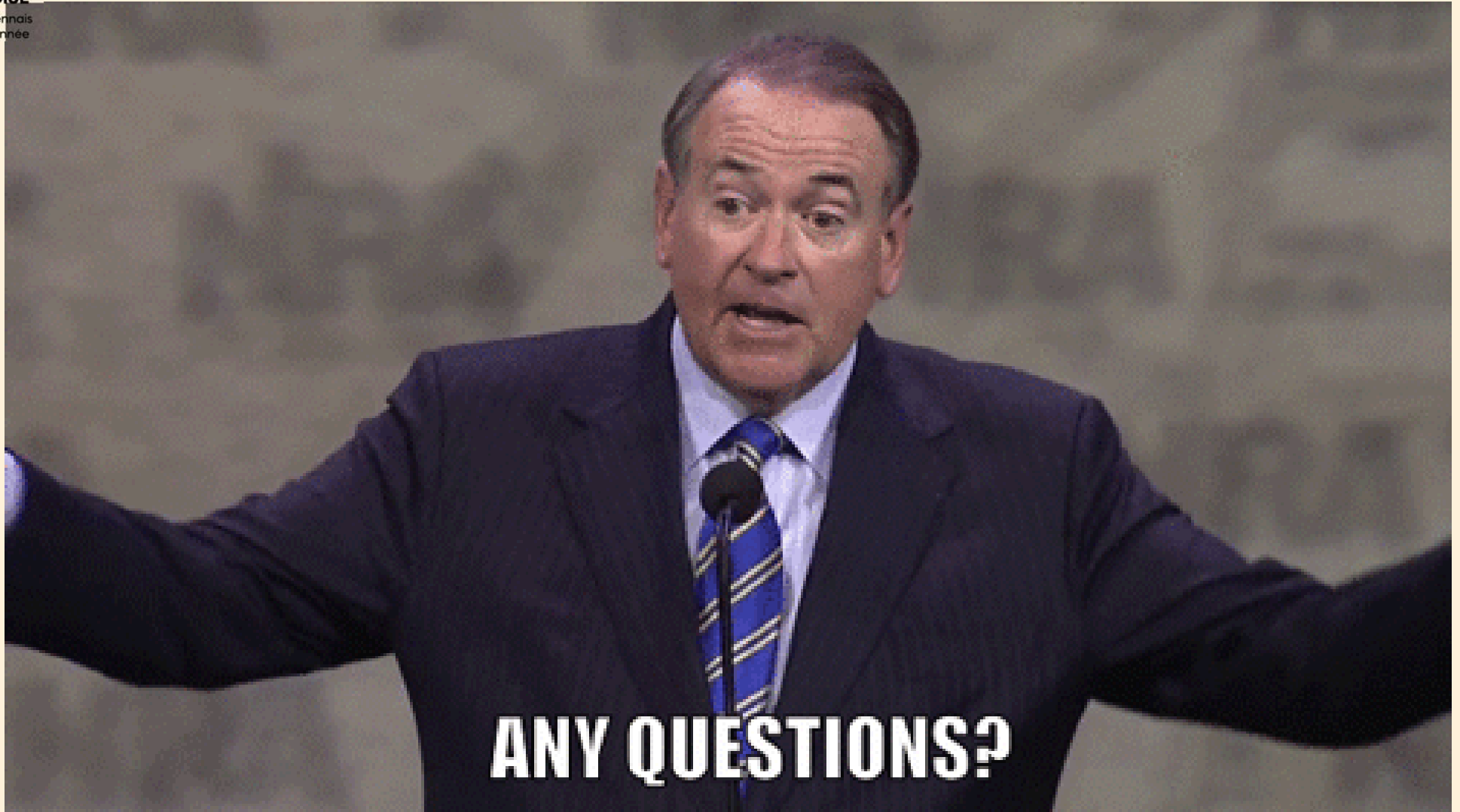
Use Guix as package manager to improve replicability  3

```
R version 4.3.0 (2023-04-21 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 11 x64 (build 22631)

Matrix products: default


locale:
[1] LC_COLLATE=French_France.utf8  LC_CTYPE=French_France.utf8
[3] LC_MONETARY=French_France.utf8 LC_NUMERIC=C
[5] LC_TIME=French_France.utf8


time zone: Europe/Paris
tzcode source: internal
```



illustrations

figure	source et crédits
--------	-------------------

Figure 1	Di Cosmo, R. et al. Curated Archiving of Research Software Artifacts: Lessons Learned from the French Open Archive (HAL). International Journal of Digital Curation 15, 16 (2020).
----------	--

Figure 2	Di Cosmo, R. et al. Curated Archiving of Research Software Artifacts: Lessons Learned from the French Open Archive (HAL). International Journal of Digital Curation 15, 16 (2020).
----------	--

Figure 3	Nathan, P. Overview of Data Governance. Derwen, Inc. https://derwen.ai/s/6fqf (2019)
----------	---



ARDoISE
Atelier rennais
de la donnée

References

1. Barker, M. *et al.* [Introducing the FAIR Principles for research software](#). *Sci Data* **9**, 622 (2022).
2. Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. & McGillivray, B. [The citation advantage of linking publications to research data](#). *PLOS ONE* **15**, e0230416 (2020).
3. Gorgolewski, K. & Poldrack, R. *A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research*. (2016). doi:[10.1101/039354](#).
4. Gabelica, M., Bojčić, R. & Puljak, L. [Many researchers were not compliant with their published data sharing statement: Mixed-methods study](#). *Journal of Clinical Epidemiology* **0**, (2022).
5. Shanahan, H. & Bezuidenhout, L. [Rethinking the A in FAIR Data: Issues of Data Access and Accessibility in Research](#). *Front. Res. Metr. Anal.* **7**, (2022).
6. Brown, N. J. L. [Fixing science means an end to gaming the system](#). *PLOS Biology* **22**, e3002816 (2024).
7. Jean-Quartier, C. *et al.* Sharing practices of software artefacts and source code for reproducible research. *Int J Data Sci Anal* (2024) doi:[10.1007/s41060-024-00017-7](#)