

How to manage, share and open research data, using the FAIR guiding principles? Feedbacks and perspectives

1

- FAIR guiding principles? FAIRness?

2

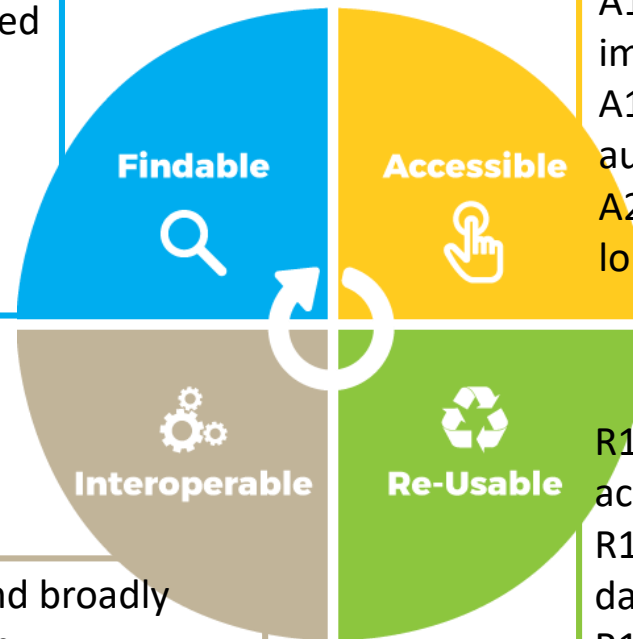
- Trying to apply the FAIR guiding principles to the LabEx DRIIHM research context: feedbacks and perspectives

3

- A concrete example of data sharing

What are the FAIR guiding principles?

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**

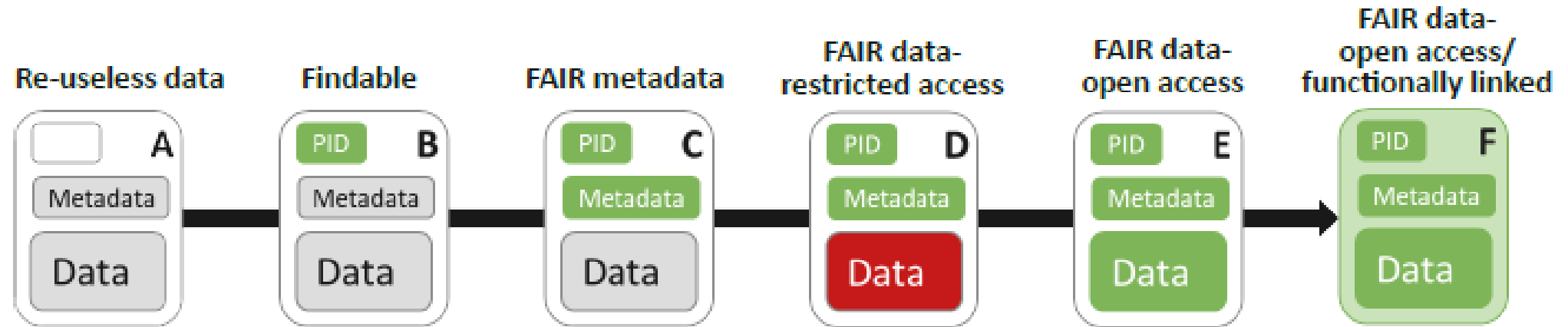


A1. (meta)data are retrievable by their identifier using a **standardized communications protocol**
 A1.1 the protocol is **open, free**, and universally implementable
 A1.2 the protocol allows for an **authentication** and authorization procedure, where necessary
 A2. **metadata are accessible**, even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

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

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**

The data "FAIRness"

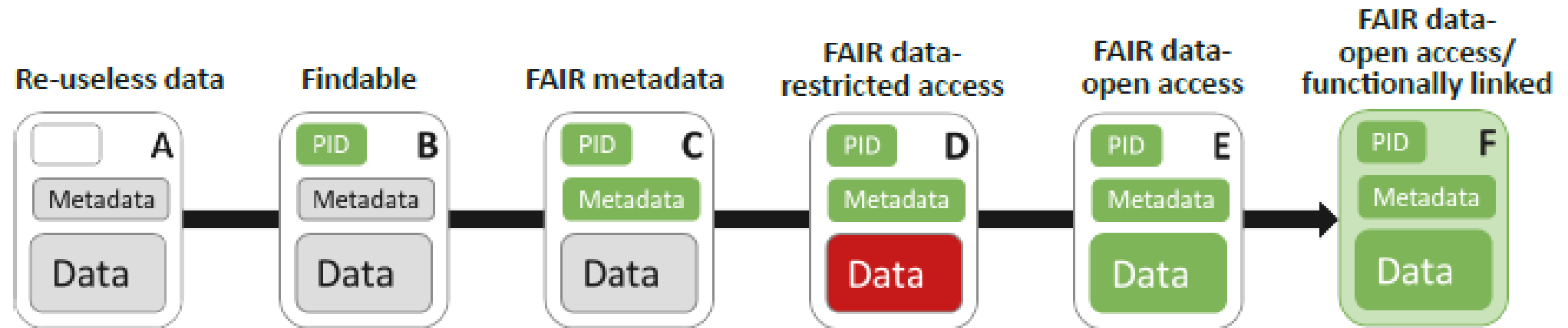


Indicators will be used to assess the level of FAIRness of the data

Green elements are FAIR and open / Red ones are FAIR and closed / PID = Persistent identifier

Adapted from Mons et al, 2017

The data "FAIRness" and the EU recommendations



Indicators will be used to assess the level of FAIRness of the data

"As open as possible..."



... as closed as necessary" (H2020 EU Commission)



Trying to apply the FAIR guiding principles to the LabEx DRIIHM research context

➤ What context? LabEx DRIIHM

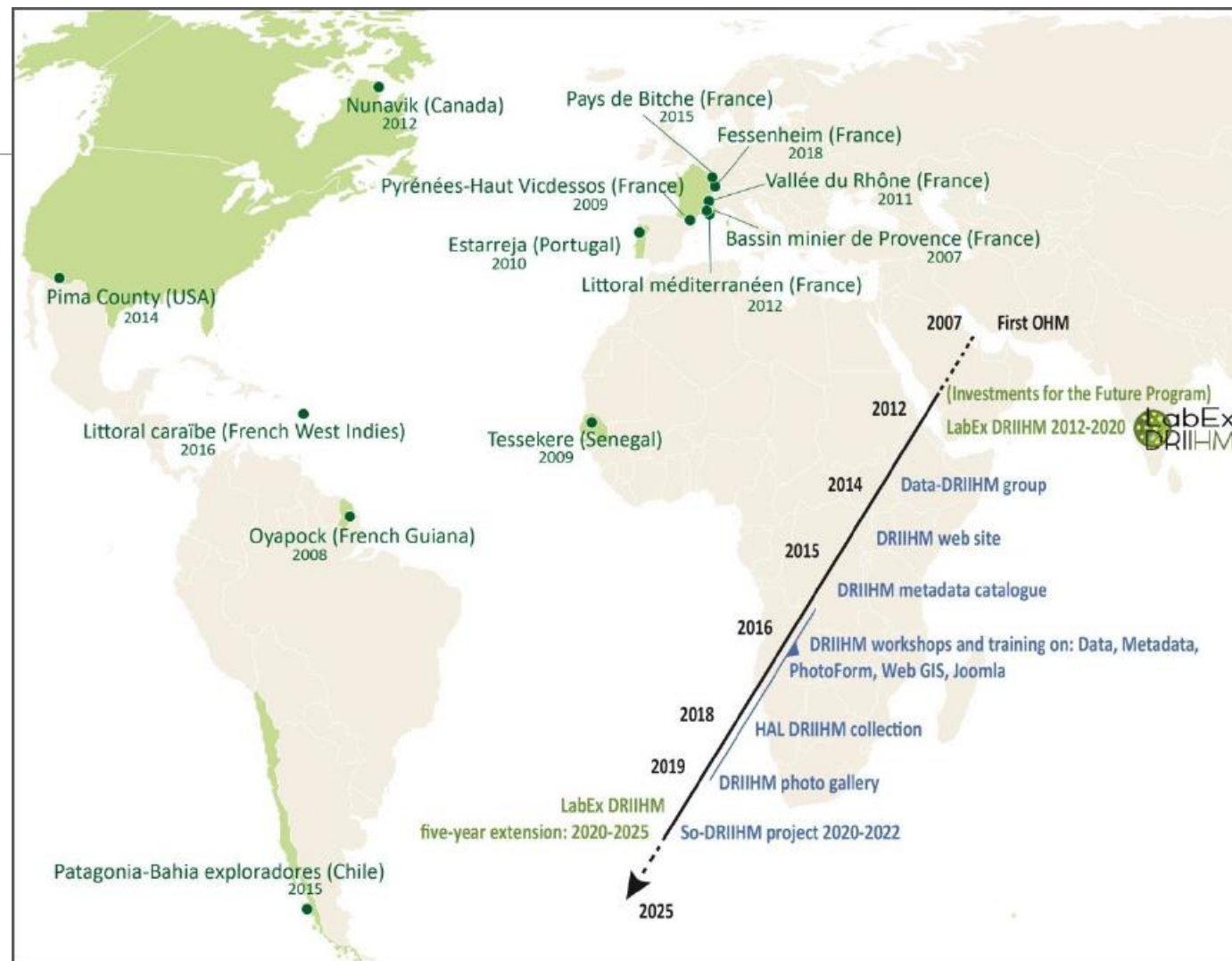
- Interdisciplinary research consortium focused on anthropogenically modified socio-ecosystems
- 13 Human-environment observatories
- ~ 1000 scientists from 65 different disciplinary fields

➤ Who?

- The Data-DRIIHM team composed by engineers and scientific coordinators, supported recently by ergonomic experts and web developers

➤ When?

- Started in 2012
- The SO-DRIIHM ANR project will accelerate the process (2020-2022)

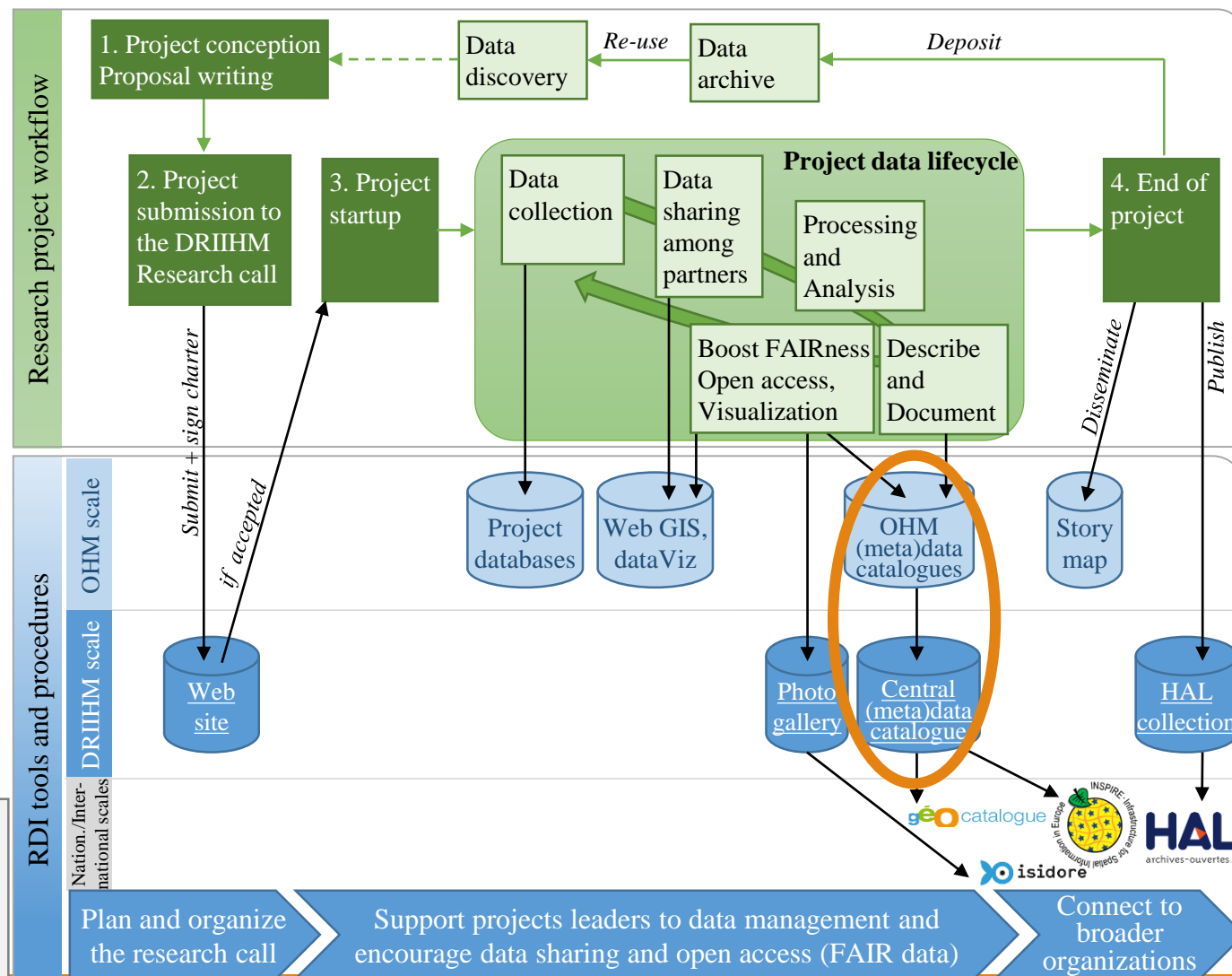


Arnaud, Lerigoleur, Jean-Charles et al., subm. in JIMIS

Trying to apply the FAIR guiding principles to the LabEx DRIIHM research context

➤ How?

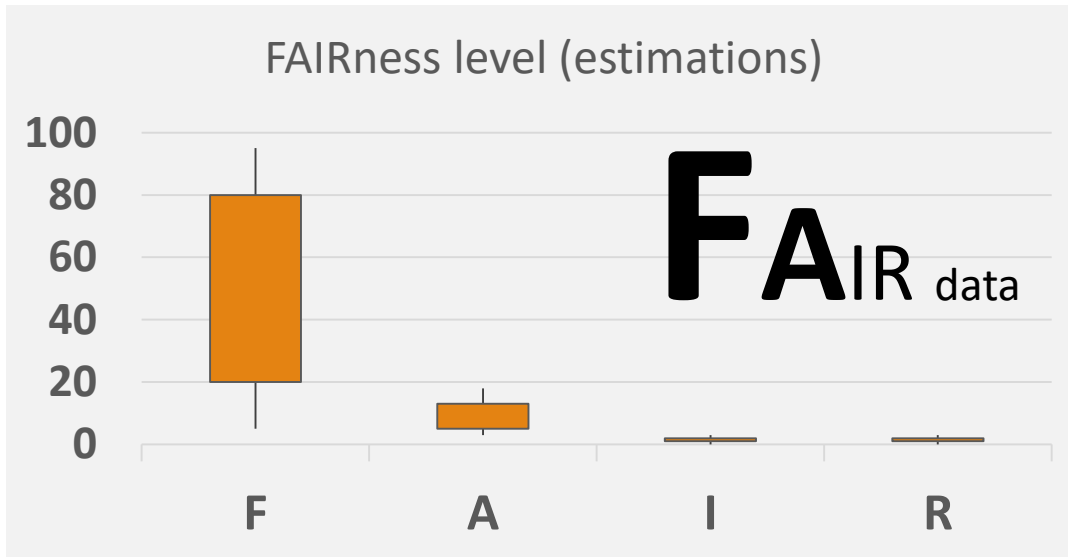
Implementation of a Research Data Infrastructure (RDI) during 2012-2019



Trying to apply the FAIR guiding principles to the LabEx DRIIHM research context

➤ How?

- A metadata catalog with some accessible data



- ➔ Geonetwork = non-ergonomic tool 😞
- ➔ Difficulties to estimate the FAIR levels

ohm vr

Search

Filter

- TYPE OF RESOURCES
 - Dataset (61)
 - Map (2)
 - Non geographic dataset (1)
 - Static map (4)
- AVAILABLE ACTIONS
 - Downloadable (3)
- TOPICS
 - Environment (2)
 - Inland waters (2)
- KEYWORDS
 - Inland waters (2)
 - Isotopes (2)
 - OHM VR (2)
 - Rhône (2)
 - Tree rings (2)
- CONTACT FOR THE RESOURCE
 - Cardiff University - Scho... (2)
 - UMR 5600 EVS - ENS d... (2)
 - University of St Andrews... (2)
- PROVIDED BY
 - 973851db-887a-42be... (66)
- YEARS
 - 2015 (6)
 - 2016 (8)
 - 2017 (6)
 - 2018 (4)
 - 2019 (2)
- FORMATS
 - CSV (2)
 - DOC (1)
 - JPEG (4)
 - JPEG, SHP, XLS (2)
 - JPG, TIF (2)

Sorted by relevancy

1 - 20 on 66

Source Water Isotope Identification Toolkit (SWIT) for detecting historical...

Isotopic characterisation of waters and tree ring cellulose along the Rhône...

Photothèque de l'OHM VR (2018)

Frise chrono-systémique de l'OHM VR (1750-2018)

Physico-chimie et écologie des casiers Girardon du Bas-Rhône (2012)

Corpus d'entretiens avec des acteurs du fleuve Rhône impliqués dans la crise...



Scientific and technological problems

Strengths

- Data-DRIIHM group
- Project submission platform
- Many tools for long-tail dataset management
- Metadata standards, interoperability
- HAL collection

Weaknesses

- Heterogeneous human resources
- Insufficient communication on sharing and opening data
- Very little open access data
- Lack of RDI evaluation indicators
- Insufficient interoperability

Opportunities

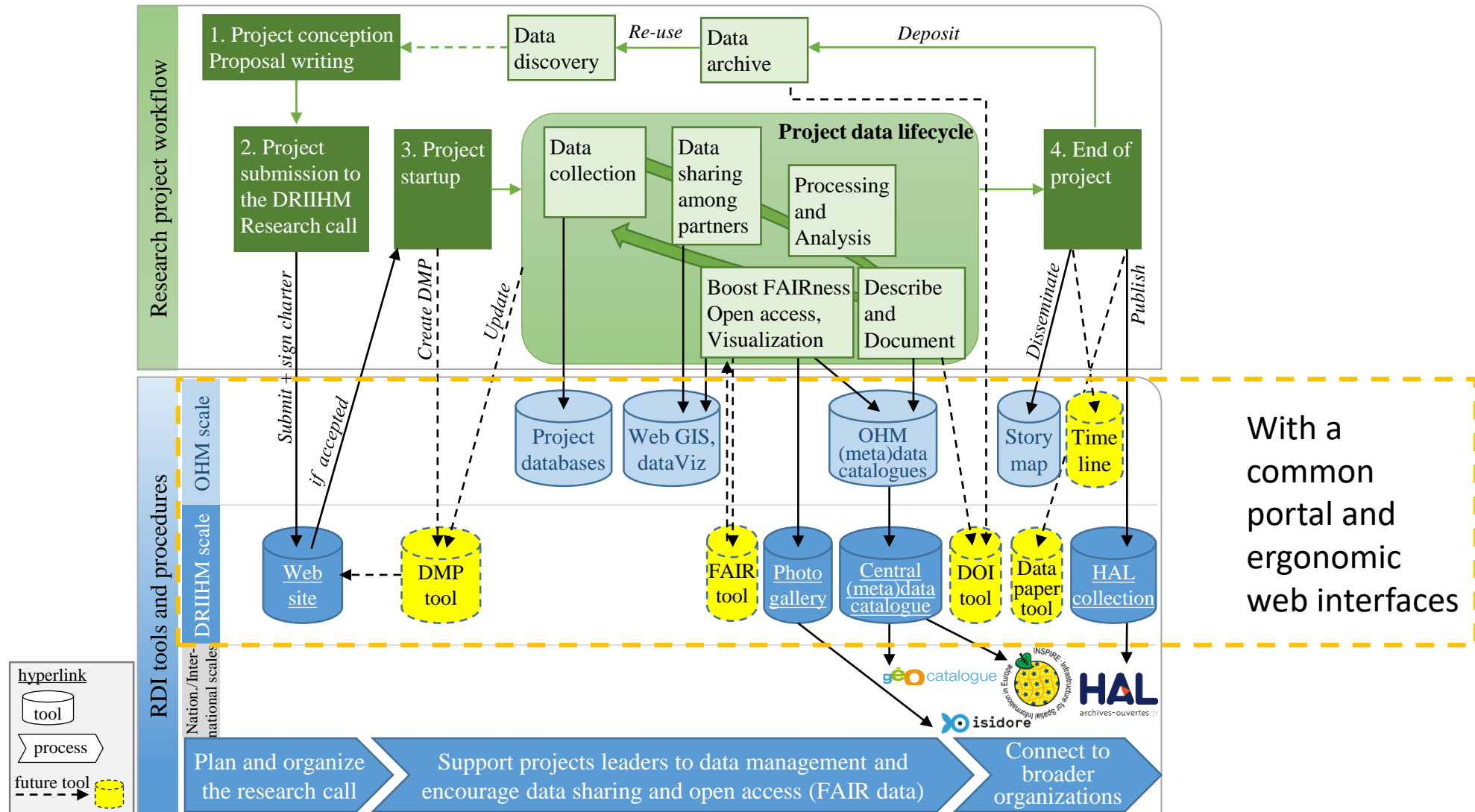
- Increasing open science awareness
- Promoting open-access data papers
- DOIs to datasets
- Story maps, Chrono-systemic timelines
- **User-friendly, interoperable e-infrastructure**

Threats

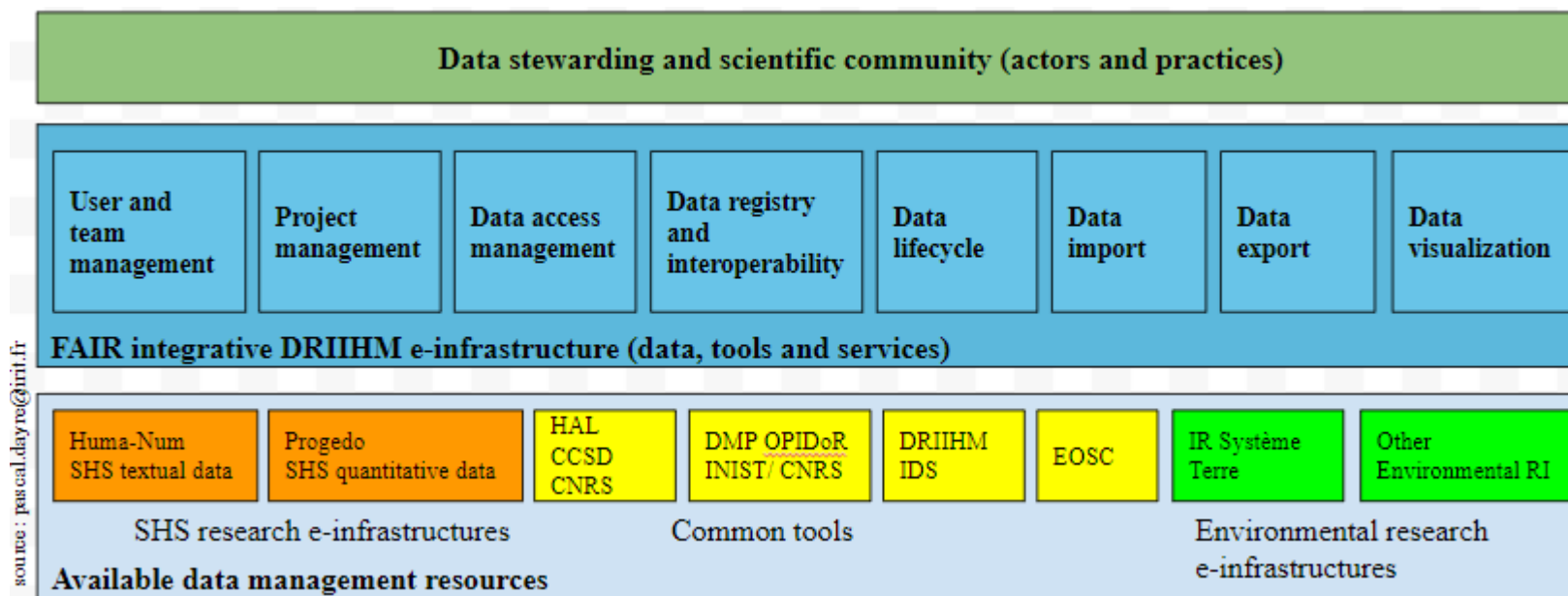
- Isolation of some OHMs
- Increasing rate of unshared data
- Lack of international visibility
- Misunderstanding researchers /data managers

SWOT matrix of the current DRIIHM data infrastructure

SO-DRIIHM project (2020-2022) objective: co-construction of an e-infrastructure



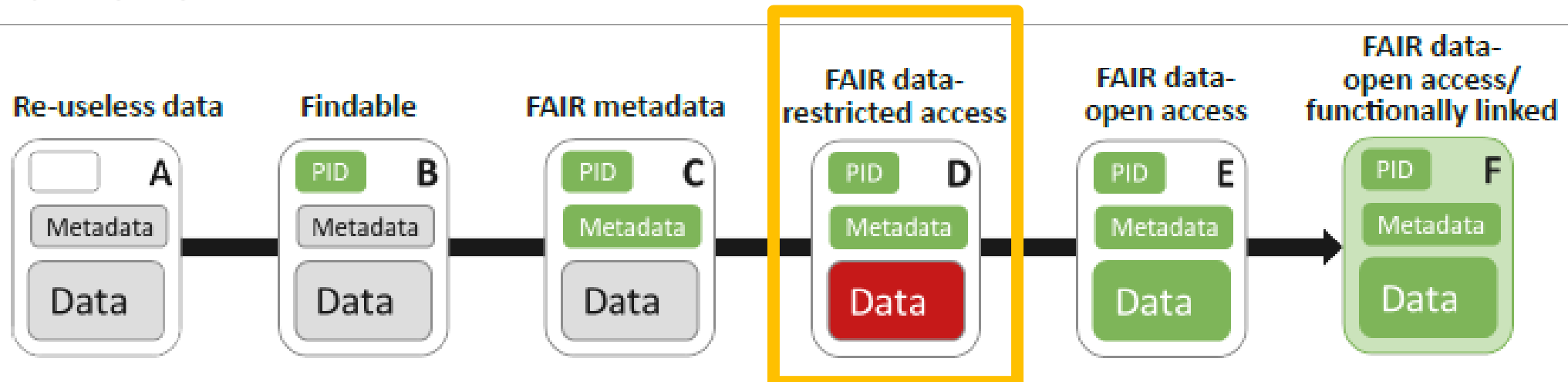
SO-DRIIHM project (2020-2022) objective: co-construction of an e-infrastructure



A virtual research environment (VRE) or virtual laboratory is an online system helping researchers collaborate

➔ Is a VRE (one of) the technical solution(s) for our DRIIHM RDI?

What objective, acceptable for all, and the means to achieve it?



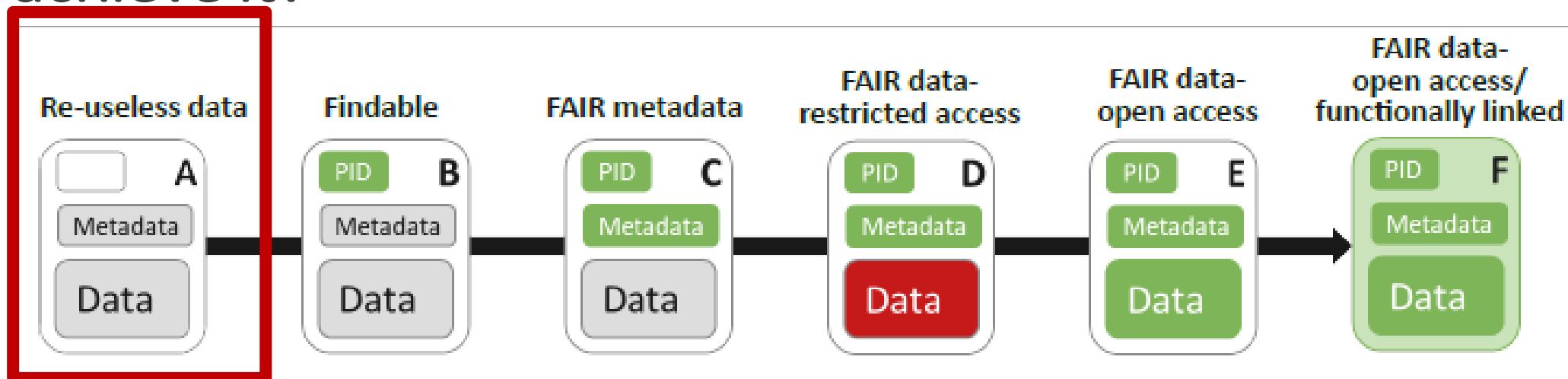
Indicators will be used to assess the level of FAIRness of the data

Green elements are FAIR and open / Red ones are FAIR and closed / PID = Persistent identifier

Adapted from Mons et al, 2017

Is this level acceptable?
Is it the minimum requirement?

What objective, acceptable for all, and the means to achieve it?



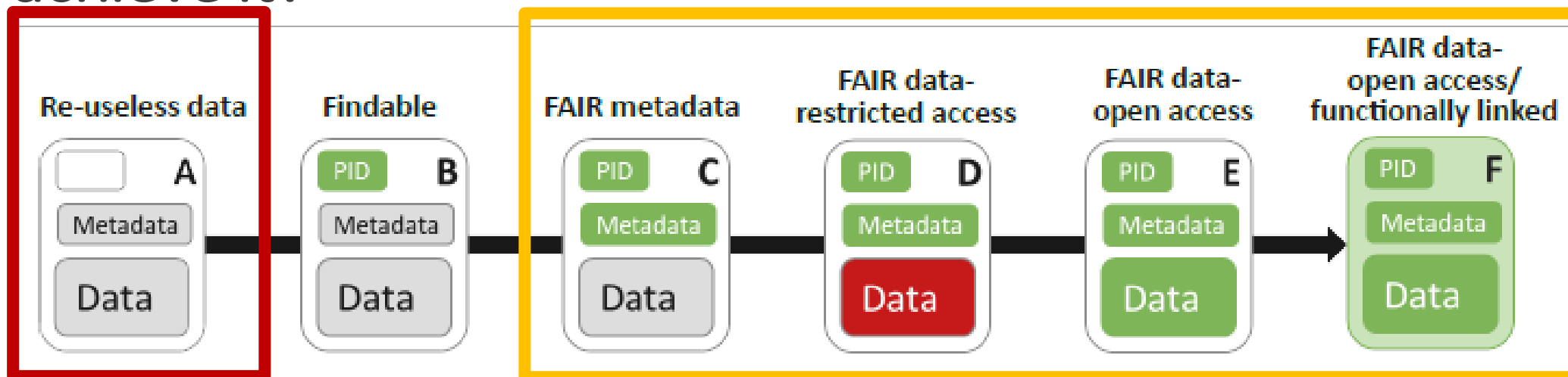
Indicators will be used to assess the level of FAIRness of the data

Green elements are FAIR and open / Red ones are FAIR and closed / PID = Persistent identifier

Adapted from Mans et al, 2017

The difficulty will be to quantify the Re-useless data (those not included in the e-infrastructure)

What objective, acceptable for all, and the means to achieve it?



Indicators will be used to assess the level of FAIRness of the data

Green elements are FAIR and open / Red ones are FAIR and closed / PID = Persistent identifier

Adapted from Mans et al, 2017

The difficulty will be to quantify the Re-useless data (those not included in the e-infrastructure)

We have to quantify the number of datasets per level of FAIRness

Evaluation of the use of the future DRIIHM e-infrastructure: Key Performance Indicators (KPIs)

- For data producers

- the number of metadata records (total and per project)
- the proportion of researchers who referenced / indexed their datasets

- For data users

- the number of visits (distinguishing new visitors from those who visit it again)
- the connexion duration
- the bounce rate (ratio between the total number of visits and the number of visits to a page)
- the number of consulted metadata
- the number of requests access to the data not the metadata (download or visualisation)

+ evaluation questionnaires filled after each training/seminar event and analysed

A concrete example of data sharing

A PhD student asked me: « The Journal of Geophysical Research-Biogeosciences constrained me to give access to my datasets (*ie.* database query results). What should I do? »

➔ First, I suggest him to write metadata and to include them into our Geo-catalogue

BUT it was not validated by the Journal

<http://ohmpyr-cat.univ-tlse2.fr/geonetwork/srv/fre/catalog.search#/metadata/d460c4be-0c0d-4629-96ce-125d6f179944>



The screenshot shows a web interface for a Geo-catalogue. The main title is "Hydrological and biogeochemical database of two Pyrenean peatlands." Below the title, there is a text box with the following content: "The database has been developed during Thomas Rosset's Ph.D at Ecolab, CNRS, University of Toulouse (LabEx DRIIHM). Useful to assess stream organic carbon fluxes, data collected come from a monitoring network which consists in several remote sensors and manual sampling operations. Data are collected at different frequency (30min to monthly). To organize reports and analysis a web app has been developed to import, visualize and export datasets." To the right of the text box, there is a small table with columns for "Nom de la donnée", "Type de données", "Métadonnées", "Date de mise à jour", and "Statut". Below the text box, there is a section titled "Téléchargements et liens" with a button "Accéder à l'interface cartographie en ligne (accès restreint)" and a button "Ouvrir le lien". Below that, there is a section titled "Ressources associées" with a button "Fiche parent". On the right side of the page, there is a section titled "Aperçu" with a line graph showing data over time. Below the graph, there is a section titled "Contact" with the text "UMR 5602 - Géographie de l'Environnement (GEODE)". At the bottom right, there is a section titled "Étendue spatiale" with a map showing the location of the peatlands near Toulouse.

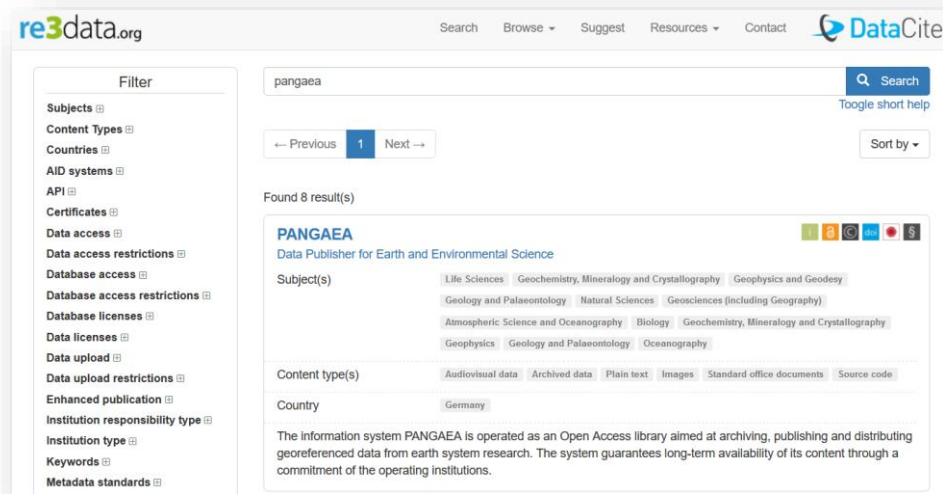
A concrete example of data sharing

→ Then 2 solutions were undertaken:

① **attribute a GEODE DOI** with a dedicated landing page and an internal data repository (Univ. Toulouse 2)

② **deposit the dataset in an external repository** such as PANGAEA (geoscience specific data repository)

	
Local storage Speed of process	Sustainability?
Sustainability Re3data repository	1-3 weeks delay



The screenshot shows the re3data.org search interface. The search term 'pangaea' is entered in the search bar. The results show 8 results, with the first result being 'PANGAEA Data Publisher for Earth and Environmental Science'. The subject(s) listed are Life Sciences, Geochemistry, Mineralogy and Crystallography, Geophysics and Geodesy, and Geology and Palaeontology. The content type(s) listed are Audiovisual data, Archived data, Plain text, Images, Standard office documents, and Source code. The country listed is Germany.

→ Solution ② adopted!

→ Process:

- 1) submit a request to PANGAEA
- 2) PANGAEA validates (or not)
- 3) add metadata and deposit the dataset

A concrete example of data sharing

➔ This PANGAEA landing page conforms to best practices for landing pages (according to DataCite)

The screenshot shows the PANGAEA landing page for a dataset. The page includes the PANGAEA logo, navigation links (SEARCH, SUBMIT, ABOUT, CONTACT), and a full bibliographic citation. A red box highlights the citation text: "Hüning, Anne K; Melzner, Frank; Thomsen, Jörn; Gutowska, Magdalena A; Krämer, Lars; Kohno, Mika; Rosenstiel, Philip; Pörtner, Hans-Otto; Philipp, Eva E R; Lucassen, Magnus (2013): Seawater carbonate chemistry and mantle gene expression patterns of the Baltic Sea blue mussel in a laboratory experiment. PANGAEA, doi: https://doi.org/10.1594/PANGAEA.833670". A red box also highlights the DOI URL in the citation. Another red box points to the DOI URL in the page meta data (DC.identifier) shown in the browser's developer console. A third red box highlights the "Download Data" section, which includes options to download the dataset as tab-delimited text or view it as HTML. A vertical URL on the right side of the page is also visible: https://support.datacite.org/docs/landing-pages.

Full bibliographic citation

DOI displayed as URL

Access to the item

DOI tagged in page meta (DC.identifier in this case)

https://support.datacite.org/docs/landing-pages

A concrete example of data sharing

→ Result:

[SEARCH](#) [SUBMIT](#) [ABOUT](#) [CONTACT](#)



PANGAEA.
 Data Publisher for Earth & Environmental Science

Citation:

Rosset, Thomas; Gandois, Laure; Le Roux, Gaël; Teisserenc, Roman; Durantez Jimenez, Pilar; Camboulive, Thierry (2019): Stream organic matter exports from a montane peatland (2015-01-01 to 2018-01-01). *PANGAEA*, <https://doi.org/10.1594/PANGAEA.899200>,

Supplement to: Rosset, Thomas; Gandois, Laure; Le Roux, Gaël; Teisserenc, Roman; Durantez Jimenez, Pilar; Camboulive, Thierry; Binet, S (accepted): Peatland contribution to stream organic carbon exports from a montane watershed. *Journal of Geophysical Research-Biogeosciences*

Always quote above citation when using data! You can download the citation in several formats below.

[RIS Citation](#)
[Bibtex Citation](#)
[Copy Citation](#)
[Facebook](#)
[Twitter](#)
[Show Map](#)
[Google Earth](#)

Abstract:

High frequency (30 min) data of discharge, fluorescence of the organic matter and turbidity used to calculate respectively the exports of dissolved and particulate organic carbon from a mountainous watershed containing a peatland. Measurements were done both at the inlet and at the outlet of the peatland for three years from 2015, 1st January to 2018, 1st of January. This depository integrates low frequency organic carbon concentration data which were used for optical sensors calibrations and tracking the origin of the organic carbon in the watershed.

Keyword(s):

Dissolved organic carbon 🔍; Particulate organic carbon 🔍; Peatland 🔍; Pyrenees 🔍; Stream organic carbon 🔍

Coverage:

Median Latitude: 42.802275 * Median Longitude: 1.425238 * South-bound Latitude: 42.801427 * West-bound Longitude: 1.422660 * North-bound Latitude: 42.803240 * East-bound Longitude: 1.426735
 Date/Time Start: 2013-10-03T10:00:00 * Date/Time End: 2018-01-01T00:00:00

License:

 Creative Commons Attribution 4.0 International

Size:

4 datasets



Download Data

Download **ZIP** file containing all datasets as tab-delimited text (use the following character encoding:)

<https://doi.pangaea.de/10.1594/PANGAEA.899200>

A concrete example of data sharing

→ Result:

We have now to harvest PANGAEA to expand our DRIIHM RDI



PANGAEA.
Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

Citation: **Rosset, Thomas; Gandois, Laure; Le Roux, Gaël; Teisserenc, Roman; Durantez Jimenez, Pilar; Camboulive, Thierry (2019):** Stream organic matter exports from a montane peatland (2015-01-01 to 2018-01-01). *PANGAEA*, <https://doi.org/10.1594/PANGAEA.899200>,
Supplement to: Rosset, Thomas; Gandois, Laure; Le Roux, Gaël; Teisserenc, Roman; Durantez Jimenez, Pilar; Camboulive, Thierry; Binet, S (accepted): Peatland contribution to stream organic carbon exports from a montane watershed. *Journal of Geophysical Research-Biogeosciences*

PANGAEA Web Services Server

The following services are available:

- [PANGAEA Search Web Services](#) (proprietary)
- [PANGAEA Background Services](#) (proprietary)
- [PANGAEA Elasticsearch instance](#)
 - [PANGAEA main index](#) (*panmd metadata scheme*)
 - [Data Portals index](#) (*pansimple.xsd metadata scheme*)
- PANGAEA value-added services: Data Download
 - [filter by dataset DOI and parameter value](#)
 - [filter by geocodes and parameters](#)
- [OAI-PMH](#) (Open Archives Initiative Protocol for Metadata Harvesting 2.0)
- [XML Schema Repository](#)

<https://ws.pangaea.de/>

Download Data
Download ZIP file containing all datasets as tab-delimited text (use the following character encoding: UTF-8: Unicode (PANGAEA default))

Conclusions

□ Trying to apply the FAIR guiding principles to the LabEx DRIIHM research context

- Not so easy to implement them to an Interdisciplinary research consortium focused on anthropogenically modified socio-ecosystems
- "As open as possible, as closed as necessary" EU H2020 recommandation
- Is the "FAIR data-restricted-access" an acceptable objective to achieve as a minimum requirement?
- How to quantify the Re-useless data? How to quantify each F, A, I and R levels?
- Which kind of e-infrastructure? Is a VRE (one of) the technical solution(s) for our DRIIHM RDI?

□ Data sharing

- Persistent identifier = DOI? Other?
- Landing page best practice
- External or internal data repository?