



HAL
open science

Success stories of FAIR data publication on behalf of the ANR SO-DRIIHM Open Science project

Emilie Lerigoleur, Laurent Bouveret, Ablaye Diop, Jean-Luc Jung, Iwan Le Berre, Mathieu Massaviol, Florence Mazier, Papa Ibnou Ndiaye, Anna Niang, Corinne Pardo, et al.

► To cite this version:

Emilie Lerigoleur, Laurent Bouveret, Ablaye Diop, Jean-Luc Jung, Iwan Le Berre, et al.. Success stories of FAIR data publication on behalf of the ANR SO-DRIIHM Open Science project. International Symposium of LabEx DRIIHM 2023, Jun 2023, Strasbourg, France. , 2023, 10.34972/driihm-a7445b . hal-04128197

HAL Id: hal-04128197

<https://hal.science/hal-04128197>

Submitted on 14 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Success stories of FAIR data publication on behalf of the ANR SO-DRIIHM Open Science project

E. Lerigoleur^{1*}, L. Bouveret², A. Diop³, J.-L. Jung⁴, I. Le Berre⁵, M. Massaviol⁶, F. Mazier¹, P. I. Ndiaye³, A. Niang³, C. Pardo⁶, M.-L. Trémélo⁷

¹ CNRS - GEODE (UMR5602), Toulouse; ² OMMAG, Port Louis; ³ Univ. Cheikh Anta Diop - ESS (IRL3189), Dakar; ⁴ MNHN - ISYEB (UMR7205), Dinard; ⁵ Univ. Bretagne Occidentale - LETG (UMR6554), Brest; ⁶ CNRS - ECCOREV (FR3098), Aix-en-Provence; ⁷ CNRS - ESPACE (UMR7300), Aix-en-Provence

SO-DRIIHM project in brief

The ANR SO-DRIIHM project encourages the sharing and opening of data within, between and beyond OHMs.

- Promote Open Science** and enhance Open Science practices, showing the challenges and benefits of structuring, accessing, sharing and opening research data
- Enhance the DRIIHM existing infrastructure** to a more user-friendly and interoperable e-infrastructure with new functionalities that will be prioritized according to users' expectations and motivations (co-design process)



What is FAIR Data?

Findable Accessible Interoperable Reusable

- F Findable:** Metadata and data should be findable for both humans and computers
- A Accessible:** Once found, users need to know how the data can be accessed
- I Interoperable:** Data needs to work with applications or workflows for analysis, storage and processing
- R Reusable:** The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

3 success stories in making data as FAIR as possible

TRAFIC project core team

- J.-L. Jung: responsible
- L. Coché: trainee in ecology
- I. Le Berre & E. Lerigoleur: data stewards
- R. David & B. Madon: data experts
- OMMAG & other NPOs and whale-watchers



Wild animals project core team

- P.-I. Ndiaye: responsible
- A. Diop & A. Niang: PhD students
- P. Duboz: OHMi coordinator
- E. Lerigoleur: data steward



PASTSERV project core team

- F. Mazier: responsible
- S. Haunold, C. Plancher, C. Bonnet, H. Belly, T. Houet: data producers and data analyst
- E. Lerigoleur: data steward



1. Identify the datasets useful for dissemination

The Kakila database (cetacean biodiversity in Guadeloupe) = **6 linked tables** (xlsx format)



3 tables (xlsx format):

- Large mammals biodiversity in Ferlo (North of Senegal)
 - occurrence data
 - trapping data of large mammals in Ferlo
- Bird biodiversity in Ferlo (occurrence data)



1 table containing absolute cover percentage of plant species from **vegetation survey** in the Vicdessos mountainous area (Pyrenees, France) (xlsx format)



2. Identify the most appropriate data repository // Check conditions



The **PNDP** portal (<https://data.pndb.fr/data>) offers:

- ✓ A national data portal in the field of biodiversity **F A**
- ✓ A permanent identifier (DOI) **F**
- ✓ Open licenses (eg. CC BY) **R**
- ✓ A detailed description of the data using the EML (Ecological Metadata Language) metadata standard and the possibility to specify provenance and the data dictionary (variables) **F**
- ✓ Interoperability with other thematic portals **I**
- ✓ Free of charge, many formats accepted, ... **I R**



The **GBIF** portal (<https://www.gbif.org/>) offers:

- ✓ An international biodiversity data portal **F A**
- ✓ A DOI **F**
- ✓ Open licenses (eg. CC BY) **R**
- ✓ A detailed description of the data using EML standard metadata **F**
- ✓ Use of the Darwin Core Standard (DwC) to structure and interconnect text files with a common terminology → imposed data format, but this standardization simplifies the publication process, and it also makes it easy for users to discover, search, evaluate and compare datasets **I R**
- ✓ Interoperability with other thematic portals **I**
- ✓ Free of charge, open formats ... **I R**



GBIF has been chosen to publish the dataset in the standard Darwin Core format.

The **Data.InDoRES** portal (<https://data.indores.fr/>) has been chosen to publish the source file with a structure commonly used by the research community. It offers:

- ✓ An institutional data repository (CNRS-INEE/MNHN) for long tail environmental data **F A**
- ✓ A DOI **F**
- ✓ Open licenses (eg. CC BY) **R**
- ✓ Standardized core metadata (Dublin core, DDI) **F I**
- ✓ Free of charge, many formats accepted, ... **I R**

3. Deposit & dissemination // Documenting and cataloguing data thanks to metadata

Deposit tool: MetaShARK (*Metadata Shiny Application for Resources and Knowledge*) hosted by PNDP <https://www.pndb.fr/en/tools/partager>

Time spent: **data curation + + + + + / data deposit + +**

DOI of the dataset: <https://doi.org/10.48502/8bb5-pk85>
DOI of the data paper: <https://doi.org/10.3897/BDJ.9.e69022>

Deposit tool: IPT Integrated Publishing Toolkit hosted by the GBIF France <https://ipt.gbif.fr/>

Time spent: **data curation + / data deposit + +**

DOI of the 3 datasets: <https://doi.org/10.15468/k6vut5>
<https://doi.org/10.15468/tr6zfp>
<https://doi.org/10.15468/u484g6>

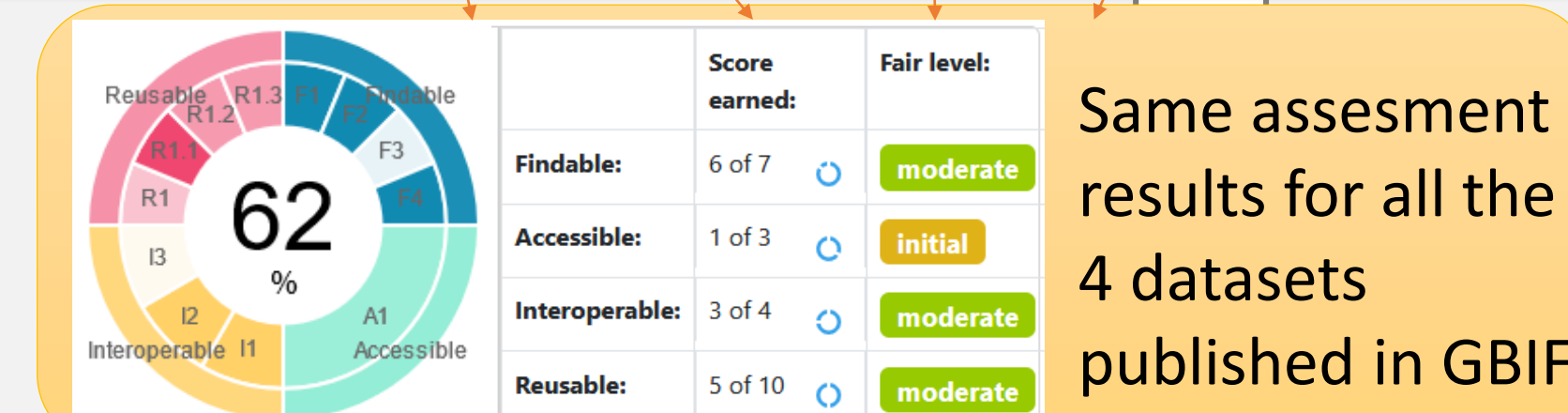
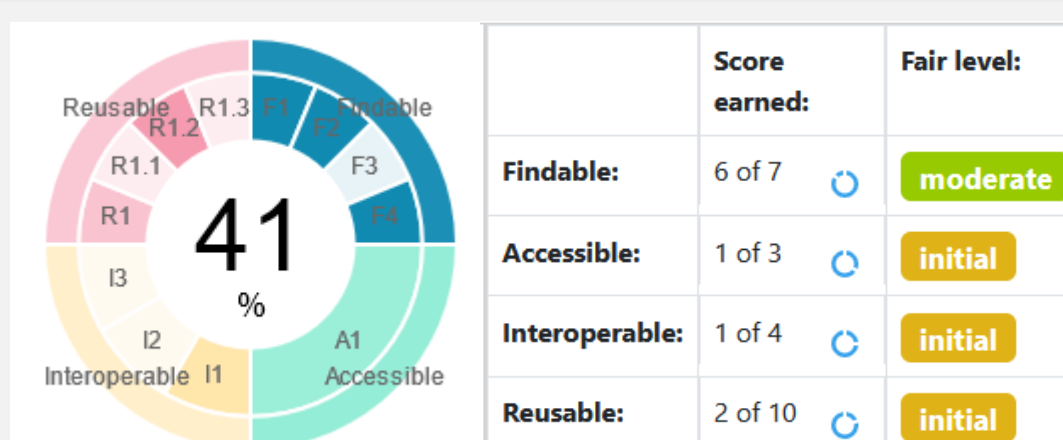
Deposit tools: IPT for GBIF and the dataverse web interface for Data.InDoRES

Time spent: **data curation + + + / data deposit + + +**

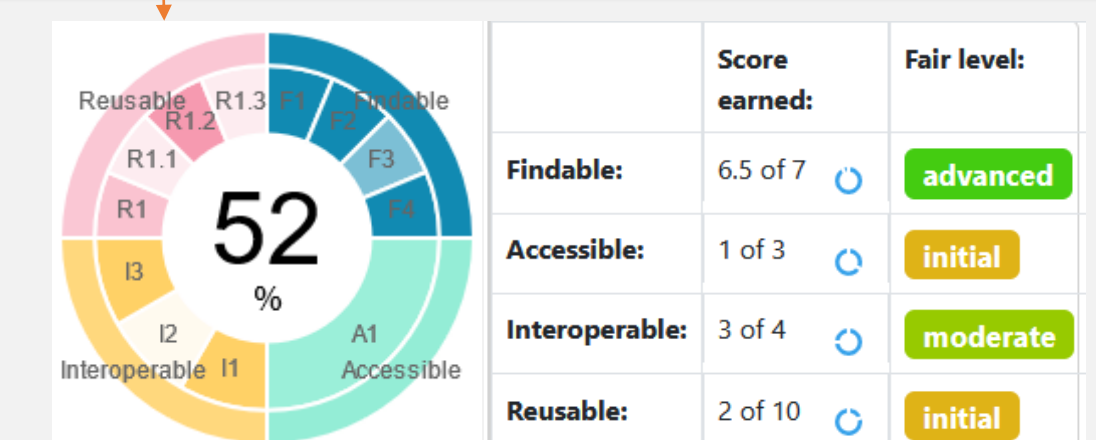
DOI of the datasets:
- GBIF Darwin Core file: <https://doi.org/10.15468/4q47d3>
- Data.InDoRES source file: <https://doi.org/10.48579/PRO/9S11PN>

4. Check the FAIRness level of the data using F-UJI tool

Note: F-UJI is a web service to programmatically assess FAIRness of research data objects. It is not perfect, it is being improved! <https://f.uji.net>



Same assesment results for all the 4 datasets published in GBIF



- +** Not so hard to achieve!
A DOI for each dataset
Guaranteed medium-term preservation
Fairly good level of data FAIRness

! Pay particular attention to the choice of data repository, which will affect the completion of the FAIR criteria

- A significant amount of dedicated learning time
Recurring need for data curation

Interested in data sharing? Interested in making your data FAIR?
Would you like to contribute to the SO-DRIIHM project as a beta tester for our future open science support platform? → **Contact us** @ datadriihm-groupe@services.cnrs.fr or emilie.lerigoleur@univ-tlse2.fr

Publish an improved version of the data and metadata