Science ouverte
Université Grenoble Alpes

CELLULE DATA
GRENOBLE ALPES

Making the most of scientific data and codes: publication and dissemination/sharing

Masterclass on Open Sciences and Publishing – Grenoble
03rd-04th of June, 2024

Plan

Definition
Why?
How to disseminate FAIR?
Which part of your research disseminate?
Where to disseminate your research
- *Publications*
- *Data*
- *Code*
- *Dissertations*
- *Linking publications, data, code and software*

How to disseminate your research ?
- *Focus on legal issues*

Help and resources

Definition

Research Data

Research data is any information that have been collected, observed, generated or created to do research.
Research data may be arranged or formatted to be made suitable for communication, interpretation and processing.

Examples :

- Observation data
- Experimental data
- Simulation data
- Compiled data
- Survey data
- Recorded data

Definition

Codes

- Set of instructions making up a computer program in a programming language.
- Set of text files readable by a user and executable by a machine.
- Representation of a software program so that a user can make modifications.

Why?

Integrity and reusability

- Ethics and scientific integrity
- Reproducibility of results
- Re-usability of data

Increased visibility

- Research Data accessible and citable independently from the article
- Linking data to publications

Legal aspect

- Mandatory for french public research

Why?

Valuable or unique data

- High production costs
- High processing costs
- Unique capture
  - e.g. video-based research…
- Difficulties in accessing the terrain

« As Open as Possible, as Closed as Necessary »

Why?

Publishers' requirements :

Examples

Plos One
Elsevier
Springer / Nature

Requirements of funders

- **Horizon Europe 2021-2027** : « Open science is a policy priority for the European Commission and the standard method of working under its research and innovation funding programmes as it improves the quality, efficiency and responsiveness of research. »
- ANR

Requirements of governments and institutions

Plan National Science Ouverte 2021-2024
CNRS
CEA
UGA

How to FAIR?

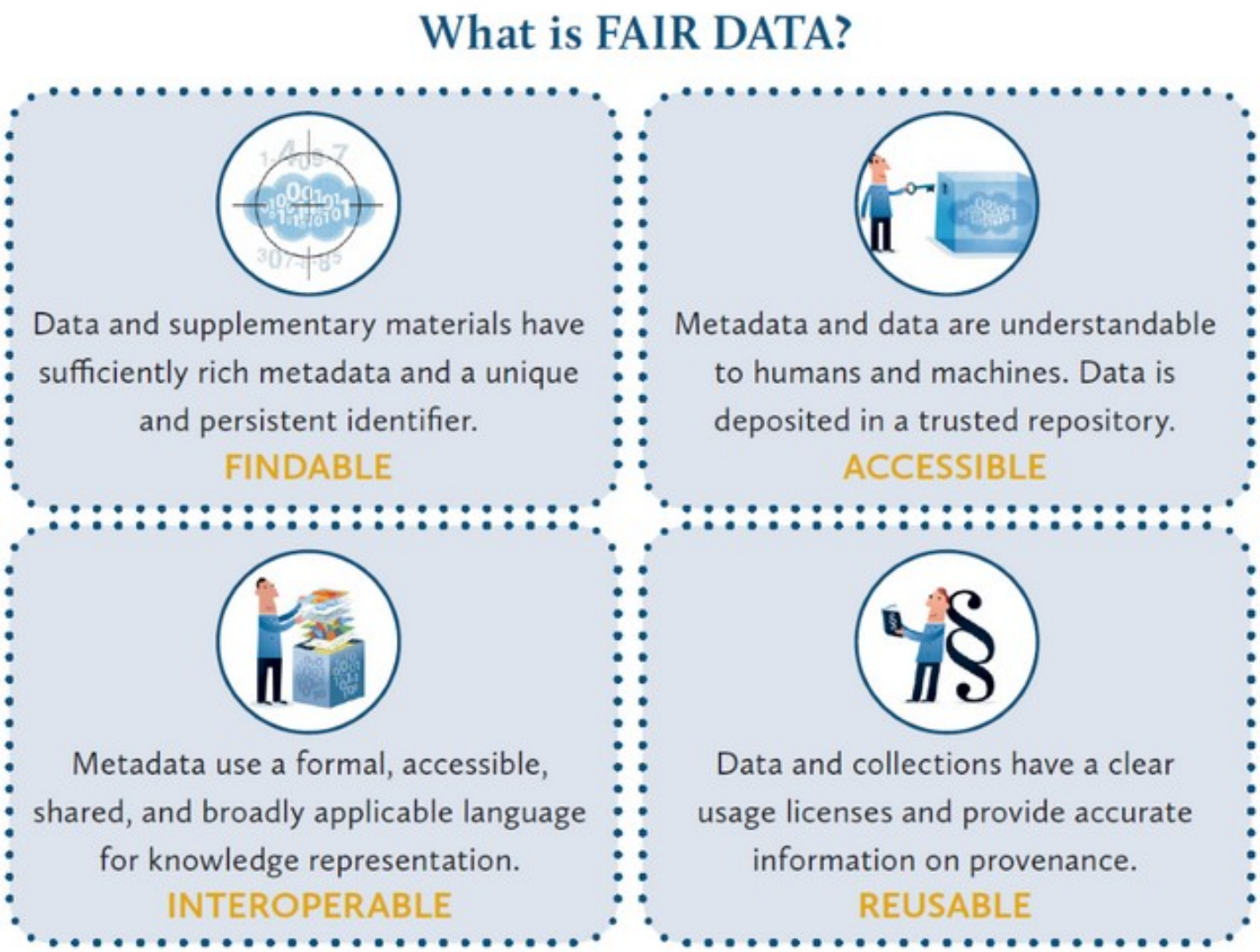


Illustration from: Implementing FAIR Data Principles: The Role of Libraries. Liber Images in the FAIR Data graphic are from Digitalbevaring.dk / Jørgen Stamp (CC BY 2.5 Denmark license)

What?

- **How to select research data**
  - Which data can be reproduced or re-used?
  - Is the data unique?
    - If so, make them available
  - Was data processing costly/difficult (time or resources)?
  - Was data processing innovative?
  - Are tools/codes/software required to re-use the data?
    - If possible, disseminate related codes/software
    - Or chose, if possible, an open software to process the data *(e.g. txt vs docx)*

What?

- Disseminating « negative » or « non conclusive » data?
- « Messy » data *(e.g. bad audio files)*
- Environmental impact: limiting data volume
- RGPD
- Ethics

Where?

Different ways

- Research data repositories
- Article including research data
- Supplementary materials
- Data paper

Where? Data

Research Data repositories :

Data repositories are tools for sharing and preserving research data

- National or Institutional Repositories
- Domain-specific or Disciplinary Repositories
- General-purpose or Open Repositories

Where? Data

General

Zenodo (CERN)

National

Recherche Data Gouv (France)

Data Archiving and Networked Services (DANS)

Institutional

Data INRAE

Data.sciencespo

Where? Data

How to select a repository ?

- Choosing a domain-specific repository
- Recommendations from funder, publisher
- Recommendations from research project, partners
- Recommendations from institution or organization
- Legal aspects

Where? Data

UGA Recommandations

- Choosing a domain-specific repository

  If not possible :

  - **Data Repository Grenoble Alpes** (Recherche Data Gouv)
    - Identifier Doi
    - All data types
    - Multidisciplinary
    - Restricted access possible
    - Link to publications
    - Rich metadata
    - Online help

Where? Data

How to select a repository

Status and policy

- Public / private ?
- Server location?
- Certification? Recognized?
- Moderation? *(e.g. Nakala « tests »)*
- Business model (cost per deposit?)
- Duration of data preservation ?

Where? Data

How to select a repository

- Data types
- Format types
- DOI **attribution**
- Metadata quality
- Version management
- Link with publication and codes
- Volume
- Types of access (embargo, access on request, etc.)
- Licensing

- Simplicity
- Support address

Where? Data

Domain-specific

Environmental sciences

- Global Biodiversity Information Facility(GBIF)
- Centre de Données astronomiques de Strasbourg (CDS)
- DataONE: Data Observation Network for Earth

Materials Science

- Materials Cloud Archive
- Crystallography Open Database (COD)

Chemistry

- PubChem
- ChEMBL

Medical sciences, biology

- wwPDB (Protein Data Bank)
- GenBank
- Institut Français de Bioinformatique (IFB)

Where? Data

Domain-specific

Social sciences

Progedo >>> **Plateforme universitaire des données UGA**

Inter-university Consortium for Political and Social Research (ICPSR)

Qualitative Data Repository (QDR)

Linguistic Data Consortium

Ortolang

Open science framework (OSF)

Humanities

- Nakala (Human Sciences, e.g. Margot Lherbet Repo)
- Archaeology Data Service

Where? Data

How to find a repository?

Directory of repositories

Re3data (Registry of Research Data Repositories)

Repository Finder (DataCite)

CoreTrustSeal (list of certified repositories)

Tools

Find a Repository (French) - Université de Bordeaux – global

« Selecting a thematic repository for scientific data: methods and analysis of existing resources », COmité pour la Science Ouverte, 2024

CAT OPIDOR – catalogue of services and repositories for research data. (Inist -CNRS)

Where to disseminate? Data

Domain-specific tools :

DATACC'– Research Data Management support platform in Physics and Chemistry (fr/eng)

Repositories list in biomedical studies (National Institutes of Health)

Where to disseminate? Data

Pros/Cons

Mode de publication	Recherche et citabilité	Paternité et crédits auteurs	Volumétrie	Réutilisabilité
Données intégrées	★★☆☆	★★★★	★☆☆☆	★☆☆☆
Matériel supplémentaire	★★☆☆	★★★★	★★☆☆	★☆☆☆
Data paper	★★★★	★★★★	★★★★	★★★☆

DoRANum. Données de la recherche : apprentissage numérique [En ligne].France : DoRANum; 2017. Comment publier des données de recherche [modifié le 28 mai 2018 ; consulté le 17 septembre 2018]. Disponible : https://doranum.fr/data-paper-data-journal/comment-publier-donnees-recherche/

Where to disseminate? Data

Datapapers (data articles / data descriptors)

Definition :
Articles that describe data sets, rather than analysis results.
In traditional journals or in "data journals".

Practice :
Data can be deposited in a repository, recommended by the publisher or chosen by the author.

Benefits :
Publication like any published scientific article.
The datapaper is peer-reviewed

Recommendation :
Deposit the post-print of the publication in an open archive (HAL, arXiv, etc.).

Where to disseminate? Data

Datapapers

Examples (datajournals) :
- Data Science Journal – codata
- Scientific Data (Nature)

- Lists :
- "Where to publish" of Cirad (select data papers)
- The Global Biodiversity Information Facility's (Gbif) list
- Datacc' How and where to publish your data paper?

Where to disseminate?
Codes and software

Software Heritage

Universal code and software archive
Supported by UNESCO

Why?

Facilitates research (by domain)
Promotes software
Citation
Facilitates reuse

How to disseminate code: best practices

- Use a software forge harvested by SoftWare Heritage
- Describe the code (metadata) with code-meta
- Deposit the code in HAL with a link to the SWID (perennial code identifier)
- Publish your code and software: Software Heritage list (2021)

Where to disseminate?

Linking publications, data and code with doi :

https://www.nature.com/articles/s41557-019-0329-3#relevant-content-header

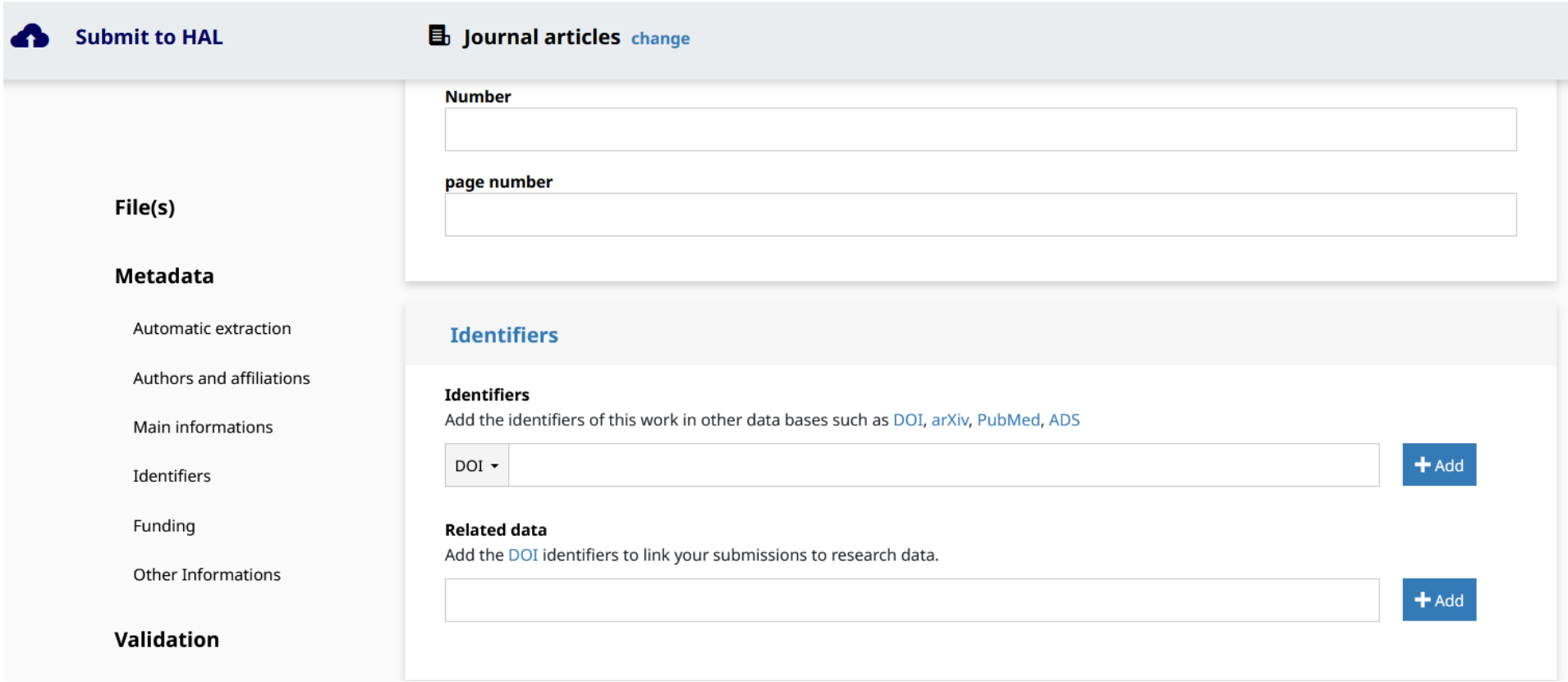
- preprint on BioRxiv https://www.biorxiv.org/content/10.1101/476432v3

- code on gitlab : https://github.com/fraser-lab/solution_scattering

- Data on NIH Figshare:
https://nih.figshare.com/articles/dataset/X-ray_scattering_curves_SAXS_WAXS_used_for_the_analysis_described_in_Temperature-Jump_Solution_X-ray_Scattering_Reveals_Distinct_Motions_in_a_Dynamic_Enzyme_/9177143

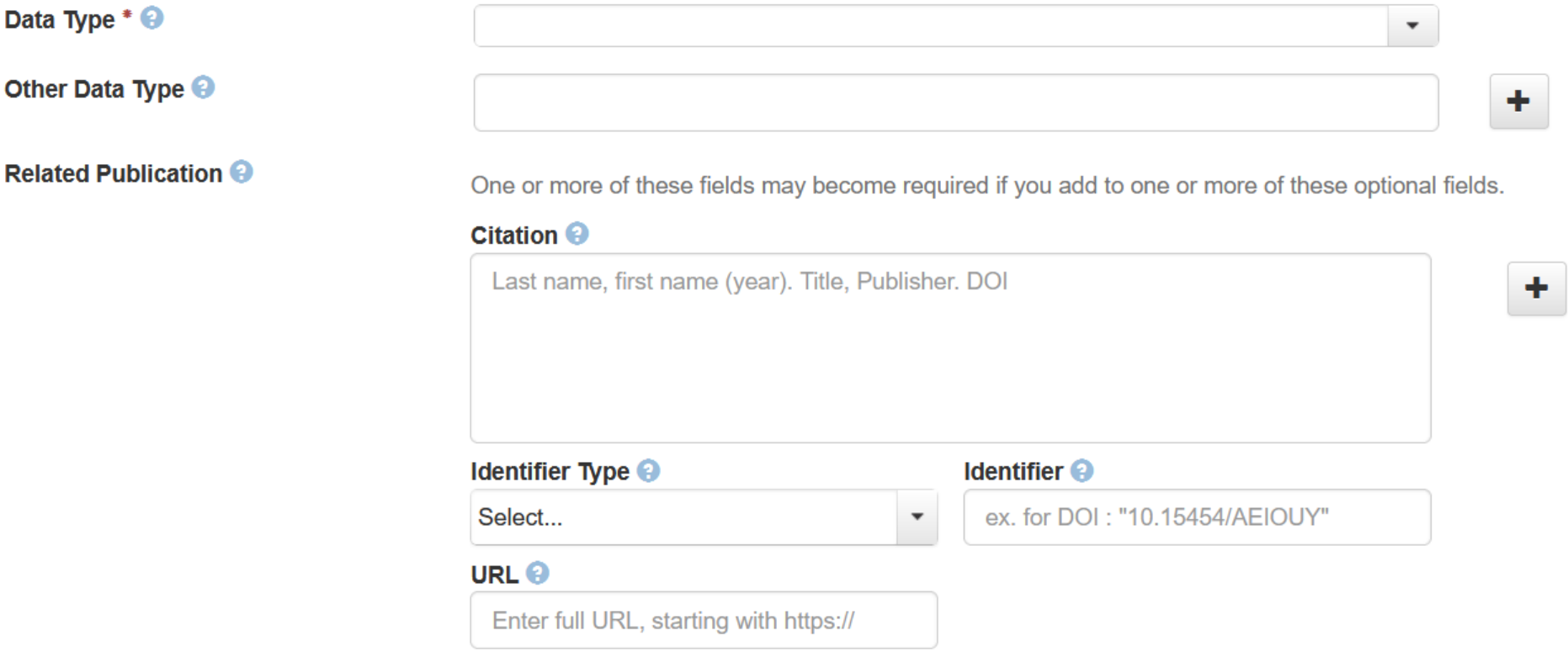
Where to disseminate?

Linking publications with HAL



Where to disseminate?

Linking publications with Recherche Data Gouv



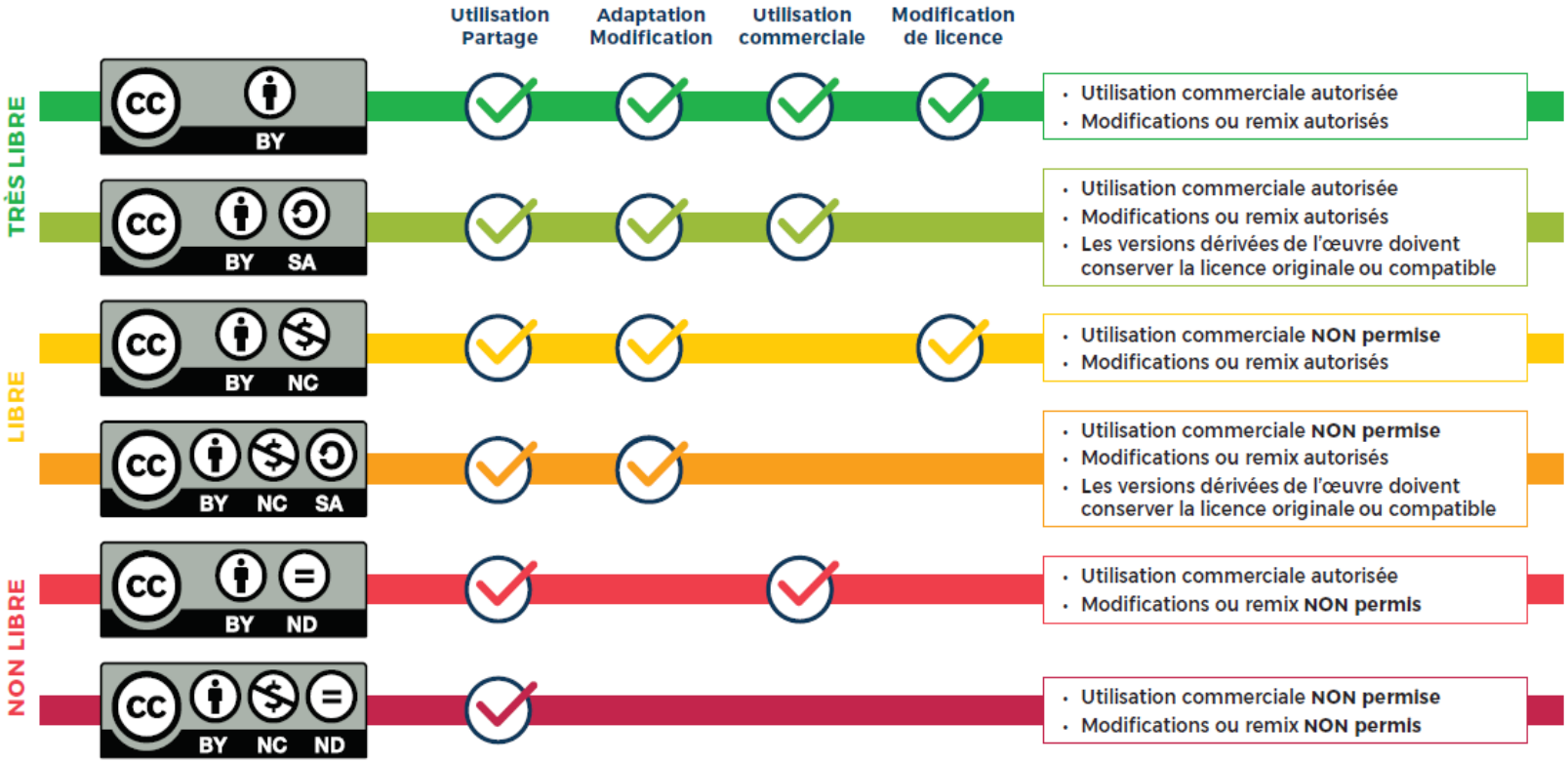
How to disseminate?

Data :

- Define the use of deposited data via a license

- Licences **Creative Commons**
  - BY : Attribution
  - SA : Share Alike
  - NC : Non Commercial
  - ND : Non Derivative

Les licences Creative Commons



BY
ATTRIBUTION
Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.
L'auteur doit être cité

SA
PARTAGE DANS LES MÊMES CONDITIONS
Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.
Partage sous licence compatible

NC
POUR USAGE NON COMMERCIAL
Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.
Pour usage non commercial

ND
PAS DE MODIFICATION
Vous pouvez retenir, réutiliser et redistribuer.
Aucune modification permise

fabrique REL
RESSOURCES ÉDUCATIVES LIBRES

L'auteur de l'œuvre doit être cité en toutes circonstances pour l'ensemble de ces licences.

How to disseminate?

Databases :
Open Data Commons Open Database License (OdbL)

Code :

- Free to run the program, for any purpose.
- Free to study how the program works, and adapt it to your needs.
- Free to redistribute copies.
- Free to improve the program and publish your improvements

How to disseminate?

Code:
- no copyleft
- weak copyleft
- strong copyleft

For example GNU GPL = GNU General Public License.

Type	Exemple de licences
Sans copyleft	BSD license, Apache License 2 MIT
Copyleft faible	GNU library or « Lesser »General Public License (LGPL)
Copyleft fort	GNU General Public, License EUPL

In France, list of licenses like Etalab 2.0

Help

Tools :

- **Help on disseminating research data (French) - Cirad - (in some cases)**

  Choose an open source license
  License Selector (code & data)
  Licentia by Inria

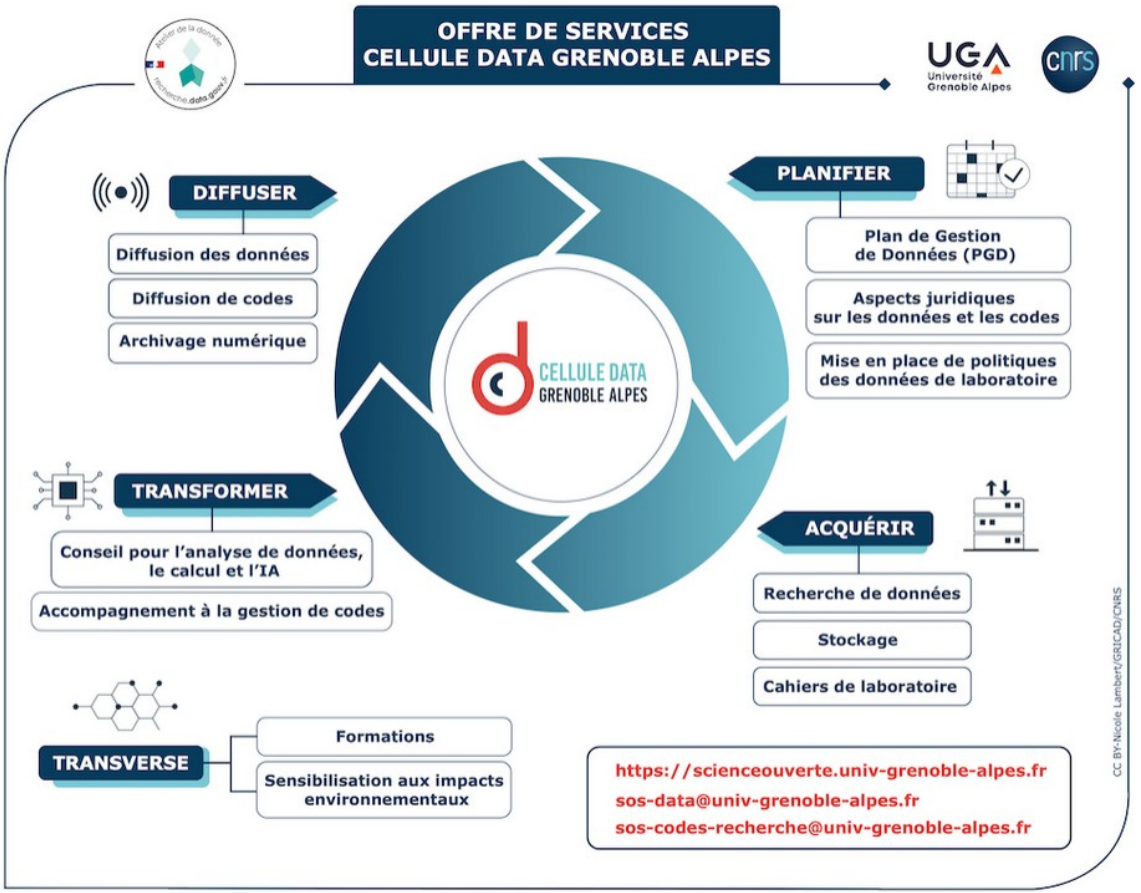
Help

Data, codes and software :

Cellule Data Grenoble Alpes (CDGA)
Services
Network of data referents

A support address : sos-data@univ-grenoble-alpes.fr

Help


OFFRE DE SERVICES
CELLULE DATA GRENOBLE ALPES
UGA Université Grenoble Alpes
cnrs
DIFFUSER
Diffusion des données
Diffusion de codes
Archivage numérique
PLANIFIER
Plan de Gestion de Données (PGD)
Aspects juridiques sur les données et les codes
Mise en place de politiques des données de laboratoire
CELLULE DATA GRENOBLE ALPES
TRANSFORMER
Conseil pour l'analyse de données, le calcul et l'IA
Accompagnement à la gestion de codes
ACQUÉRIR
Recherche de données
Stockage
Cahiers de laboratoire
TRANSVERSE
Formations
Sensibilisation aux impacts environnementaux
https://scienceouverte.univ-grenoble-alpes.fr
sos-data@univ-grenoble-alpes.fr
sos-codes-recherche@univ-grenoble-alpes.fr
CC BY-Nicole Lambert/GRICAD/CNRS

Help

- The **UGA's Open Science website**
- « webseries **Passport : introduction to Open Science** »
  - 5 videos
  - 3 booklets
    - Passport for Open Science – A Practical Guide for PhD Students
    - Code et software
- DoraNum
- The Turing Way
- INIST :
  Introduction to managing and sharing research data
  Class "Understanding Open Science" (French)
- CIRAD's website, Coopist, Managing data
- Guide to best practices in research data management (CNRS)