User Guide - Cat.InDoRES Catalog

Preface

Before publishing any dataset in the data.InDoRES repository, it is necessary to complete its metadata record in cat.InDoRES. Below is a step-by-step guide to assist you in entering your metadata.

This guide was originally developed within the framework of the Zone Atelier Network. We would like to thank the CIGAL Metadata Group, and in particular Guillaume Ryckelynck and Grégoire Skupinski (Urban Environmental Zone Atelier, LIVE Laboratory), for their support and contributions. Special thanks also go to Dominique Andrieu (Loire Zone Atelier, MSH Val de Loire) for his thorough reviews and valuable advice during the development of the initial version of this guide.

This guide is regularly updated to comply with current standards and norms in force within the European Union, ensuring interoperability.

Introduction

Metadata entry, although a legal requirement—particularly under the European INSPIRE Directive, remains a challenge for anyone wishing to share information about their dataset.

Since 2013, as part of a partnership with **CIGAL** (**C**oopération pour l'**I**nformation **G**éographique en **AL**sace) , the goal of creating an inventory of geographic information led to the formation of a "Metadata" working group. This group developed a metadata entry form aligned with current standards and directives. It was quickly decided to adopt the ISO 19115/19139 standards and the INSPIRE Directive to meet national and European requirements for geographic data. A field mapping table was then created, resulting in a simple and user-friendly form.

Building on CIGAL's work, a new metadata form has been developed—fully compliant with standards, ensuring seamless interoperability and precise data description, tailored for research teams.
Available in Excel format, it includes a dedicated help section, compiled here into a comprehensive guide. For each field, you'll find clear definitions, completion tips, and concrete examples.
The form can be exported to any Inspire- or ISO 19139-compliant portal (e.g., GeoNetwork, GeoSource) and meets national and European metadata requirements, *ie* standard Dublin Core.

For any questions, feel free to contact us at this address: contact@indores.fr

Amandine Hénon & Noémie Tomadini
UAR 2047 DoHNEE Muséum national d'Histoire naturelle

1. Data description

1.1 Data language

🔷 **Definition**

This refers to the language used to describe the data in the metadata record. Three languages are available in the dropdown list: French, English, and German.

🔷 **Recommendations**

The default value is set to "French," as it is the most commonly used language. If your dataset description is available in multiple languages, it's best to create one metadata record per language—avoid mixing languages within the same form.

🔷 **Example**

```
french
```

1.2 Data title

🔷 **Definition**

This is the name that identifies and describes your dataset clearly and concisely. It should be explicit, distinctive, and unique—avoid acronyms. A well-crafted title helps users quickly understand the content and purpose of your data.

🔷 **Recommendations**

It’s recommended to include in the title:
-The data type (e.g., map, thematic data, reference dataset),
-The official or common name of the data (avoiding unexplained acronyms),
-The version or year of production,
-The name of the geographic area covered.
➡️ This ensures clarity, uniqueness, and easy identification.

🔷 **Example**

```
"Piezometric map of the lower aquifer in the Bregnier-Cordon area,
November 22–23, 1979" includes:
Type: Map
Theme: Piezometric (clearly describes the data)
Geographic scope: Lower aquifer, Bregnier-Cordon area
Date: November 22–23, 1979 (data production period).
```

1.3 Abstract

🔷 **Definition**

The abstract provides a clear description of the dataset, using either an official definition (if available) or a commonly accepted one. It should clearly explain the information or phenomenon represented, making the data understandable to users. This is an essential part of the metadata record.

🔷 **Recommendations**

Any information that helps contextualize the data is welcome. The more details you provide, the better your dataset will be understood. Include relevant background, purpose, or usage notes to enrich its overall understanding.

1.4 Dates and Update Frequency

🔷 **Definition**

These details clarify key dates in your dataset's lifecycle:
-**Creation**: When the data was originally produced.
-**Publication**: When it was first released or shared.
-**Update**: The date of the latest revision or version.
-**Frequency**: How often updates occur (e.g., annual, as needed).

🔷 **Recommendations**

At least one date must be provided, in YYYY-MM-DD format using hyphens (keyboard key 6) as separators.

➡️ The creation date is required at minimum. The update date only makes sense when paired with the creation date.

➡️ For "living" data (updated daily or multiple times per week), clearly specify the update frequency—but still include the original creation date.

➡️ If the frequency isn't listed, choose "Unknown" (default). If no updates are planned, select "Not planned".

> ⚠️ *The creation date refers to when the dataset itself was created—not the date the metadata was written.*

🔷 **Exemple**

```
Creation date: 1981-07-23
Update Frequency: Continuous
```

1.5 Contacts for the data

🔷 **Definition**

This field identifies the individuals or organizations to contact for any information about the data. These may be specific persons or institutional contacts.

🔷 **Recommendations**

As with other fields, the name must be entered in uppercase. The "Role" dropdown offers several options:

- **Provider**: The organization that supplies the resource, delivering it directly or via a distributor.
- **Manager**: Responsible for maintaining and updating the data.
- **Owner**: Holds the intellectual property rights.
- **Contact Point**: The primary person to reach for information.
- **Author**: Holds moral rights over the resource.

Multiple entries are allowed if different roles apply, but at least one **Contact Point** is required. Always provide full details: name in uppercase, first name capitalized, organization (avoid acronyms), and email. The role is essential.

> ⚠️ *For students (Master's or PhD), please use a long-term email address (non-educational) to ensure you can be reached in the future.*
>
> ⚠️ *Generic addresses (e.g., accueil@mnhn.fr) are strongly discouraged—use a personal, professional contact instead.*

🔷 **Example**

```
• Name : DUPONT Marc
• Organization : Muséum national d'Histoire naturelle, UAR 2047 DoHNEE
(Données en Histoire Naturelle, Ecologie, Environnement)
• Email : exemple@mnhn.fr
• Role : Contact Point
```

1.6 Thesaurus and Keywords

🔷 **Definition**

Using keywords and thesauri greatly improves your dataset's visibility. Here are the available thesauri (with auto-complete support):

- GEMET ou GEneral Multilingual Environmental Thesaurus;
- Biodiversity, Key concepts in biodiversity sciences;
- Spatial Data Service Category - This thesaurus is based on the taxonomy of geographic services defined in the EN ISO 19119 standard
- Agri-food vocabulary - This thesaurus covers key concepts in agri-food science and technology, supporting precise and consistent indexing of datasets in this domain.
- Ethnologie - This is a controlled vocabulary used to index bibliographic references in the FRANCIS database under the "Ethnology" category;
- Chimie - This vocabulary includes over 9,500 chemistry concepts aligned with major ontologies like ChEBI, RXNO, and FIX, enabling precise, interoperable annotation of chemical data;
- Sociologie - A controlled vocabulary used to index bibliographic references in the FRANCIS database under "Sociology";
- Vocabulaire thématique de géographie - A selected indexing vocabulary used to build the International Geographical Bibliography;
- Vocabulaire des sciences administratives - A controlled vocabulary used to index bibliographic references in the FRANCIS database for "International Public Administration Science.
- Geonames - A free geographic database under Creative Commons license, used to define geographic keywords in the "Régions_FR" and "Pays et subdivisions" fields.

🔷 **Recommendations**

➡️ For free keywords, use **lowercase**, include **accents**, and write in **singular** form. If the term comes from another thesaurus, add its name or even the keyword's URI if known.

1.7 INSPIRE Themes

🔷 Definition

This field classifies your data under one or more themes from a standardized European list defined by the INSPIRE Directive improving discoverability and interoperability. Choose the most relevant theme(s) to ensure your dataset is easily found and correctly categorized.

🔷 Recommendations

Whenever possible, select just **one** INSPIRE theme—only use multiple if truly necessary. Ensure consistency with the "International Categories" field. The available values follow the official INSPIRE list, based on the GEMET thesaurus:

- **Coordinate Reference Systems**: Define geographic locations using coordinates (x, y, z) or latitude, longitude, and altitude, based on a fixed geodetic reference.
- **Geographic Grid Systems**: Harmonized multi-resolution grids with standardized cell size, origin, and positioning.
- **Geographical Names**: Official names of places, regions, cities, or other geographic features of public or historical interest.
- **Administrative Units**: Areas bounded by administrative borders, used for local, regional, or national governance.
- **Addresses**: Location identifiers based on street name, house number, and postal code.
- **Cadastral Parcels**: Land parcels defined in cadastral registers or equivalent systems.
- **Transport Networks**: Road, rail, air, and water networks with associated infrastructure, including intermodal connections and the trans-European transport network.
- **Hydrography**: Water bodies (marine, inland), river basins, and related features, aligned with the EU Water Framework Directive.
- **Protected Sites**: Areas designated under international, EU, or national law for conservation.
- **Altitude**: Digital terrain models, including land elevation, seafloor depth (bathymetry), and coastline.
- **Land Cover**: Physical surface cover—artificial areas, agriculture, forests, wetlands, water bodies.
- **Orthoimagery**: Georeferenced aerial or satellite images of the Earth's surface.
- **Geology**: Rock composition, structure, aquifers, and geomorphology—key for understanding Earth’s subsurface.

- **Statistical Units**: Geographic areas used for collecting and disseminating statistical data.
- **Buildings**: Location and footprint of residential, commercial, and public structures.
- **Soils**: Characterized by depth, texture, organic content, erosion, and water retention—vital for agriculture and environment.
- **Land Use**: Functional classification of land—residential, industrial, recreational, etc.—based on current or planned socio-economic activity.
- **Human Health & Safety**: Geographic distribution of health issues linked to environmental factors (pollution, noise, chemicals) and well-being indicators (stress, fertility decline, epidemics).
- **Utility & Public Services**: Infrastructure for waste, energy, water, and public social services like hospitals, schools, and civil protection sites.
- **Environmental Monitoring Facilities**: Locations and operations of sites measuring environmental conditions—emissions, biodiversity, ecosystem health—by or for public authorities.
- **Production & Industrial Sites**: Industrial plants (under EU pollution control directives), water abstraction, mining, and storage facilities.
- **Agricultural & Aquaculture Facilities**: Farms, greenhouses, irrigation systems, and fish farming installations.
- **Population Distribution / Demography**: Geographic spread of people, with demographic and activity data grouped by grid, region, or administrative unit.
- **Management/Restriction Zones**: Regulated areas for reporting or control—e.g., protected water zones, noise limits, mining permits, coastal management, and nitrate-vulnerable areas.
- **Marine Regions**: Seas and saline water bodies divided into areas with shared physical characteristics.
- **Biogeographical Regions**: Zones with similar ecological conditions and shared species or habitats.
- **Oceanographic Features**: Physical ocean conditions—currents, salinity, wave height, temperature.
- **Energy Resources**: Locations of fossil fuels, hydropower, solar, wind, and bioenergy, including depth/height data when relevant.
- **Natural Risk Zones**: Areas prone to natural hazards—floods, landslides, earthquakes, volcanic eruptions, wildfires—with potential societal impact.
- **Atmospheric Conditions**: Geographic data on weather and climate, from measurements, models, or both, including sensor locations.

- **Meteorological Geographical Features**: Weather conditions—precipitation, temperature, wind speed/direction, and evapotranspiration—along with measurement locations.
- **Habitats & Biotopes**: Ecologically distinct areas (terrestrial or aquatic) defined by specific abiotic and biotic conditions that support life—natural or semi-natural.
- **Species Distribution**: Geographic occurrence of animal and plant species, grouped by grid, region, or administrative unit.
- **Mineral Resources**: Locations of metallic ores, industrial minerals, and related data on depth or elevation.

You'll find below the correspondence table between international categories (ISO) and the European INSPIRE thematic classification of data:

> ⚠️ ***When describing a dataset, it's essential to ensure consistency between these two value lists.***

	INSPIRE Themes	Category ISO
Annexe 1	Coordinate reference systems	--
	Geographical grid systems	--
	Geographical names	Location
	Administrative units	Boundaries
	Addresses	Location
	Cadastral parcels	Planning / Cadastre
	Transport networks	Transportation
	Hydrography	Inland Waters
	Protected sites	Environment
Annexe 2	Elevation	Elevation
	Land cover	Imagery / baseMaps / earth cover
	Orthoimagery	Imagery / baseMaps / earth cover
	Geology	Geoscientific information
Annexe 3	Statistical units	Boundaries
	Buildings	Structure
	Soil	Geoscientific information
	Land use	Planning / Cadastre
	Human health and safety	Health
	Utility and governmental services	Utilities / Communication
	Environmental monitoring facilities	Structure
	Production and industrial facilities	Structure
	Agricultural and aquaculture facilities	Farming
	Population distribution and demography	Society
	Area management / restriction / regulation zones & reporting units	Planning / Cadastre
	Natural risk zones	Geoscientific information
	Atmospheric conditions	Climatology / Meteorology / Atmosphere
	Meteorological geographical features	Climatology / Meteorology / Atmosphere
	Oceanographic geographical features	Oceans
	Sea regions	Oceans
	Bio-geographical regions	Biota
	Habitats and biotopes	Biota
	Species distribution	Biota
	Energy Resources	Economy
	Mineral Resources	Economy

1.8 International categories

🔷 **Definition**

This field classifies the data into one or more categories from a closed, international list, helping users find the data more easily. It's important to associate the dataset with the most relevant theme(s).

🔷 **Recommendations**

It's recommended to select just one category, except in rare cases. Consistency with the next field (INSPIRE Themes) is also important. The list below follows the ISO 19115 standard:

- **Agriculture** (livestock and/or crops)
- **Biota** (flora and fauna in a natural ecosystem)
- **Boundaries** (national, provincial, or administrative boundaries)
- **Climatology/Meteorology** (atmospheric processes and phenomena)
- **Economy** (economic activities and employment)
- **Elevation** (altitude, seafloor and land topography)
- **Environment** (natural resources, conservation, protection)
- **Geoscientific Information** (earth sciences data)
- **Health** (public health, healthcare services, epidemiology)
- **Imagery/Base Maps/Earth Cover** (land cover, aerial and satellite imagery, thematic maps)
- **Military** (military bases and infrastructure)
- **Inland Waters** (rivers, lakes, glaciers, hydrographic systems)
- **Location** (postal zones, addresses)
- **Oceans** (marine environment components and characteristics)
- **Planning/Cadastre** (land-use maps, zoning plans, risk prevention planning)
- **Society** (social and cultural characteristics)
- **Structure** (human-made structures: museums, churches, factories)
- **Transportation** (infrastructure for moving people and goods)
- **Utilities/Communication** (distribution networks: telecom, energy, water; waste collection)

1.9 Temporal Extent

🔷 **Definition**

The Temporal Extent refers to the time period covered by the dataset—when the data was collected, observed, or applies. It can include:

➡️ A single date (e.g., a satellite image from 2023-05-15)

➡️ A range (e.g., 2009-10-15 _ 2009-11-01)

🔷 **Recommendations**

As with other date fields, the date must be expressed in the format: **YYYY-MM-DD**, using hyphens (from the keyboard's minus key) as separators. For temporal extents, the default reference is the Gregorian calendar. If another system is used (e.g., geological eras), specify it in the **Description** field.

🔷 **Example**

```
Start date : 2009-10-15
Final date : 2009-11-01
```

1.10 Geographic Extent

🔷 **Definition**These fields define the geographic extent—the area where data was collected. It's represented as a bounding box in WGS 84 (World Geodetic System 1984), using decimal degrees with at least two decimal places, referenced to the Greenwich meridian. The extent includes:

➡️ A descriptive name (e.g., city, nature reserve), if available

➡️ west (min longitude), east (max longitude), south (min latitude), north (max latitude)

🔷 **Recommendations**

The geographic extent should cover the full area for which the producer guarantees data completeness. Even confirmed absence of features is meaningful—so the bounding box may sometimes be larger than the actual data coverage.

> ⚠️ ***For accuracy, the bounding rectangle should be as tight as possible to reflect the true data footprint.***

Multiple non-contiguous areas can be described using several bounding boxes.

🔷 **Example**

```
City : Strasbourg
East (max longitude): 7,84
West (min longitude): 7,69
North (max latitude): 48,65
South (min latitude): 48,89
```

2. Data Quality

2.1 Described Level

🔷 **Definition**

This technical field specifies the scope of the dataset description—whether it applies to a standalone dataset or a collection of datasets.

🔷 **Recommendations**

If the data is part of a larger series (e.g., annual land cover maps), choose *Collection of data*. If it's a single, self-contained layer (e.g., a specific map), choose *Dataset*.

🔷 **Example**

```
Dataset
```

2.2 Quality Statement

🔷 **Definition**

Data quality description doesn't judge whether data is "good" or "bad," but ensures its quality level matches the intended use and user needs. It clearly outlines limitations through completeness and methodology. The text should include **data lineage**, production methods, funding sources, and key project partners—providing full transparency on how and why the data was created.

🔷 **Recommendations**

It is recommended to be as precise as possible when describing data quality. Information to include may consist of:

- **Completeness**: Which part is covered by the dataset? What is excluded?
- **Planimetric positioning accuracy**: Provide additional details beyond spatial resolution, specifying optimal conditions for data use and exploitation.
- **Quality control**: Is a quality control process applied? If so, describe its general rules and results. Are there attributes indicating data quality (e.g., a precision code)?
- **Production method**: Is there a technical specification document or project

requirements file?

- **Technical usage limitations**: Indicate any known limitations or unsuitable applications for the dataset.
- **Lineage and history of the data**: A free-text section describing the data's history, key production phases, and source datasets used (e.g., reference data).
- **For 3D data**: Specify vertical extent and the vertical reference system.
- **Temporal extent**: Include relevant time period information.

It is also possible to link to additional documentation.

> ***In summary, this text helps contextualize the data to prevent misuse and ensure results are properly understood.***

3. Data Distribution

3.1 Distribution Format(s)

🔷 **Definition**

These details indicate the format(s) in which the data is available from the provider. It's recommended to specify the format version (e.g., GeoPackage 1.3, Shapefile ESRI, CSV UTF-8).

🔷 **Recommendations**

Prioritize widely used, open formats. When possible, write the full format name along with the file extension (e.g., "Comma-Separated Values (CSV)."

🔷 **Example**

```
Shapefile (shp) 1.0
```

3.2 Legal Constraints on Access and Use

> ***The information below covers the legal aspects of data access and use, such as licenses, copyright, and usage restrictions. Technical limitations (e.g., scale, accuracy) should already be described in the data quality section, as previously explained.***

3.2.1 Public Access Constraints

🔷 **Definition**

This information specifies the administrative and legal conditions for accessing and using the data, such as required authorizations, licenses, or usage restrictions.

🔷 **Recommandations**

The proposed values follow the ISO 19115 standard (liste B.5.24. MD_RestrictionCode). It is a closed list—only the defined terms may be used:

> ⚠ *By default, the value will be set to "Intellectual Property Right / Patrimonial Right" as defined in ISO 19115.*

N°	Title	Reason for the restriction
0	No restrictions on public access	There are no restrictions.
1	Copyright / Moral Rights	Restriction related to the exercise of moral rights.
2	Patent	Restriction resulting from the existence of a patent.
3	Patent pending	Restriction resulting from the filing of a pending patent.
4	Trademark	Restriction related to the existence of a trademark registration.
5	Licence	Restriction resulting from the existence of a license.
6	Intellectual Property Right / Patrimonial Right□	Restriction related to the exercise of property rights.
7	Restricted	⚠ **This value has no identified application in France.**

🔷 Exemple

```
"Intellectual Property Right / Patrimonial Right"
```

3.2.2 Use Constraints

🔷 Definition

This information specifies the administrative and legal conditions for accessing and using the data.

🔷 Recommendations

The list of available values follows the ISO 19115 standard (liste B.5.24. MD_RestrictionCode _ see table above).

> ⚠ *By default, the value will be set to "License". In France, the Loi pour une République numérique mandates open public access to research data by default, unless restricted for legal reasons (e.g., GDPR, national defense). The CC BY 4.0 license is ideal for scientific datasets: it allows unrestricted reuse (including commercial and modified use), provided the producer, source, and any modifications are clearly attributed.*

🔷 Example

```
« Licence »
```

3.2.3 Other Conditions and Legal Notices for Use

🔷 **Definition**

This information informs the user about the administrative and legal conditions for data use. It complements the previously provided details on public access constraints related to INSPIRE and other restrictions. It should specify, in particular, conditions regarding licensing, formal agreements, legal notices, usage limitations in commercial or competitive contexts, etc.

🔷 **Recommendations**

Technical limitations on use and exploitation related to data quality should be described in the dedicated field.
→ If no additional legal conditions for use or access apply, state: *"No further legal conditions apply."*
→ If conditions are unknown, state: *"Conditions unknown."*
It is recommended to clearly indicate any required legal notices and source attributions that must appear on any materials distributing or using the data.

🔷 **Example**

```
Terms of Use and Legal Notices:
- "Commercial use, direct or indirect, is prohibited"
- "Legal notices and sources to be credited on any materials using the
data: ..."
```

3.3 Access to Data and Associated Documents

🔷 **Definition**

This field lists links to the data or related resources, such as direct download URLs, technical specifications, commitment agreements, or relevant scientific publications. Each URL must be accompanied by a clear name and description indicating the nature and content of the resource.

🔷 **Recommendations**

Use complete, **public**, and persistent URLs (no relative paths).

> ⚠️ *Although the Excel template provides only three fields for "Access to Data and Associated Documents," you may include as many online resources as needed. Feel free to send us a more comprehensive list.*

4. Illustration

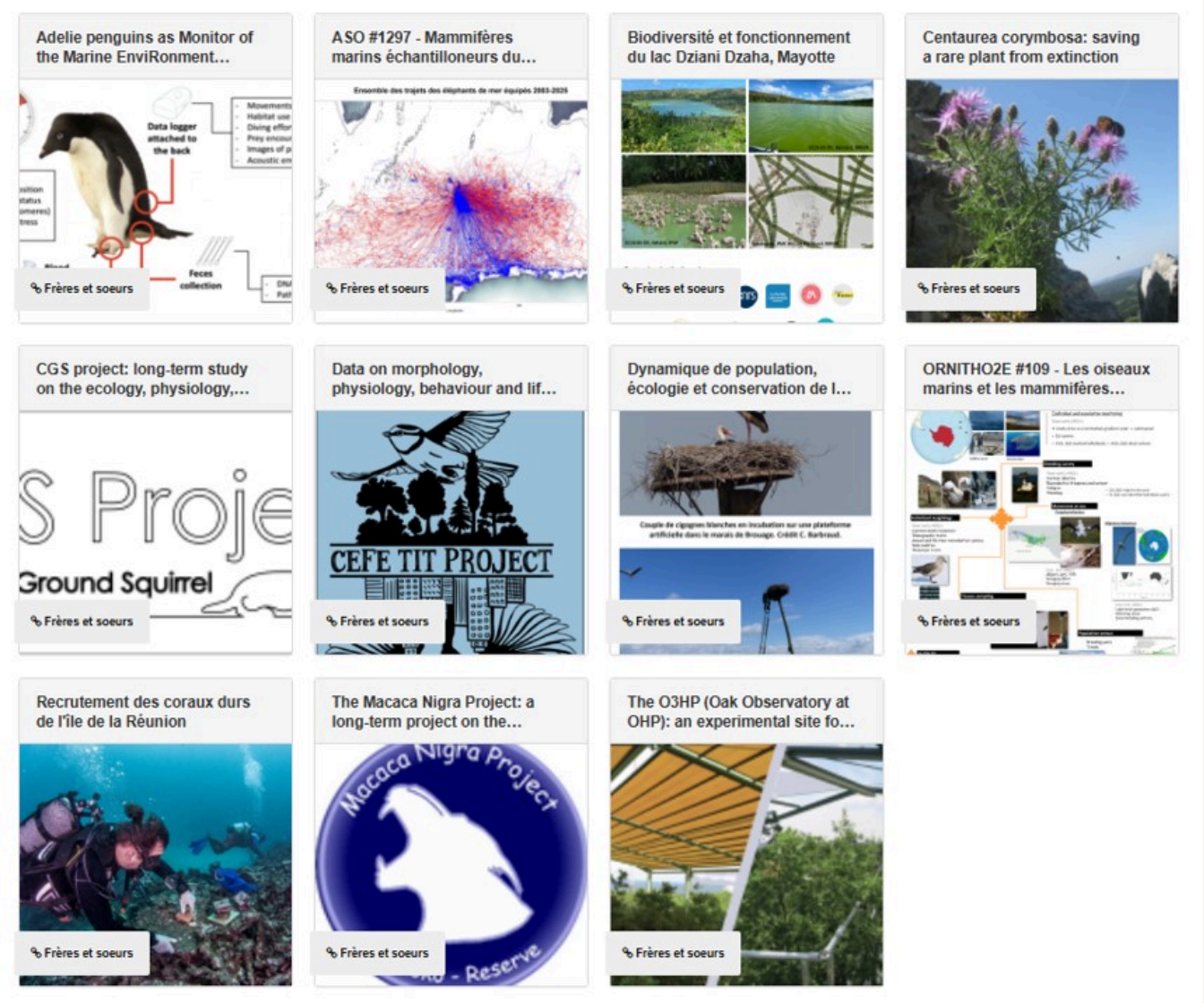
🔷 Definition

This field provides a link to an image representing the dataset. It may be a photo of the study subject, a logo, a graphic, or a map.

🔷 Recommendations

Provide a sufficiently large image (20 cm high at 72 dpi) in .jpeg format. Send the illustration along with the metadata form, including a caption and any required credits.

➡️ Not sure which image to choose? Feel free to send several—we can create a composite or select the most suitable one.



5. Metadata Information

5.1 Metadata Contact

🔷 Definition

This field identifies the person or organization responsible for the metadata record.

🔷 Recommendations

At least one contact must be provided. Include full details: last name in uppercase, first name capitalized, full organization name (avoid acronyms), and a valid email address.

> ⚠️ *Students (Master's/PhD): Use a long-term email (non-educational) to ensure you can be reached in the future.*
>
> ⚠️ *Avoid generic emails (e.g., accueil@mnhn.fr).*

🔷 Example

- Name : DUPONT Marc
- Organization : Muséum national d'Histoire naturelle, UAR 2047 DoHNEE (Données en Histoire Naturelle, Ecologie, Environnement)
- Email : exemple@mnhn.fr
- Role : Point of contact