

An ULiège research data repository – the ULiège Dataverse

1st ULiège Open Science Day – Oct 27th, 2023



Judith Biernaux, PhD

RISE

Place du XX Août, 7 (Bât. A1)

B-4000 Liège

Tél : +32 (0)4 366 55 14

Jbiernaux@uliege.be



Open data is the destination...

Increasing pressure to make research data available for re-use

Most EU funding programmes refer to the « **as open as possible, as closed as necessary** » principle.

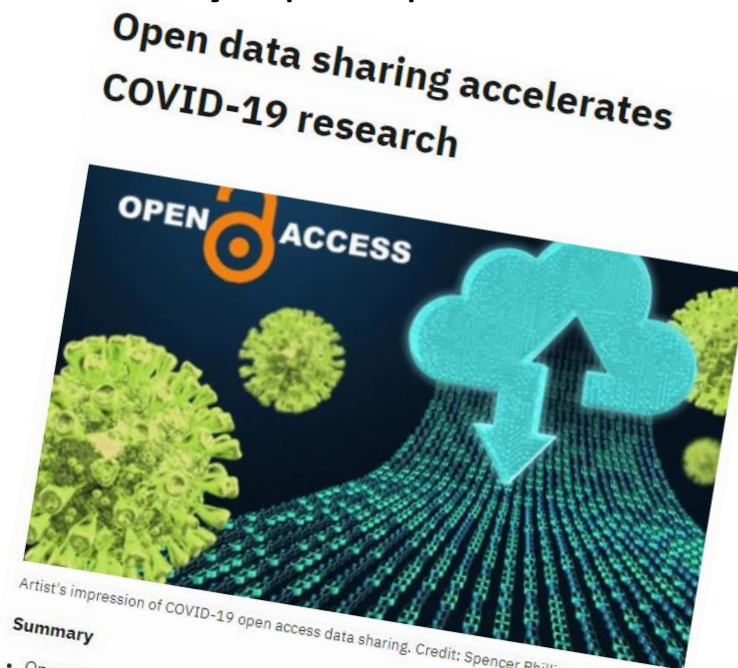
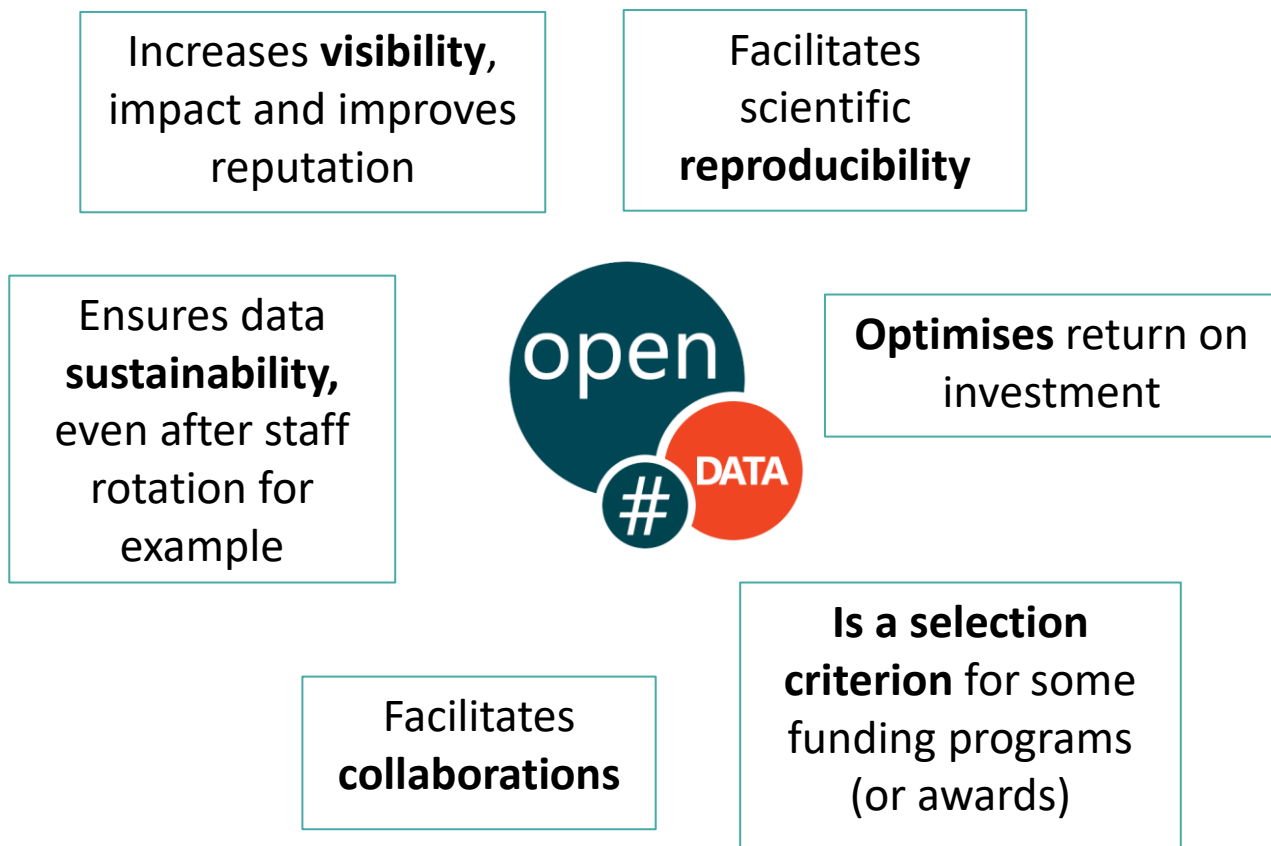
The aim is therefore to practice as much **open data** as possible.

Open data is the destination...

Increasing pressure to make research data available for re-use

Most EU funding programmes refer to the « **as open as possible, as closed as necessary** » principle.

The aim is therefore to practice as much **open data** as possible.



Artist's impression of COVID-19 open access data sharing. Credit: Spencer Phillips

- Summary**
- Open access increases the visibility of research data and information, giving scientists the ability to build upon and react to existing research quickly
 - EMBL-EBI launched the European COVID-19 Data Platform to enable rapid access to datasets and results pertaining to the SARS-CoV-2 outbreak
 - Open access data sharing has greatly accelerated COVID-19 research and helps further our understanding of the biology, transmission, and spread of the SARS-CoV-2 virus

Open data is the destination...

Increasing pressure to make research data available for re-use

Most EU funding programmes refer to the « **as open as possible, as closed as necessary** » principle.

The aim is therefore to practice as much **open data** as possible.

However, there is a right way to practice open data

Open data is the destination...

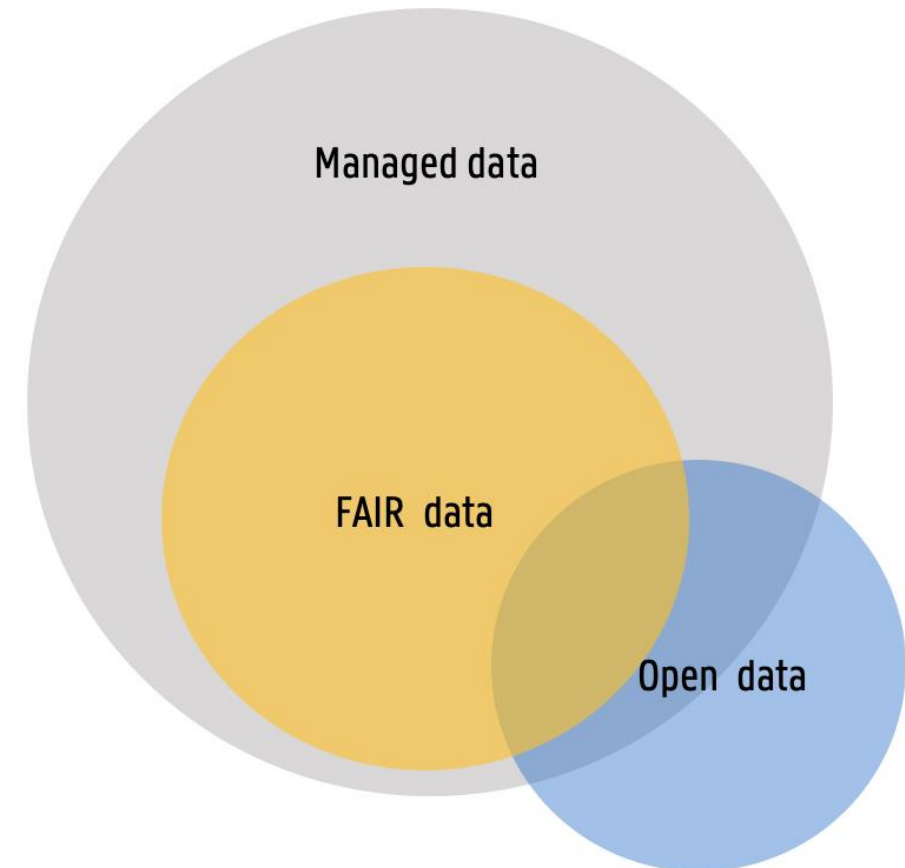
Increasing pressure to make research data available for re-use

Most EU funding programmes refer to the « **as open as possible, as closed as necessary** » principle.

The aim is therefore to practice as much **open data** as possible.

However, there is a right way to practice open data

- Only if no applicable regulations or contractual obligations prevent the publication
- According to the **FAIR principles**, so that it is of added value to the scientific community



...FAIR data is the journey

Findable

Accessible

Interoperable

Reusable

...FAIR data is the journey

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets.

Metadata, PID such as a DOI, keywords, abstract, author, date, versions...

Accessible

Interoperable

Reusable

...FAIR data is the journey

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets.

Metadata, PID such as a DOI, keywords, abstract, author, date, versions...

Interoperable

Accessible

Once the user finds the required data, they need to access it. The protocol is open, free, and allows for an authentication and authorisation procedure, where necessary

Metadata are accessible, even when the data are no longer available

Reusable

...FAIR data is the journey

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets.

Metadata, PID such as a DOI, keywords, abstract, author, date, versions...

Interoperable

The data usually need to be integrated with other data. It needs to interoperate with different workflows, systems, software, applications, languages, institutions, countries, ...

Accessible

Once the user finds the required data, they need to access it. The protocol is open, free, and allows for an authentication and authorisation procedure, where necessary

Metadata are accessible, even when the data are no longer available

Reusable

...FAIR data is the journey

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets.

Metadata, PID such as a DOI, keywords, abstract, author, date, versions...

Interoperable

The data usually need to be integrated with other data. It needs to interoperate with different workflows, systems, software, applications, languages, institutions, countries, ...

Accessible

Once the user finds the required data, they need to access it. The protocol is open, free, and allows for an authentication and authorisation procedure, where necessary

Metadata are accessible, even when the data are no longer available

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this :

- The format is non proprietary
- The documentation is sufficient
- There is an appropriate license
- The provenance of the data is clear
- The (meta)data meets community standards

...FAIR data is the journey

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automation. Metadata, PID, author, date, version, etc.

Accessible

Once the user finds the data, they need to be able to access it. Open access allows for an open procedure, where the data are not behind a paywall.

Publishing data in a dedicated data repository usually checks some of the boxes
 But how to choose a repository?

Interoperable

The data usually need to be used with other data. It needs to interoperate with different workflows, systems, software, applications, languages, institutions, countries, ...

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this :

- The format is non proprietary
- The documentation is sufficient
- There is an appropriate license
- The provenance of the data is clear
- The (meta)data meets community standards

How to choose a data repository?

A good data repository:

- Is **recognized** by your peers and matches the **community standards**
- Provides a persistent **identifier** such as a DOI or handle
- Comes with a few possibilities for open data **licenses**
- Has high documentation **metadata standards** with controlled vocabularies (therefore discipline-specific is usually better)
- Lets you **keep all your rights**
- Has a certification such as CoreTrustSeal

Note : the editor might suggest that you share the datasets related to a paper as an annex to the paper you are submitting. Not the best move



How to choose a data repository?

Discipline-specific repository

Some examples :
The QDR or Bequali (HSS), CDS (astro), NCBI (genomics), ...
Catalogs of directories : Re3data, FAIRsharing

(But if you have to look it up in a catalog maybe it is not really community-wide?)

All-purpose repository

Zenodo, OpenAire, OSF, Figshare...



How to choose a data repository?

Discipline-specific repository

Some examples :

The QDR or Bequali (HSS), CDS (astro), NCBI (genomics), ...

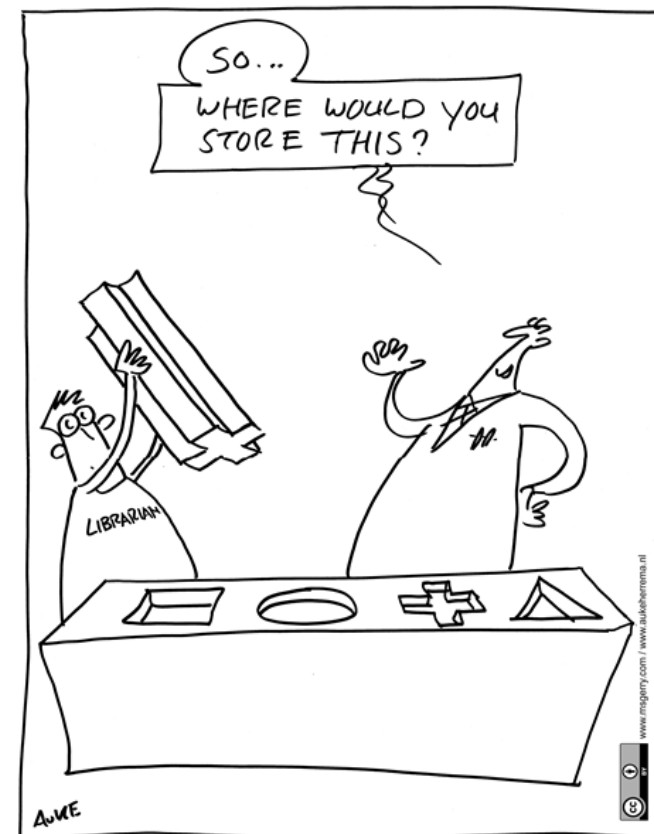
Catalogs of directories : Re3data, FAIRsharing

(But if you have to look it up in a catalog maybe it is not really community-wide?)

All-purpose repository

Zenodo, OpenAire, OSF, Figshare...

Dataverse ULiège

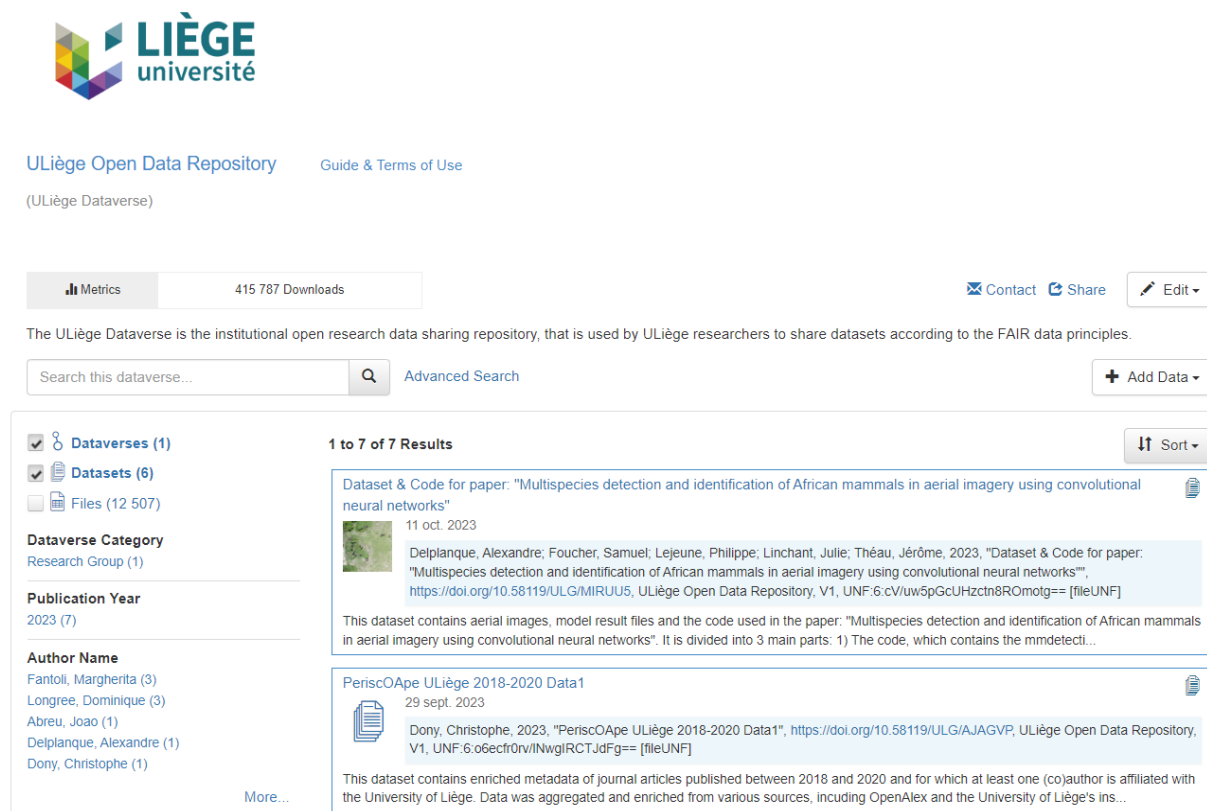


LEARNING HOW TO ARCHIVE DATA

How to choose a data repository?

Dataverse works like “ORBi” but is dedicated to sharing open research data

dataverse.uliege.be - [Demo time](#) !



The screenshot shows the ULiège Open Data Repository interface. At the top, there is the ULiège université logo and navigation links for "ULiège Open Data Repository" and "Guide & Terms of Use". Below this, it indicates "(ULiège Dataverse)" and shows a metrics bar for "415 787 Downloads". There are buttons for "Contact", "Share", and "Edit". A descriptive sentence states: "The ULiège Dataverse is the institutional open research data sharing repository, that is used by ULiège researchers to share datasets according to the FAIR data principles." Below this is a search bar with the text "Search this dataverse...", an "Advanced Search" link, and an "Add Data" button. The main content area is divided into two columns. The left column contains filters: "Dataverses (1)", "Datasets (6)", "Files (12 507)", "Dataverse Category" (Research Group (1)), "Publication Year" (2023 (7)), and "Author Name" (Fantoli, Margherita (3), Longree, Dominique (3), Abreu, Joao (1), Delplanque, Alexandre (1), Dony, Christophe (1)). The right column shows "1 to 7 of 7 Results" with a "Sort" dropdown. Two dataset entries are visible: 1) "Dataset & Code for paper: 'Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks'" dated 11 oct. 2023, by Delplanque, Alexandre; Foucher, Samuel; Lejeune, Philippe; Linchant, Julie; Théau, Jérôme, 2023. 2) "PeriscOApe ULiège 2018-2020 Data1" dated 29 sept. 2023, by Dony, Christophe, 2023.

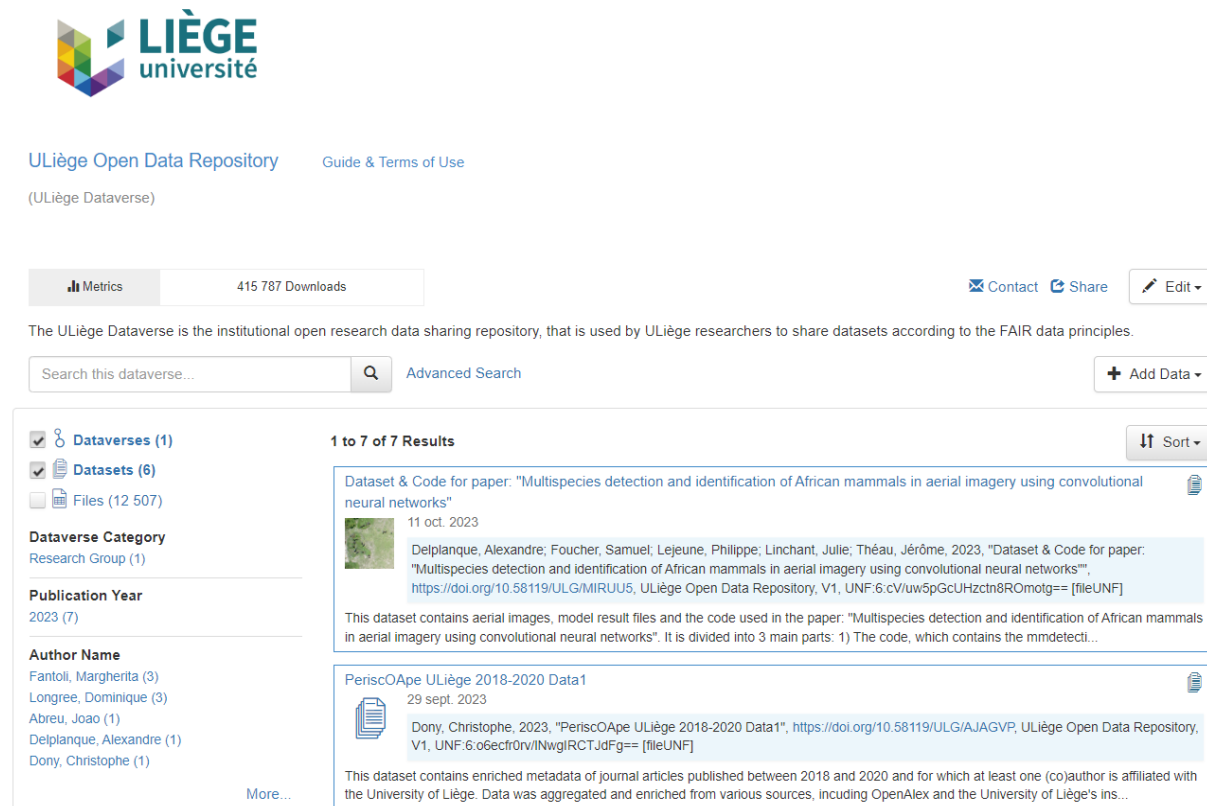
How to choose a data repository?

Dataverse works like “ORBi” but is dedicated to sharing open research data

dataverse.uliege.be - [Demo time](#) !

Key points :

- **Any ULiège researcher** may publish a dataset in the ULiège Dataverse : only an ULiège and ORCID identifiers are needed
- **Checklists and processes** are provided in the User Documentation tab
- The ULiège Dataverse is **operated in-house** by ULiège (RISE, ULL and SEGI)
 - Support is available when needed
 - Before any publication, the admins will proceed to a supporting verification



The screenshot shows the ULiège Open Data Repository website. At the top, there is the LIÈGE université logo and navigation links for "ULiège Open Data Repository" and "Guide & Terms of Use". Below the logo, it says "(ULiège Dataverse)". A metrics bar shows "415 787 Downloads" and options for "Contact", "Share", and "Edit". A search bar is present with the text "Search this dataverse..." and a "Q" icon, along with a link to "Advanced Search" and a "+ Add Data" button. The main content area displays a list of results. On the left, there are filters for "Dataverses (1)", "Datasets (6)", and "Files (12 507)". Below these are sections for "Dataverse Category" (Research Group (1)), "Publication Year" (2023 (7)), and "Author Name" (Fantoli, Margherita (3), Longree, Dominique (3), Abreu, Joao (1), Delplanque, Alexandre (1), Dony, Christophe (1)). The main results area shows "1 to 7 of 7 Results" and a "Sort" dropdown. The first result is titled "Dataset & Code for paper: 'Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks'" and is dated "11 oct. 2023". It includes a small image and a description: "Delplanque, Alexandre; Foucher, Samuel; Lejeune, Philippe; Linchant, Julie; Théau, Jérôme, 2023, 'Dataset & Code for paper: 'Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks'', https://doi.org/10.58119/ULG/MIRUU5, ULiège Open Data Repository, V1, UNF:6:cVuw5pGcUHHzctn8ROm0tg== [fileUNF]". A summary below the title states: "This dataset contains aerial images, model result files and the code used in the paper: 'Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks'. It is divided into 3 main parts: 1) The code, which contains the mmdetecti...". The second result is titled "PeriscOApe ULiège 2018-2020 Data1" and is dated "29 sept. 2023". It includes a document icon and a description: "Dony, Christophe, 2023, 'PeriscOApe ULiège 2018-2020 Data1', https://doi.org/10.58119/ULG/AJAGVP, ULiège Open Data Repository, V1, UNF:6:06ectf0rvlINwglRctJdFg== [fileUNF]". A summary below the title states: "This dataset contains enriched metadata of journal articles published between 2018 and 2020 and for which at least one (co)author is affiliated with the University of Liège. Data was aggregated and enriched from various sources, including OpenAlex and the University of Liège's ins...".

How to choose a data repository?

Dataverse works like “ORBi” but is dedicated to sharing open research data

dataverse.uliege.be

How to use Dataverse:

Follow the procedures 😊 - [Demo time](#) !

Create your account using your ORCiD

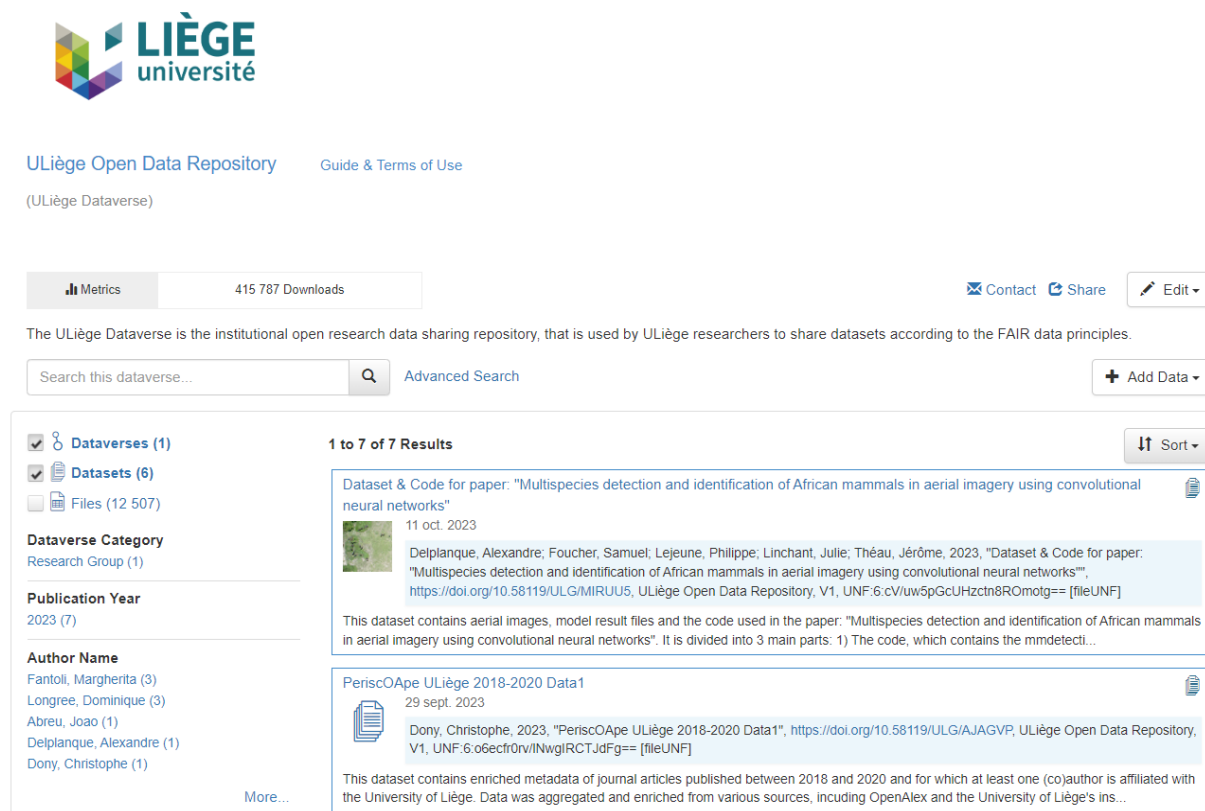
Prepare your files and documentation

Submit your dataset

Your file will go through a reviewing process...

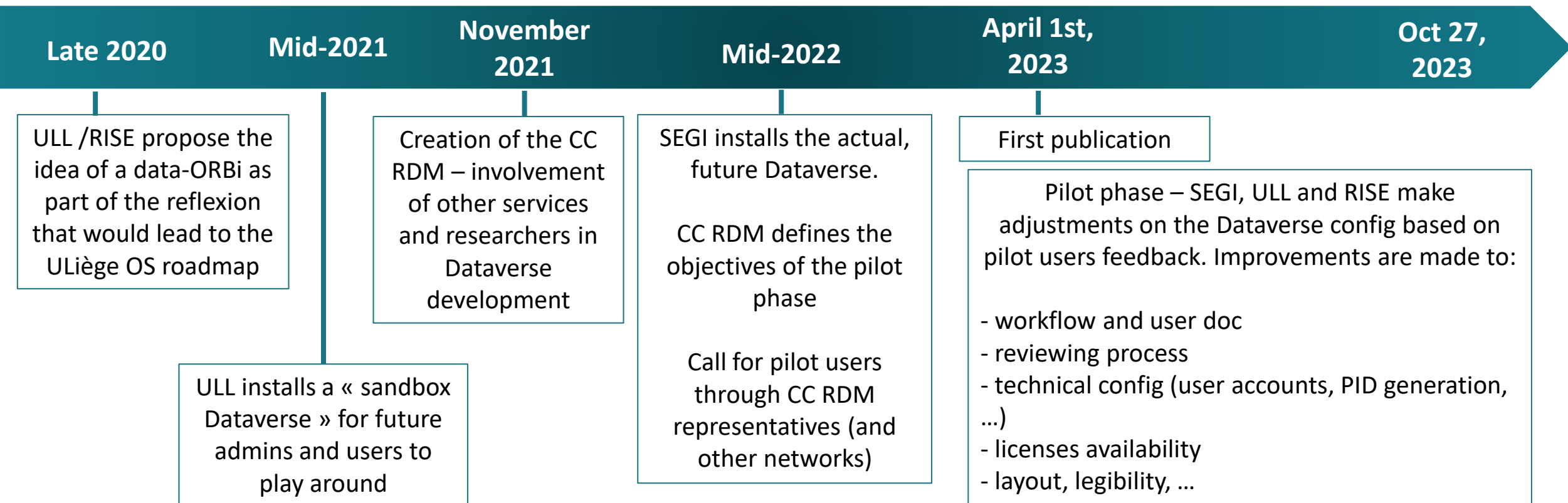
... it either gets published or sent back for fixes

Ask dataverse.admin@uliege.be for any help you need



The screenshot shows the ULIège Open Data Repository website. At the top, there is the LIÈGE université logo. Below it, the text "ULIège Open Data Repository" and "Guide & Terms of Use" are visible. A navigation bar includes "Metrics" (415 787 Downloads), "Contact", "Share", and "Edit". A search bar with "Search this dataverse..." and "Advanced Search" is present, along with an "Add Data" button. The main content area displays a list of datasets. The first result is titled "Dataset & Code for paper: 'Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks'", dated 11 oct. 2023, by Delplanque, Alexandre; Foucher, Samuel; Lejeune, Philippe; Linchant, Julie; Théau, Jérôme. The second result is "PeriscOApe ULIège 2018-2020 Data1", dated 29 sept. 2023, by Dony, Christophe. The left sidebar shows filters for "Dataverses (1)", "Datasets (6)", and "Files (12 507)", along with sections for "Dataverse Category", "Publication Year", and "Author Name".

The story of the Dataverse of madness



The story of the Dataverse of madness

Are we there yet ?

The end of the pilot phase does not mean the end of the story :
there is still plenty left to do

- Users are still more than welcome to provide feedback
- The platform, user guides, ... can be refined as we discover and understand the need of more users (detailed list of metadata examples, dedicated licences documentation, better navigation, ...)
- Some milestones remain on the roadmap :
 - ULiège web pages (ORBi is the inspo 😊)
 - Harvesting
 - Certification

There is always room for improvement



The Dataverse team

The Dataverse team is composed of the following superheroes

Paul Thirion
Alessandro Infantino
Jérôme Eeckhout
François Paquot
Laurent Debra



Image : Marvel

The Dataverse team is thankful to the pilot users

Joao Abreu
Matthieu Verstraete
Dominique Longrée
Margherita Fantoli
Alexandre Delplanque
Philippe Lejeune
Christophe Dony

Christophe Phillips
Christophe Geuzaine
Angeliki Konstantinidou
Daniela Vintila
Christian Degueldre
Jean-Michel Lafleur
Sébastien Hendrickx
Martin Grignard
Alice Mayer