



# Digital Object, FAIR principles for research products, Scientific Data Repositories

Massimiliano Assante, PhD

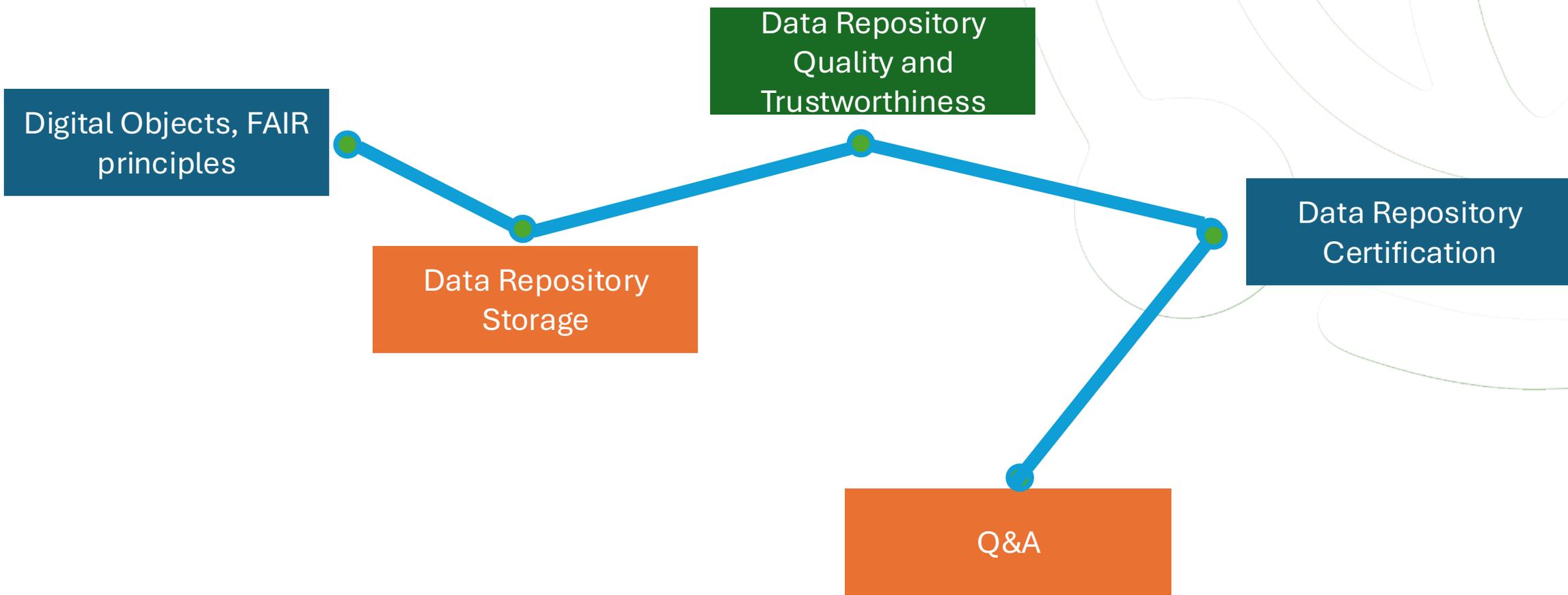
Institute of Information Science and Technologies (ISTI) - CNR

 <https://orcid.org/0000-0002-3761-1492>

**IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System**  
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-  
Mission 4 “Education and Research” - Component 2: “From research to business” - Investment  
3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”



# Today's Journey



## Digital Objects, what are they?

A **Digital Object** is any research result in its digital form that can be stored (and possibly shared).

**Examples:** articles, dataset, software, images, videos, reports, conference poster or presentations, lectures, etc

# FAIR principles in a nutshell



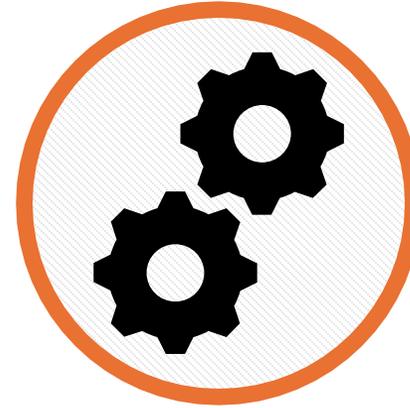
## Findable

Other can find your data



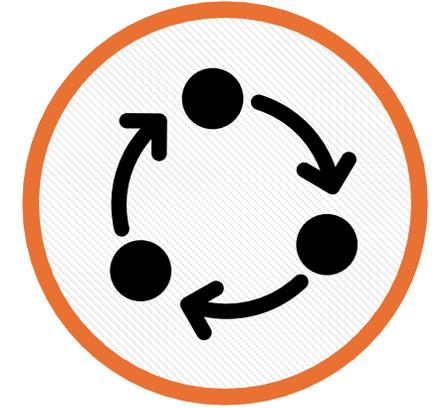
## Accessible

Your data is accessible to others



## Interoperable

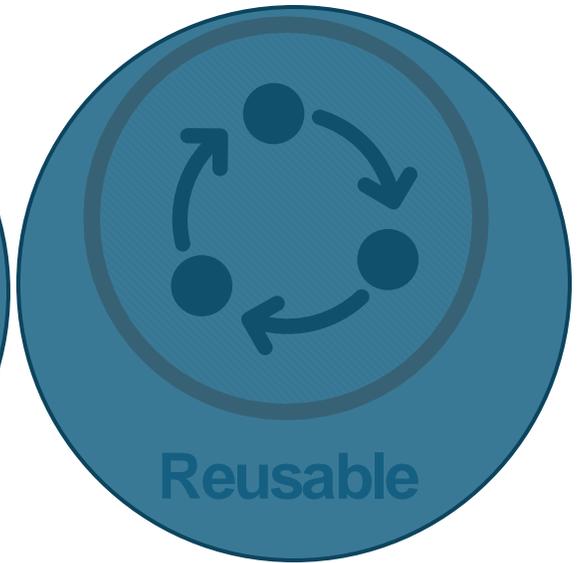
Your data can be integrated with other data and/or they can be easily used and read by machines.



## Reusable

Your data can be reused by others in new research

# Accessible



Your data is  
accessible to  
others

# Accessible

🌐 Can I make my data accessible to others?

🌐 Who will be granted access?

🌐 How?

How do you give access to your data?

Through a Data Repository

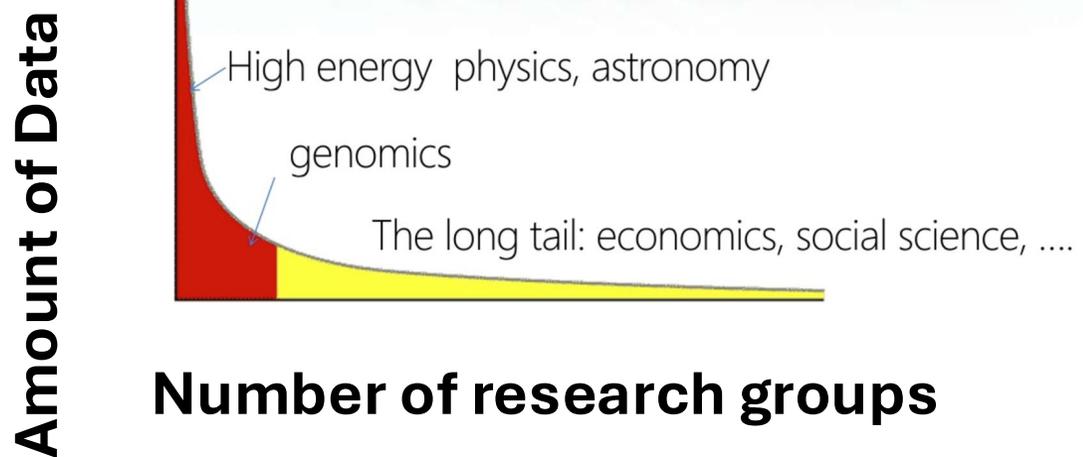
# Data Repositories

Defined as “a subtype of a sustainable information infrastructure which provides long-term storage and access to research data” ([re3data.org](https://re3data.org)).

Provide sustainable preservation to the data created by researchers.

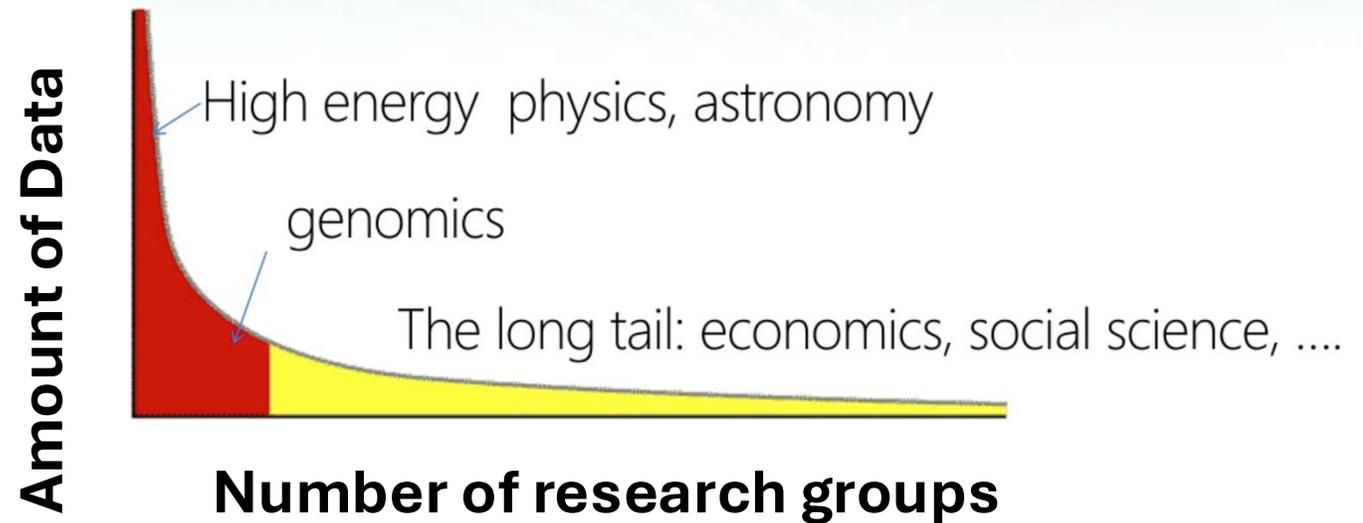
Ensure research data is accessible beyond the life of a grant, research project, or individual careers.

- 🌐 Largely diffused within communities that produce a huge amount of data, such as physics (e.g., LHC at CERN), genetics (e.g., GenBank at NCBI Data), or environmental sciences (e.g., British Atmospheric Data Centre).
- 🌐 However, the demand for data repositories is emerging also within the so-called “*long-tail*” of science.



# Data Repository, scientific communities: long-tail

🌐 However, the demand for data repositories is emerging also within the so-called “*long-tail*” of science.



# Data Repositories Storage: Records deposited

## Metadata

A set of data describing the digital object(s) you are depositing

The screenshot shows a Zenodo record page. The title is 'MOD01 - Research Data Management & Open Science: Introduction and Motivations - including EO policies and mandates'. The author is 'Lazzeri, Emma'. The record is dated July 6, 2020. A table of files is shown with one file: 'INSr\_Lazzeri\_MOD1\_20200706.pptx' (52.9 MB). The record is indexed in OpenAIRE. The DOI is 10.5281/zenodo.3939213. The keywords are 'Open Science', 'research data management', and 'open access'. The grants listed are 'European Commission - OpenAIRE-Advancing Open Scholarship (777541)', 'EDSCsecretariat.eu - EOSCsecretariat.eu', and 'EOSC-Pillar - Coordination and Harmonisation of National Initiatives, Infrastructures and Data services in Central and Western Europe (857650)'. The license is Creative Commons Attribution 4.0 International.

Name	Size
INSr_Lazzeri_MOD1_20200706.pptx	52.9 MB

## Payload

The digital object(s) you are uploading to be stored. Payload includes attached files such as the file containing the data and the accompanying material(s), readme file, etc

## Metadata

- Generally stored on Relational Database Management Systems;
- Daily or hourly backup cycles;
- Some kind of indexing and search feature is generally supported, e.g. Lucene, ElasticSearch, Solr etc.

## Payload

- Dedicated services offering petabytes disk clusters (Data storage services);
- Digital object replicas located on different disk servers;
- MD5 checksums for data integrity;
- Different type of data persistence (e.g. SSD, Magnetic Disk, Tapes) for long-term storage cost optimisation.

# An example of data storage technology

## EOS Open Storage: CERN storage technology used at the LHC

EOS provides a service for storing large amounts of physics data and user files, with a focus on interactive and batch analysis.

### Flexible



EOS is a storage solution for central data recording, analysis and processing++

### Adaptable and Scalable



EOS supports thousands of clients with random remote I/O patterns with multi protocol support  
WebDAV, CIFS, FUSE, XRootd, GRPC.

### Over 350 PB at CERN



Designed for high capacity and low latency.



### Security

EOS offers a variety of authentication methods: KRB5, X509, OIDC, shared secret, and JWT and proprietary token authorisation.



### Sync & Share

EOS provides Sync&Share functionality for the **CERNBox** front-end services.



### Tape Storage

EOS includes tape storage in combination with the **CTA** Cern Tape Archive software.

## Which repository to use?

1. use a disciplinary repository if there is one;
2. use the institutional repository, if you have one where the data will also be available for the long term;
3. search the global [re3data.org](https://re3data.org) for a suitable repository;
4. Can't find one? use the catch-all repository [Zenodo](https://zenodo.org)

# 3. Where to find a suitable research data repository

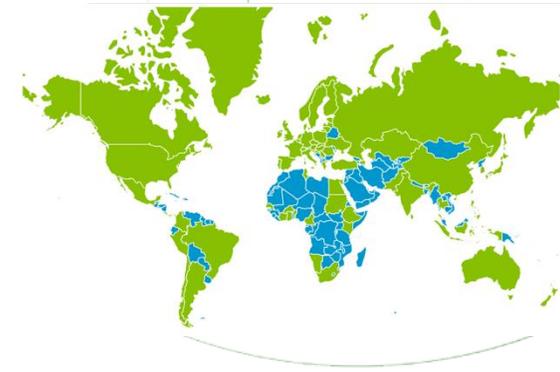
[www.re3data.org](http://www.re3data.org)

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Search **Browse** ▾

Browse by subject  
Browse by content type  
Browse by country



re3data.org +2700 repositories Search Browse ▾ Suggest Resources ▾ Contact

**Filter**

- Subjects ⊕
- Content Types ⊕
- Countries ⊕
- AID systems ⊕
- API ⊕
- Certificates ⊕
- Data access ⊕
- Data access restrictions ⊕
- Database access ⊕
- Database access restrictions ⊕
- Database licenses ⊕

social sciences and humanities  [Toggle short help](#)

← Previous **1** 2 3 4 5 6 7 ... 35 Next →

Found 857 result(s)

**TESS**  
Time-sharing Experiments for the Social Sciences

Subject(s) Social Sciences Social and Behavioural Sciences Humanities and Social Sciences

Content type(s) Standard office documents Plain text

# Can't find one? Try Zenodo



<https://zenodo.org>

Upload type required ▼

Publication  Poster  Presentation  Dataset  Image  Video/Audio  Software  Lesson  Other

Basic information required ▼

**Digital Object Identifier**

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register it for you so others can easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it is assigned. It is always possible to edit a custom DOI.

Reserve DOI

**Publication date \***

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

**Title \***

Required.

**Authors \***    ORCID (e.g.: )

Optional.

License required ▼

**Access right \***  Open Access  Embargoed Access  Restricted Access  Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

**License \***

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the other licenses available (Other (Open), Other (Attribution), etc.). The supported licenses in the list are harvested from [opendefinition.org](https://opendefinition.org) and [spdx.org](https://spdx.org). If you think that a license is missing from the list, please [contact us](#).

Related/alternate identifiers

Specify identifiers of related publications and datasets. Supported identifiers include: DOI, Handle, ARK, PURL, ISSN, ISBN, PubMed, arXiv, Life Science Identifiers (LSID), EAN-13, ISTC, URNs and URLs.

**Related identifiers**

[+ Add another related identifier](#)

Contributors

References

Journal

Conference

Book/Report/Chapter

Thesis

- cites this upload
- is cited by this upload
- is supplemented by this upload
- is a supplement to this upload
- is referenced by this upload
- references this upload
- is previous version of this upload
- is new version of this upload
- continues this upload
- is continued by this upload
- has this upload as part
- is part of this upload
- reviews this upload
- is reviewed by this upload
- documents this upload
- is documented by this upload
- is compiled/created by this upload
- compiled/created this upload
- is the source this upload is derived from
- has this upload as its source
- is identical to this upload
- is an alternate identifier of this upload

# Quality of Data Repositories

 Not easy to evaluate;

 Influenced by many external factors, starting with the mission of the repository:

- does it explicitly aim for long-term preservation - with the appropriate expertise and budget - or not?
- Is it dedicated to a specific research community and familiar with their data formats, or is it generic?

A Data Repository of quality should adhere to majority of the FAIR principles, it should at least support:

***Metadata Standards;***

***Persistent and Global Unique Identifiers;***

***Access rights and Licenses;***

***Data citation;***

***Be possibly Certified. (if certified the quality assessment has already been done for you)***

# Data Repository quality

Support *Metadata Standards*;

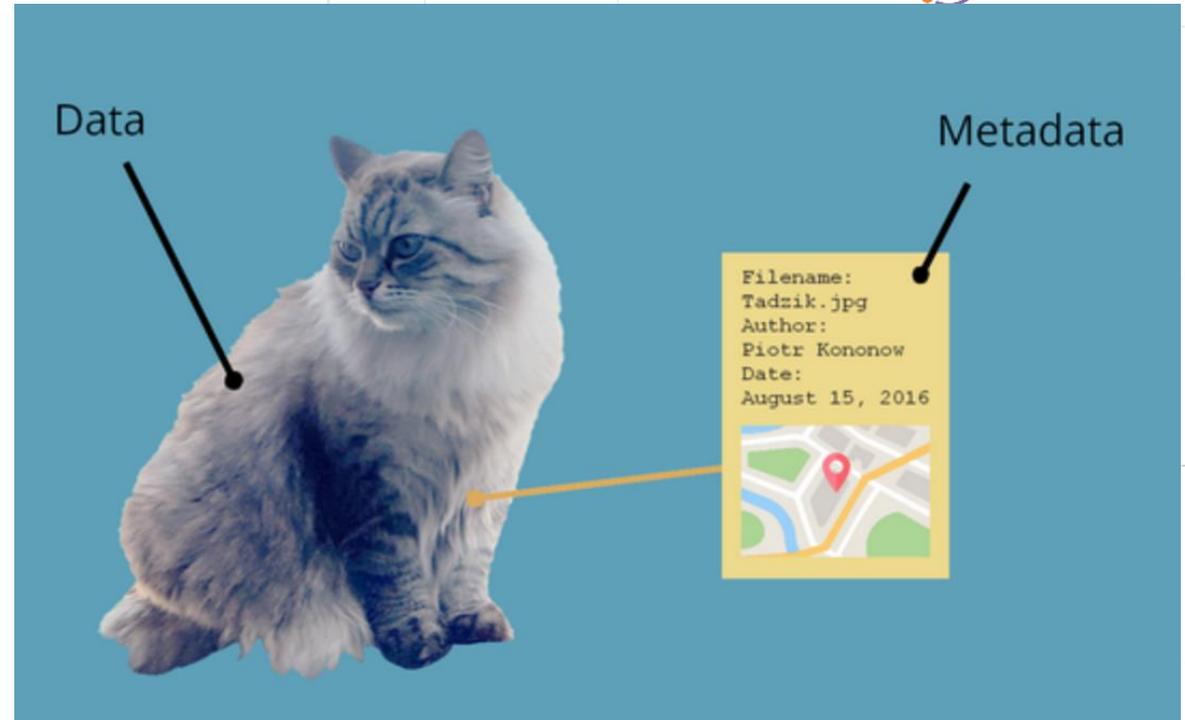
# Metadata

🌐 Data describing data

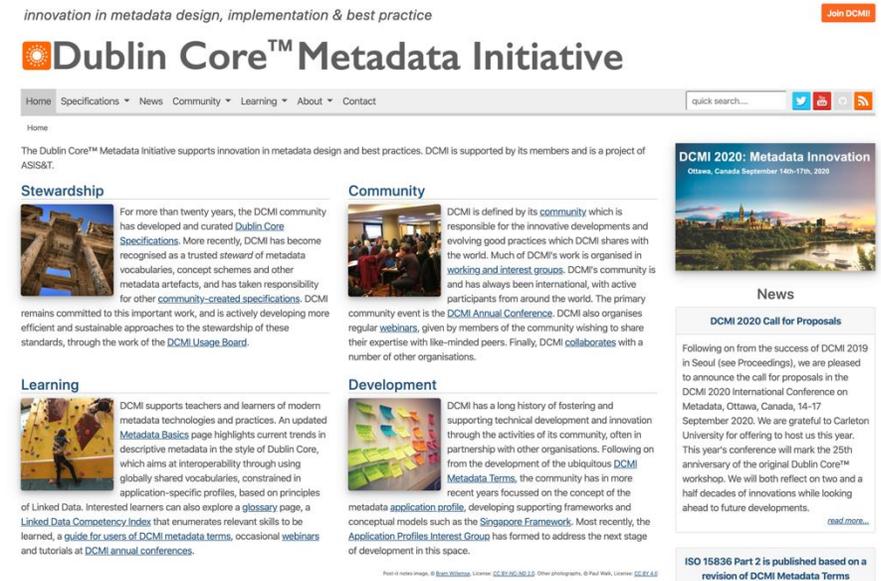
🌐 Very important for:

- Access
- Comprehension
- Process

🌐 Use your discipline specific standards: you will spend less time interpreting data and more time to actually make science!



- 🌐 **The Dublin Core**, a set of fifteen generic, widely used elements:
- 🌐 *Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Right.*
- 🌐 It was first drafted at a 1995 meeting in Dublin, Ohio (USA).



The screenshot shows the homepage of the Dublin Core™ Metadata Initiative. At the top, it says "innovation in metadata design, implementation & best practice" and "Join DCMII". The main heading is "Dublin Core™ Metadata Initiative". Below this is a navigation menu with links for Home, Specifications, News, Community, Learning, About, and Contact. A search bar is also present. The main content area is divided into several sections: "Stewardship" (describing the DCMII community's role), "Community" (describing the DCMII community's activities), "Learning" (describing resources for learning), and "Development" (describing the DCMII's role in developing standards). There is also a "News" section with a "DCMI 2020: Metadata Innovation" announcement. At the bottom, there is a footer with copyright information and a link to "ISO 15836 Part 2".

# Where to find Metadata?

## Metadata Directory:

<http://rd-alliance.github.io/metadata-directory/>



### Metadata

RDA | Metadata Directory

- View the standards
- View the extensions
- View the tools
- View the use cases
- Browse by subject areas

Contribute

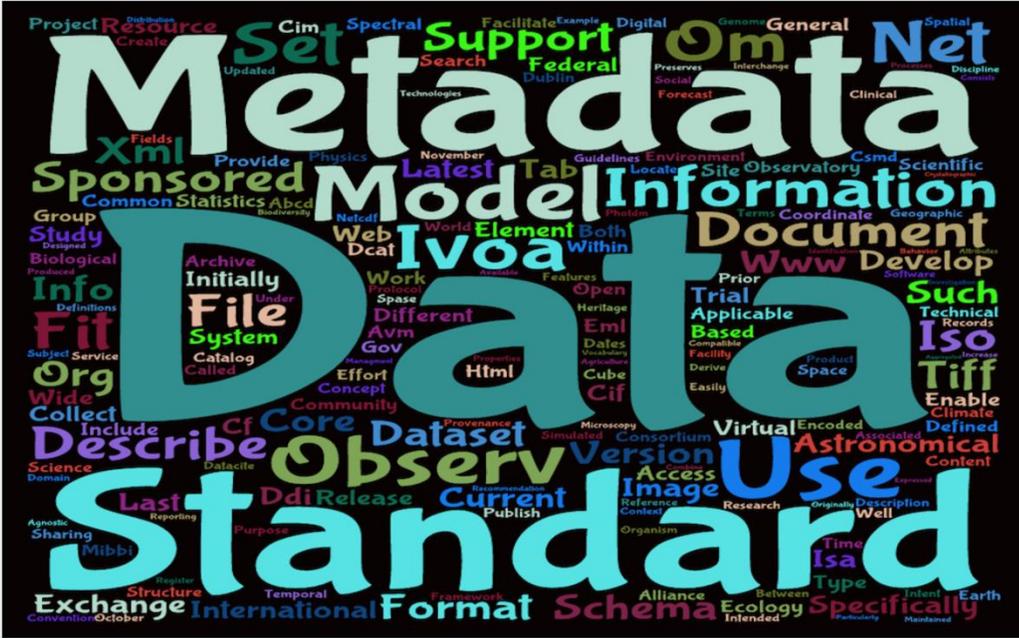
- Add standards
- Add extensions
- Add tools
- Add use cases

github

@twitter

linkedin

facebook



Metadata Standards Directory Working Group

# Where to find Metadata?

 Metadata Standards Catalogue: <https://rdamsc.bath.ac.uk>

- Contains metadata standards that may be used to document research data

Metadata Standards Catalog

Search Sign in

## Metadata Standards Catalog

The RDA Metadata Standards Catalog is a collaborative, open directory of metadata standards applicable to research data. It is offered to the international academic community to help address infrastructure challenges.

- [Read more details about the scope of the Catalog](#)
- [Read terms of use](#)
- [Read accessibility statement](#)



Metadata standards, profiles and schemes

# Data Repository quality



***Provide Persistent and Global Unique Identifiers;***

Persistent identifiers make your data

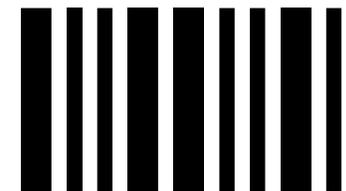


 Findable

 Accessible

# Persistent Identifiers

- 🌐 A persistent identifier (PI or PID) is a **long-lasting reference** to a document, file, web page, or other object.
- 🌐 The term persistent identifier is usually used in the context of **Digital Objects** that are accessible over the Internet.
- 🌐 Typically, such an identifier is **not only persistent but actionable**: you can plug it into a web browser and be taken to the identified source.
- 🌐 It is like the bar code used on products ...



## Persistent Identifiers: Digital Object Identifier (DOI)



- 🌐 In computing, a digital object identifier (DOI) is a persistent identifier or handle used to identify objects uniquely, standardized by the International Organization for Standardization (ISO);
- 🌐 A DOI aims to be resolvable, usually to some form of access to the information object to which the DOI refers;
- 🌐 This is achieved by binding the DOI to metadata about the object, such as a URL, indicating where the object can be found;
- 🌐 a DOI differs from identifiers such as ISBNs which aims only to identify their referents uniquely;

# DOI: an example

## Are Scientific Data Repositories Coping with Research Data Publishing

Authors: Assante M. Candela L. see more

Published: Apr 2016 in Data Science Journal

DOI: 10.5334/DSJ-2016-006

## Repositories for Open Science: The SciRepo Reference Model

Authors: Assante, Massimiliano; Candela, Leonardo; Castelli, Donatella; ... Pagano, Paolo

Published: Nov 2015 in Communications in Computer and Information Science

DOI: 10.1007/978-3-319-229-6\_26

<https://doi.org/10.5334/DSJ-2016-006>

ScienceDirect Journals & Books Help Search

View PDF Download full issue

ELSEVIER Future Generation Computer Systems Volume 101, December 2019, Pages 555-563

### Enacting open science by D4Science

M. Assante<sup>a</sup>, L. Candela<sup>a</sup>, D. Castelli<sup>a</sup>, R. Cirillo<sup>a</sup>, G. Coro<sup>a</sup>, L. Frosini<sup>a,b</sup>, L. Lelii<sup>a</sup>, F. Mangiacrapa<sup>a</sup>, P. Pagano<sup>a</sup>, G. Panichi<sup>a</sup>, F. Sinibaldi<sup>a</sup>

Show more

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.future.2019.05.063> Get rights and content

Outline Highlights Abstract Keywords

1. Introduction
2. Related works
3. D4Science open science services framework for
4. Discussion
5. Conclusion and future work

CRediT authorship contribution statement Acknowledgements

Home About Contact Content Research Integrity Search Log In Register

## DATA SCIENCE JOURNAL

Start Submission Become a Reviewer

Reading: Are Scientific Data Repositories Coping with Research Data Publishing? Share: f t g in Download A- A+

JUMP TO RELATED CONTENT

- Abstract
- 1 Introduction
- 2 Repository Selection
- 3 Published Datasets: an Overview
- 4 Analysis
- 5 Conclusion and Prospect
- Competing Interests
- Notes
- Acknowledgments
- Author contributions

### Reviews

#### Are Scientific Data Repositories Coping with Research Data Publishing?

**Authors:** Massimiliano Assante, Leonardo Candela, Donatella Castelli, Alice Tani

#### Abstract

Research data publishing is intended as the release of research data to make it possible for practitioners to (re)use them according to "open science" dynamics. There are three main actors called to deal with research data publishing practices: researchers, publishers, and data repositories. This study analyses the solutions offered by generalist scientific data repositories, i.e., repositories supporting the deposition of any type of research data. These repositories cannot make any assumption on the application domain. They are actually

## How can you assign a Persistent Identifier to your digital object?

- 🌐 Persistent identifiers need to be assigned by an entity that can ensure the persistency of the link to the object;
- 🌐 Zenodo assigns DOIs to digital objects that do not already have one

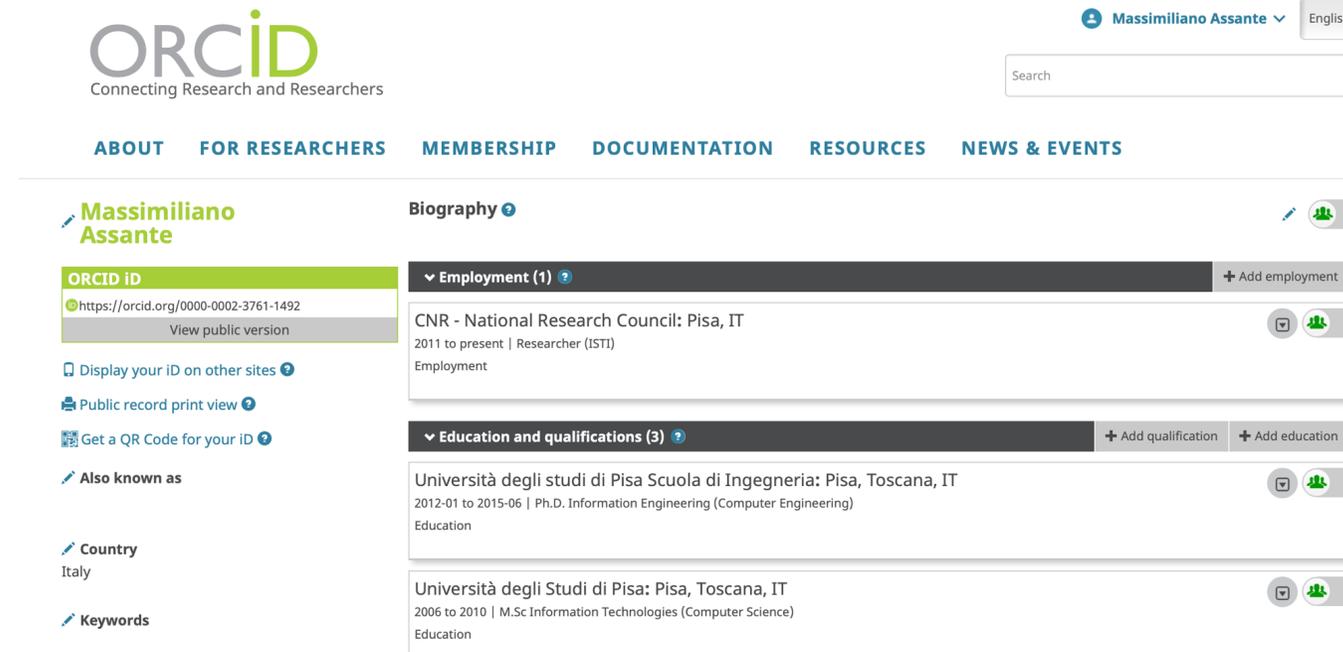
### DOI Registration Agencies



# ORCID : Open Researcher and Contributor ID

It is a non-proprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. e.g.: <https://orcid.org/0000-0002-3761-1492>

Do you have one?



The screenshot shows the ORCID iD profile for Massimiliano Assante. The profile includes the following information:

- ORCID ID:** <https://orcid.org/0000-0002-3761-1492>
- Country:** Italy
- Biography:**
  - Employment (1):**
    - CNR - National Research Council: Pisa, IT (2011 to present) | Researcher (ISTI)
  - Education and qualifications (3):**
    - Università degli studi di Pisa Scuola di Ingegneria: Pisa, Toscana, IT (2012-01 to 2015-06) | Ph.D. Information Engineering (Computer Engineering)
    - Università degli Studi di Pisa: Pisa, Toscana, IT (2006 to 2010) | M.Sc. Information Technologies (Computer Science)

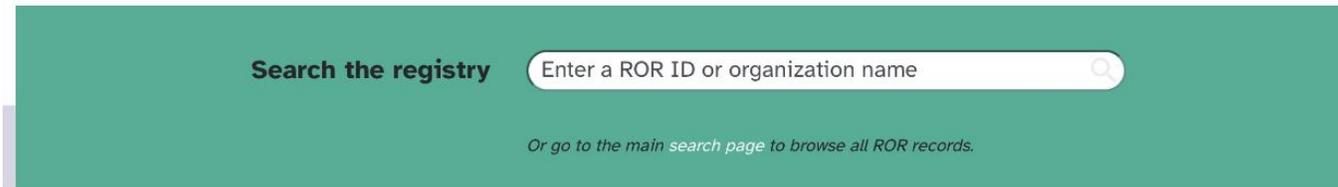
# ROR : Research Organization Registry

For Research organization identifiers (founded Jan 2019)

<https://ror.org>



**ROR is a global, community-led registry of open persistent identifiers for research organizations**



# Data Repository quality: access rights, licenses

Support *access rights, licenses*;

# Deposit vs. Access: an important difference



**Deposit:** upload a digital object (data, articles, ...) on a platform that allows to correctly describe the object through metadata and that implements long-term preservation



**Give access:** once the object has been deposited, the authors can choose the type of access that can be granted (open, restricted, closed, embargoed,...) and assigns a licence to reuse the contents (Creative Commons)



# Access Right: Open Access

**Access right \***

Open Access

Embargoed Access

Restricted Access

Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

**License \***

Creative Commons Attribution 4.0 International

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the *Other* licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from [opendefinition.org](https://opendefinition.org) and [spdx.org](https://spdx.org). If you think that a license is missing from the list, please [contact us](#).

This should be the default access right  
Always assign a licence for reuse

# Access Right: Embargoed Access

Access right \*

- Open Access
- Embargoed Access
- Restricted Access
- Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

📅 Embargo date

Required only for Embargoed Access uploads. Format: YYYY-MM-DD. The date your upload will be made publicly available in case it is under an embargo period from your publisher.

🌀 License \*

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the *Other* licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from [opendefinition.org](https://opendefinition.org) and [spdx.org](https://spdx.org). If you think that a license is missing from the list, please [contact us](#).

# Access Right: Restricted Access

Access right \*

- Open Access
- Embargoed Access
- Restricted Access
- Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Conditions \*



A rich text editor toolbar with various icons for text formatting and editing. The icons include: a lock, a padlock, bold (B), italic (I), strikethrough (ABC), subscript (x<sub>2</sub>), superscript (x<sup>2</sup>), link, unlink, list, ordered list, indent, outdent, quote, code, undo, redo, link, unlink, sum (Σ), omega (Ω), Sorgente, and refresh.

Specify the conditions under which you grant access. Based on the justifications provided on Zenodo.

Use it when you have a **valid reason** to restrict the access  
Always specify conditions under which you grant access (who, how, why can get access to your payload)  
Note: metadata is always accessible to everyone

# Access Right: Closed Access

Access right \*

-  Open Access
-  Embargoed Access
-  Restricted Access
-  Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Are you really sure you need closed access?  
consider closed or embargoed access instead!

# Licenses: Creative Commons

- 🌐 Not all of us are legal experts capable of writing proper licenses;
- 🌐 Creative Commons and Public Domain create legal certainty for everyone, who wants to use works, that are licensed respectively;
- 🌐 It is important to follow and understand the different meanings of the licenses and follow the rules for using them.

CREATIVE COMMONS LICENSES		 COPY & PUBLISH	 ATTRIBUTION REQUIRED	 COMMERCIAL USE	 MODIFY & ADAPT	 CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✓
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✓

 You can redistribute (copy, publish, display, communicate, etc.)

 You have to attribute the original work

 You can use the work commercially

 You can modify and adapt the original work

 You can choose license type for your adaptations of the work.

# Licenses: Creative Commons

- **Public Domain**

Works are not covered by copyright

- **CC-0 (no rights reserved)**

Allows creators to give up their copyright and put their works into the worldwide public domain

- **CC-BY (Attribution)**

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator

- **CC-BY-SA (Attribution – ShareAlike)**

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.

- **CC-BY-ND (Attribution – NonDerivative)**

This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use.

- **CC-BY-NC (Attribution – NonCommercial)**

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.

The chart displays the permissions for various Creative Commons licenses across five categories: Copy & Publish, Attribution Required, Commercial Use, Modify & Adapt, and Change License. Licenses are listed on the left, and their permissions are indicated by checkmarks (allowed) or crosses (not allowed) in colored boxes. A legend at the bottom explains the colors: yellow for redistribution, orange for attribution, red for commercial use, green for modification, and blue for license change.

CREATIVE COMMONS LICENSES		COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
PUBLIC DOMAIN		✓	✗	✓	✓	✓
CC BY		✓	✓	✓	✓	✓
CC BY-SA		✓	✓	✓	✓	✗
CC BY-ND		✓	✓	✓	✗	✓
CC BY-NC		✓	✓	✗	✓	✓
CC BY-NC-SA		✓	✓	✗	✓	✗
CC BY-NC-ND		✓	✓	✗	✗	✓

**Legend:**

- You can redistribute (copy, publish, display, communicate, etc.)
- You have to attribute the original work
- You can use the work commercially
- You can modify and adapt the original work
- You can choose license type for your adaptations of the work.

# Licenses: Creative Commons

## Licenses

CREATIVE COMMONS LICENSES		 COPY & PUBLISH	 ATTRIBUTION REQUIRED	 COMMERCIAL USE	 MODIFY & ADAPT	 CHANGE LICENSE
 PUBLIC DOMAIN		✓	✗	✓	✓	✓
 CC BY		✓	✓	✓	✓	✓
 CC BY-SA		✓	✓	✓	✓	✗
 CC BY-ND		✓	✓	✓	✗	✓
 CC BY-NC		✓	✓	✗	✓	✓
 CC BY-NC-SA		✓	✓	✗	✓	✗
 CC BY-NC-ND		✓	✓	✗	✗	✓

 You can redistribute (copy, publish, display, communicate, etc.)

 You have to attribute the original work

 You can use the work commercially

 You can modify and adapt the original work

 You can choose license type for your adaptations of the work.

# Data Repository quality: Data Citation

*Support Data citation;*

# Data Citation – what is it?

*The provision of a reference to dataset(s) describing data properties enabling credits and attribution, discover, interlinking, and access.*

*To support attribution, reward and encourage the reuse of data for new research questions.*

# Why it is important to cite data

- To support proper *attribution* and *credit*;
- To support *collaboration* and *reuse* of data;
- To enable *reproducibility* of findings;
- To foster *faster* and more *efficient* research progress;
- To provide the means to *share data* with future researchers.

Source:

<https://datacite.org/cite-your-data.html>

# Components of a data citation

very similar to citing publications; many "correct" formats to use. It is suggested including the following information:

- creator(s) or contributor(s);
- title of dataset;
- date of publication;
- (persistent) identifier (e.g. DOI) or URL;
- Publisher;
- version, if relevant;
- Last date accessed, if relevant;

# Data citation examples

Cite

APA 

---

Assante, M. (2015). *FigShare harvested dataset licenses at 31-12-2014* [Data set]. figshare. <https://doi.org/10.6084/M9.FIGSHARE.1311684>

Cite

Harvard 

---

Assante, M. (2015) "FigShare harvested dataset licenses at 31-12-2014." figshare. doi: 10.6084/M9.FIGSHARE.1311684.

Cite

Chicago 

---

Assante, Massimiliano. "FigShare Harvested Dataset Licenses at 31-12-2014." figshare, 2015.  
<https://doi.org/10.6084/M9.FIGSHARE.1311684>.

Locate, identify, and cite research data with the leading global provider of DOIs for research data

<https://search.datacite.org/>

26,240,291 Works

2,302,649 People

2,436 Repositories



Find what you're looking for by searching millions of records with extensive, reliable metadata.



Share your data and reuse the data of others to create the highest impact in the research community.



Cite your research sources with confidence, and receive proper credit when your work is reused.



Connect your research – publications, datasets, software, authors, institutions, and funding data all in one place.

## Create DOI on Datacite Video

<https://www.youtube.com/watch?v=yr6M6x5z3XU>

# Data Repository Certification

## *Data Repository Certification*

***(if certified*** the quality assessment has already been done for you)

# Importance of Certification

- 🌐 **National and international funders** are increasingly likely to mandate open data and data management policies that **call for the long-term storage and accessibility of data.**
- 🌐 If we want to be able **to share data, we need to store them in a trustworthy data repository.** Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them.
- 🌐 Certification can be an important contribution to **ensuring the reliability and durability of data repositories** and hence the potential for sharing data over a long period of time. **By becoming certified, repositories can demonstrate to both their users and their funders** that an independent authority has evaluated them and endorsed their trustworthiness.

# European Framework for audit and certification

3 certification instruments, with increasing degrees of complexity and depth, are available:



CoreTrustSeal (CTS)

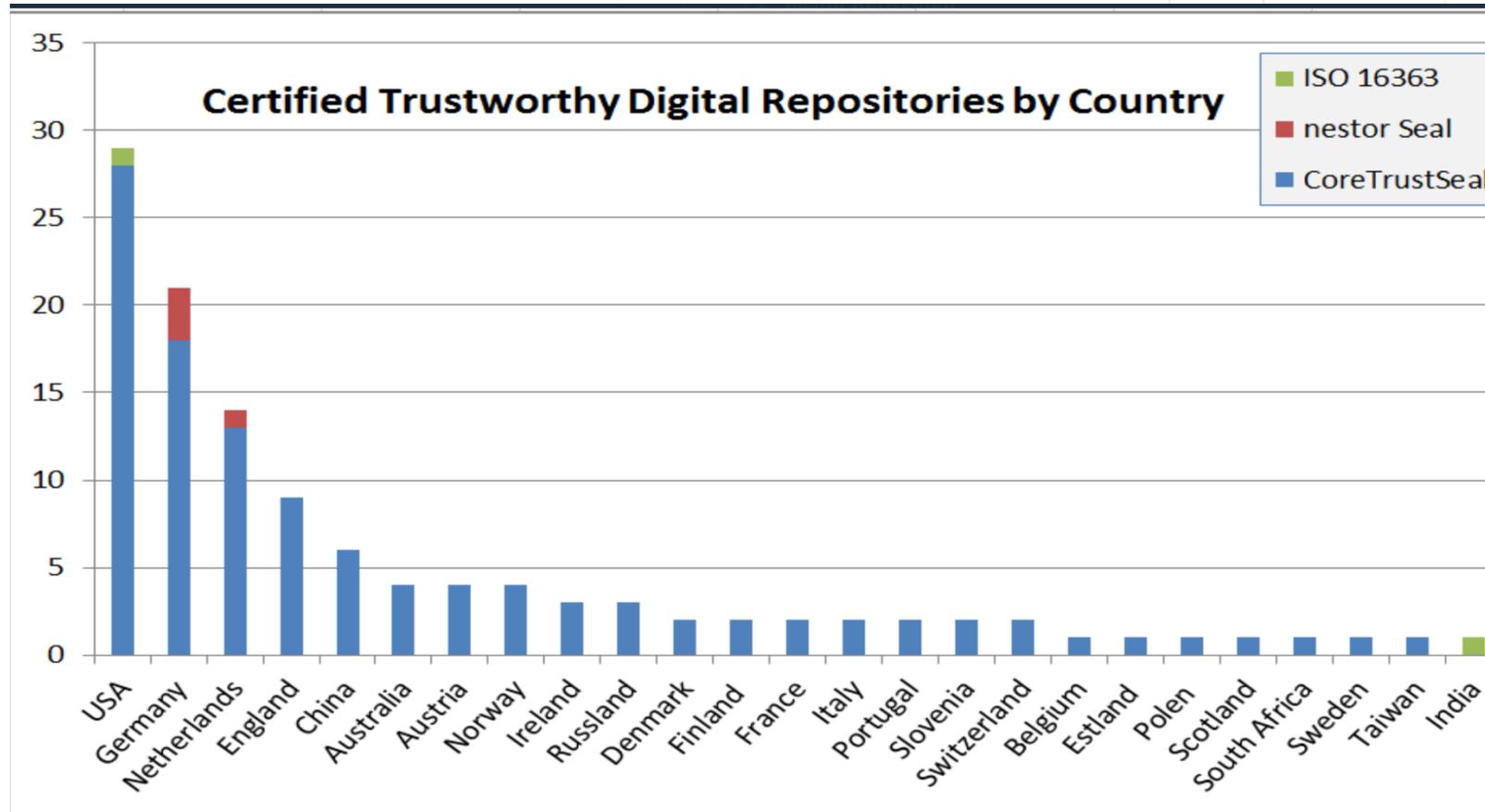


Nestor-Seal



ISO 16363

# Certified Repositories by Country at 2022



*Core Trust Seal (CTS) is the most common used*

# What is CoreTrustSeal (CTS)

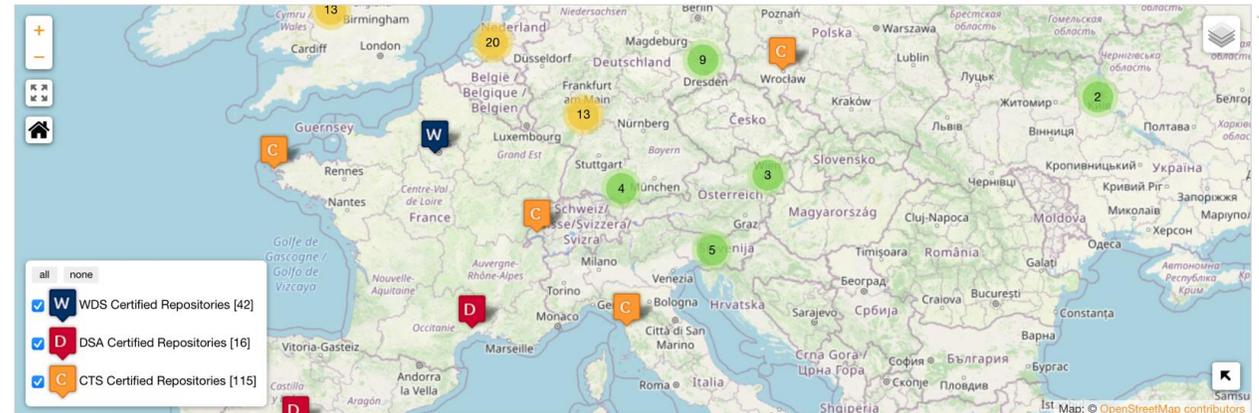


- A certification organisation launched by WDS (World Data System of the International Science Council) and DSA (Data Seal of Approval).
- All repositories having CTS cert. listed at <https://www.coretrustseal.org/why-certification/certified-repositories/>

## Core Certified Repositories

Home > Why certification > Core Certified Repositories

Applications are made public only once certification of a data repository has been approved by the CoreTrustSeal Board. Certification is against the version of the Core Trustworthy Data Repositories Requirements named in the link to the public application (e.g., 2017–2019). The CoreTrustSeal for Data Repositories is valid for three years from the certification date listed within the public application.



# Core Certification and its Benefits

Core certification involves a minimally intensive process whereby data repositories supply evidence that they are sustainable and trustworthy:

- 🌐 A repository first conducts an internal self-assessment, which is then reviewed by community peers.
- 🌐 The repository complies with the 14 requirements from CTS

## CoreTrustSeal (CTS) Requirements (1/2)

Main requirements that a repository with a CTS cert. adheres to (R0 and R1 are about context and mission of the repo):

-  **R2:** The repository maintains all applicable **licenses covering data access** and use and monitors compliance.
-  **R3:** The repository has a continuity plan to ensure **ongoing access** to and preservation of its holdings.
-  **R7:** The repository guarantees the **integrity and authenticity of the data.**
-  **R8:** The repository accepts **data and metadata** based on defined criteria to ***ensure the relevance and understandability for users.***

- 🌐 **R10:** The repository assumes responsibility for the **long-term preservation** and manages this function in a planned and **documented** way.
- 🌐 **R11:** The repository has appropriate expertise to address **technical data and metadata quality** and ensures that **sufficient information** is available for end users to make quality-related evaluations.
- 🌐 **R13:** The repository enables users to **discover the data** and **refer to them in a persistent way** through proper citation.
- 🌐 **R14:** The repository enables reuse of the data over time, ensuring that **appropriate metadata** are available to support the understanding and use of the data.

# More on CoreTrustSeal (CTS) Requirements

The screenshot shows a Zenodo record page for the document 'CoreTrustSeal Requirements 2023-2025'. The page features a blue header with the Zenodo logo, a search bar, and navigation links for 'Communities' and 'My dashboard'. On the right side of the header, there are 'Log in' and 'Sign up' buttons. Below the header, the record is identified as 'Published September 5, 2022 | Version V01.00' and includes 'Other' and 'Open' buttons. The main title is 'CoreTrustSeal Requirements 2023-2025', with a subtitle 'CoreTrustSeal Standards and Certification Board'. A brief description states: 'The CoreTrustSeal Requirements describe the characteristics required to be a trustworthy repository for digital data and metadata. Each Requirement is accompanied by Guidance text describing the response statements and evidence that applicants must provide to enable an objective review. Applicants must respond to all of the Requirements.' Under the 'Files' section, a PDF viewer displays the document 'CoreTrustSeal-Requirements-2023-2025\_v01.00.pdf', showing page 1 of 19. The PDF content is mostly blank, with a large 'CORE TRUST SEAL' logo at the bottom. The logo consists of the words 'CORE TRUST SEAL' in a bold, sans-serif font, with 'CORE' in black, 'TRUST' in orange, and 'SEAL' in black. A large orange checkmark is positioned to the right of the text, partially overlapping the word 'SEAL'. The right sidebar contains several sections: 'Views' (17K) and 'Downloads' (9K) with a 'Show more details' link; 'Versions' (Version V01.00, Sep 5, 2022, DOI 10.5281/zenodo.7051012) with a 'Cite all versions?' section; 'External resources' (Indexed in OpenAIRE); 'Communities' (CoreTrustSeal Current Requirements, CoreTrustSeal, Digital Preservation); and 'Keywords and subjects' (Certification, Trustworthy Data Repositories, CoreTrustSeal).

zenodo Search records... Communities My dashboard Log in Sign up

Published September 5, 2022 | Version V01.00 Other Open

## CoreTrustSeal Requirements 2023-2025

CoreTrustSeal Standards and Certification Board

The CoreTrustSeal Requirements describe the characteristics required to be a trustworthy repository for digital data and metadata. Each Requirement is accompanied by Guidance text describing the response statements and evidence that applicants must provide to enable an objective review. Applicants must respond to all of the Requirements.

### Files

CoreTrustSeal-Requirements-2023-2025\_v01.00.pdf

Page: 1 of 19 Automatic Zoom?

**CORE TRUST SEAL**

17K VIEWS 9K DOWNLOADS Show more details

#### Versions

Version V01.00 Sep 5, 2022  
10.5281/zenodo.7051012

**Cite all versions?** You can cite all versions by using the DOI 10.5281/zenodo.7051011. This DOI represents all versions, and will always resolve to the latest one. Read more.

#### External resources

Indexed in

OpenAIRE

#### Communities

- CoreTrustSeal Current Requirements
- CoreTrustSeal
- Digital Preservation: publications, presentations, instructions, data sets

#### Keywords and subjects

Certification Trustworthy Data Repositories CoreTrustSeal

# Repositories FAIR Certification is ongoing

WP4 will support the co-development and implementation of certification schemes for trusted data repositories enabling FAIR research data in **the EOSC**.





Thanks!

**IR000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System**  
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-  
Mission 4 “Education and Research” - Component 2: “From research to business” - Investment  
3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”









# THANKS!

**IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System**  
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-  
Mission 4 "Education and Research" - Component 2: "From research to business" - Investment  
3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

