



Management of identifiers and metadata

Mario Locati

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”



- Introduction to identifiers and metadata
- Overview of identifiers and metadata standards
- Details of each standard

Why persistent identifiers are important

Precisely identifying elements in shared scientific outputs using persistent identifiers is crucial.

It ensure:

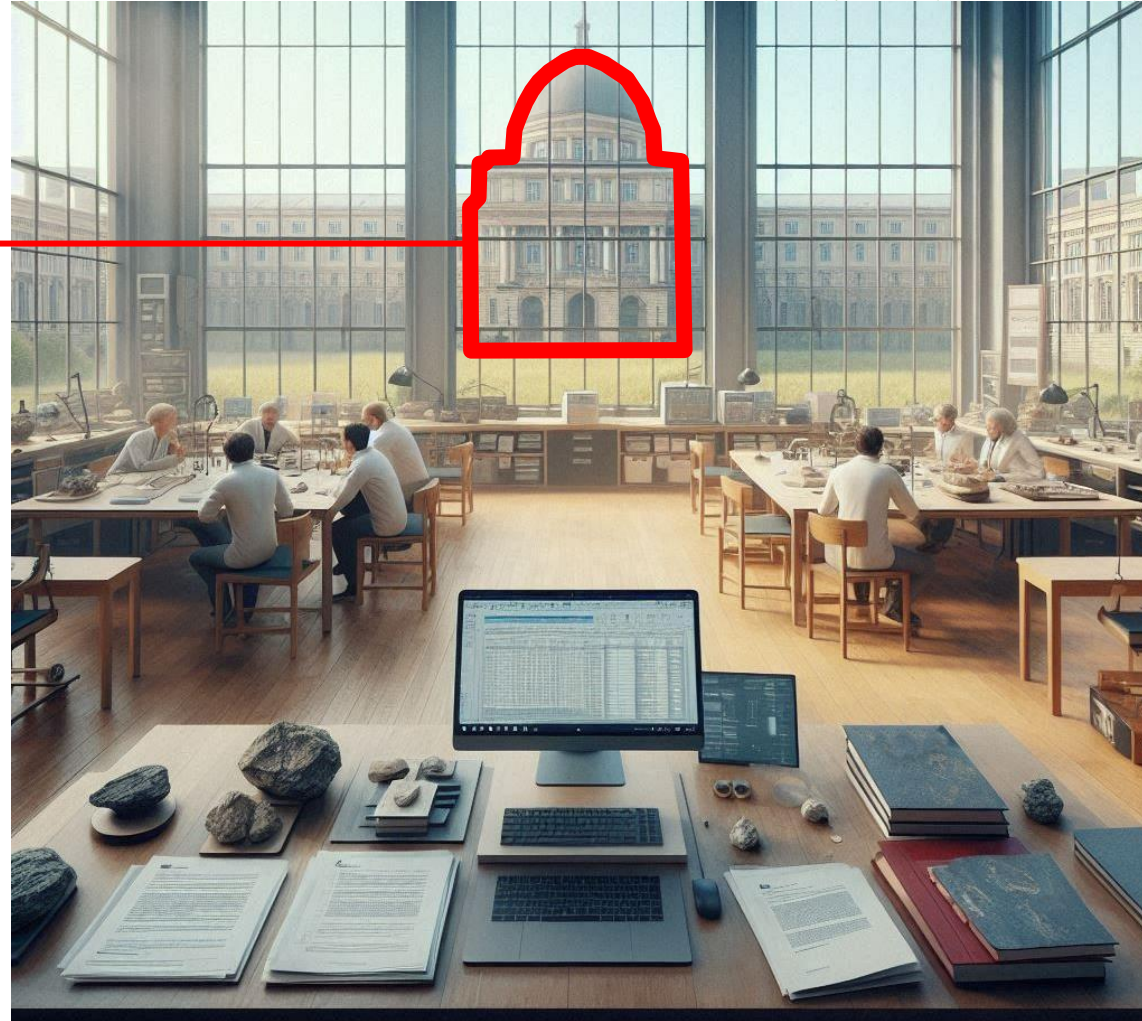
- unambiguous referencing of elements;
- consistency among all elements;
- higher reliability;
- a more efficient data sharing;
- better reproducibility;
- long-term accessibility.

Goal: identify and describe



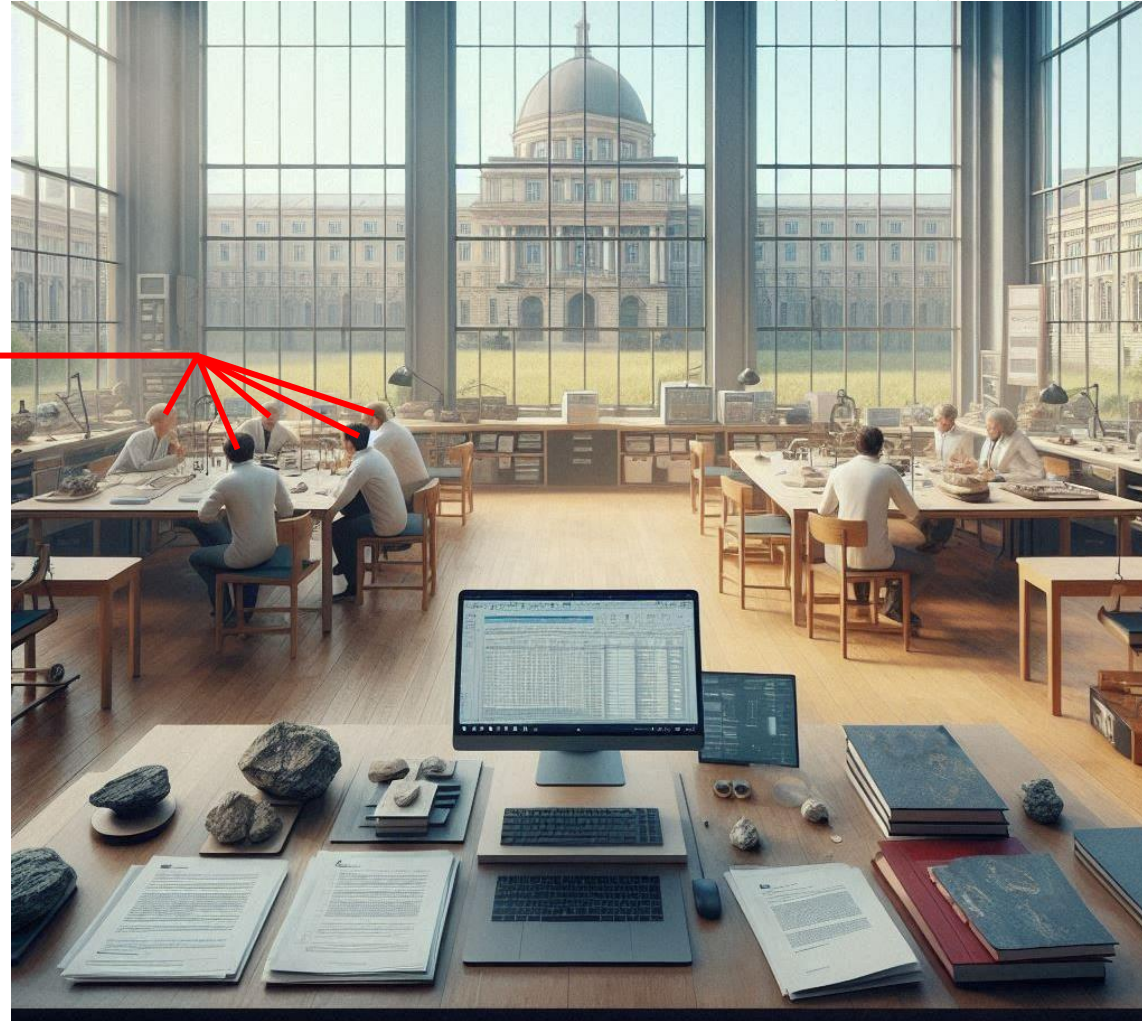
Goal: identify and describe

ROR

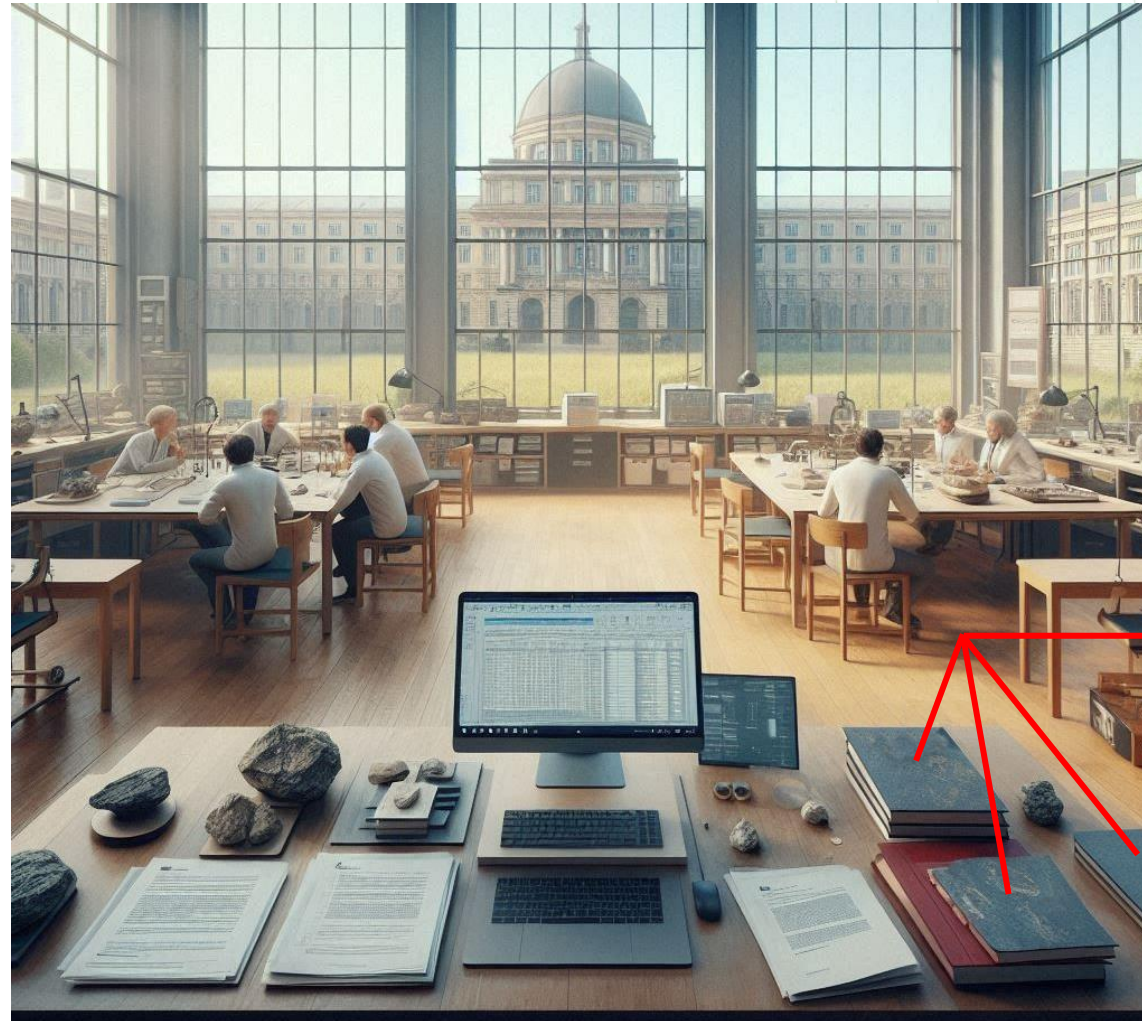


Goal: identify and describe

ORCID



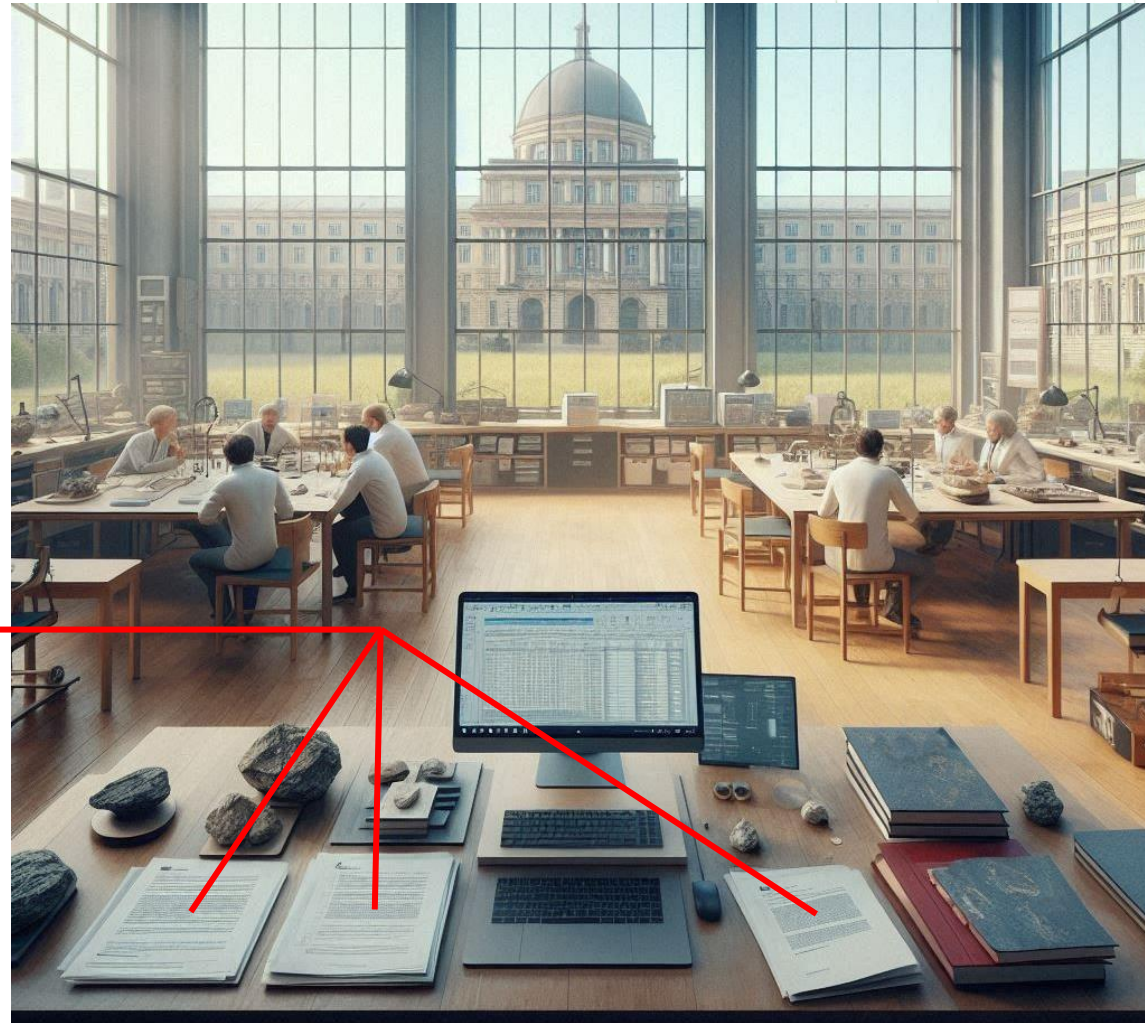
Goal: identify and describe



ISBN

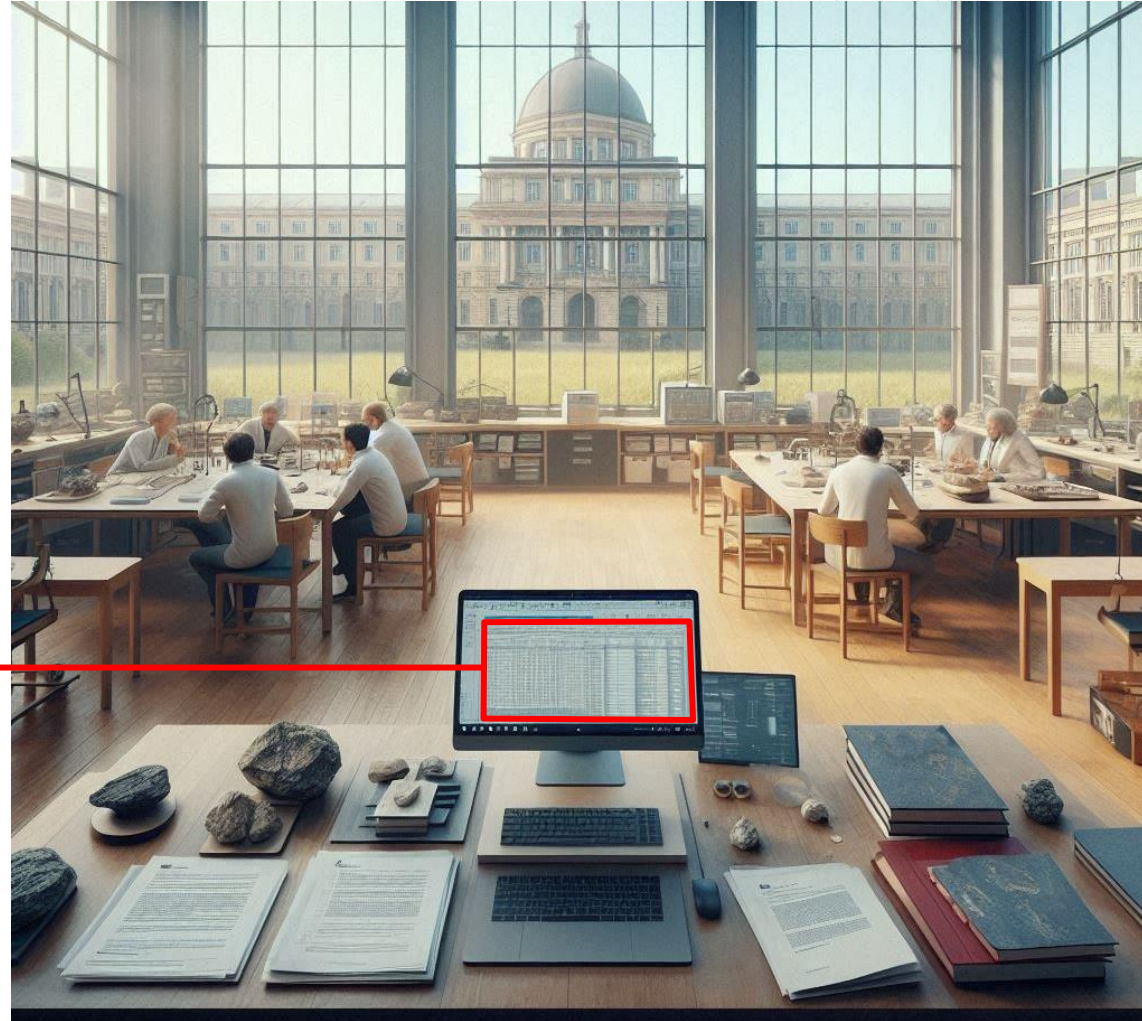
Goal: identify and describe

DOI



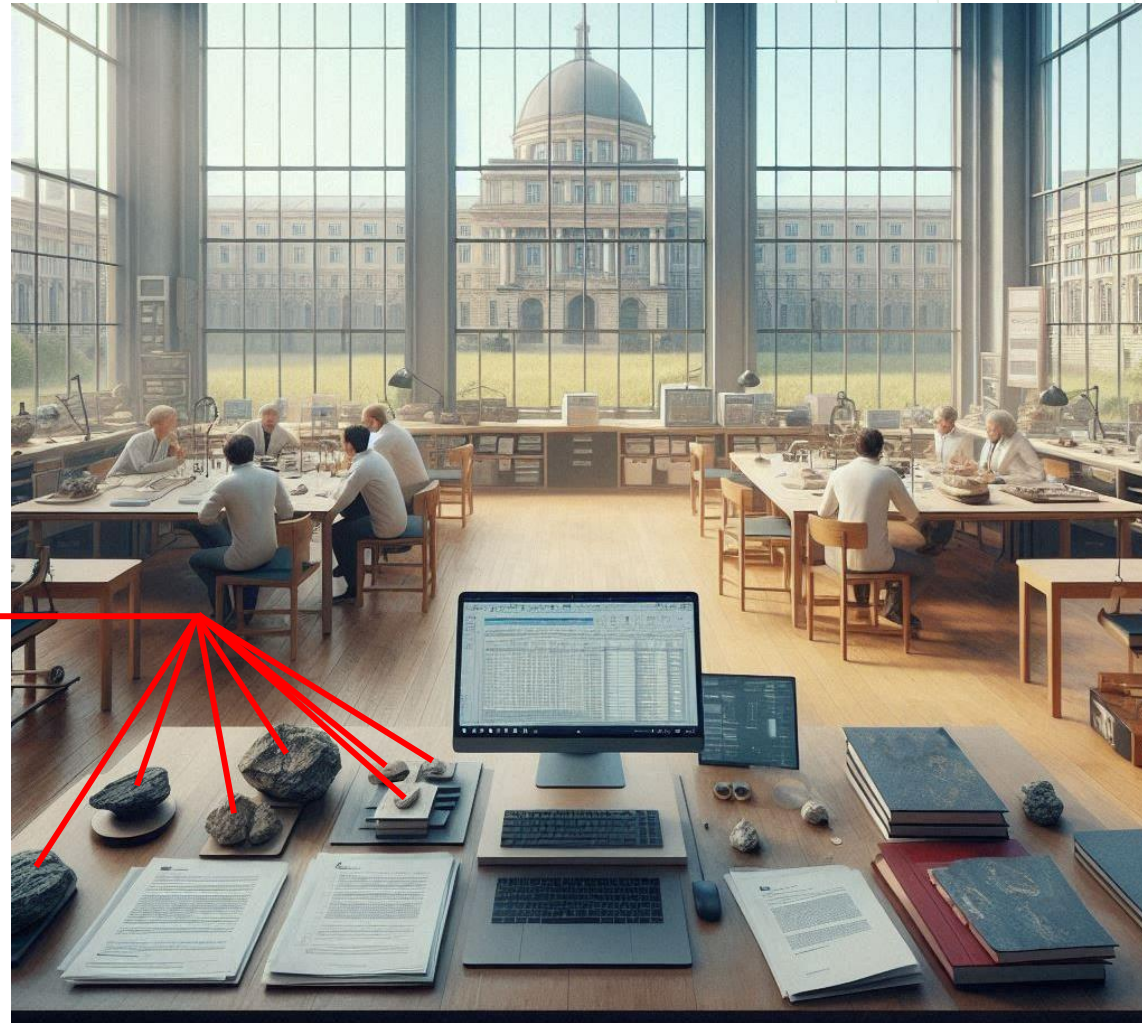
Goal: identify and describe

DOI

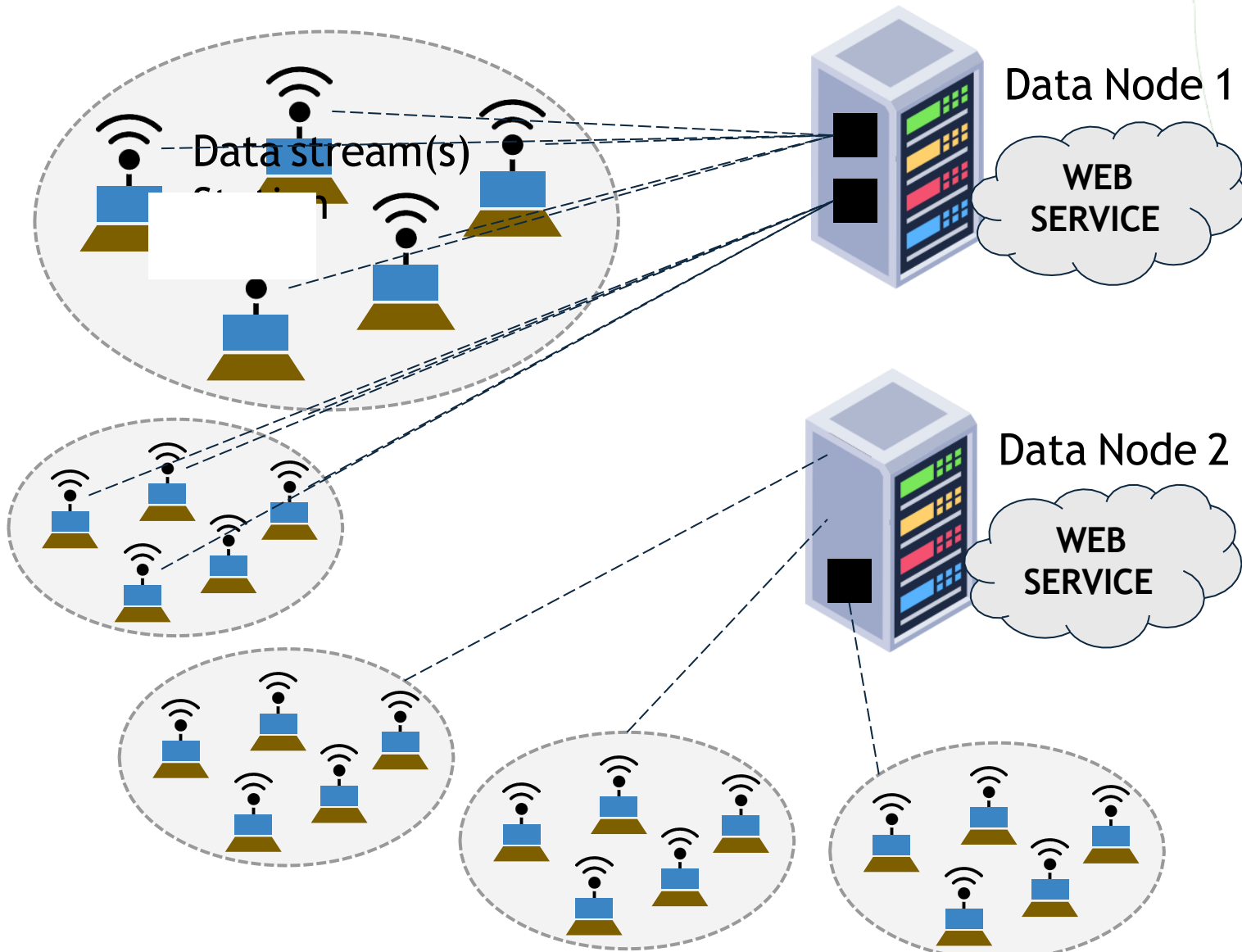


Goal: identify and describe

IGSN



Data streams from networks

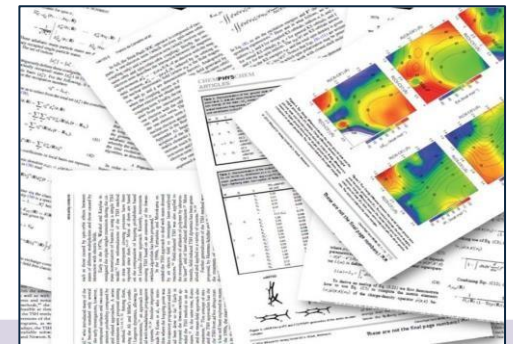


Environmental monitoring networks can be permanent or temporary.

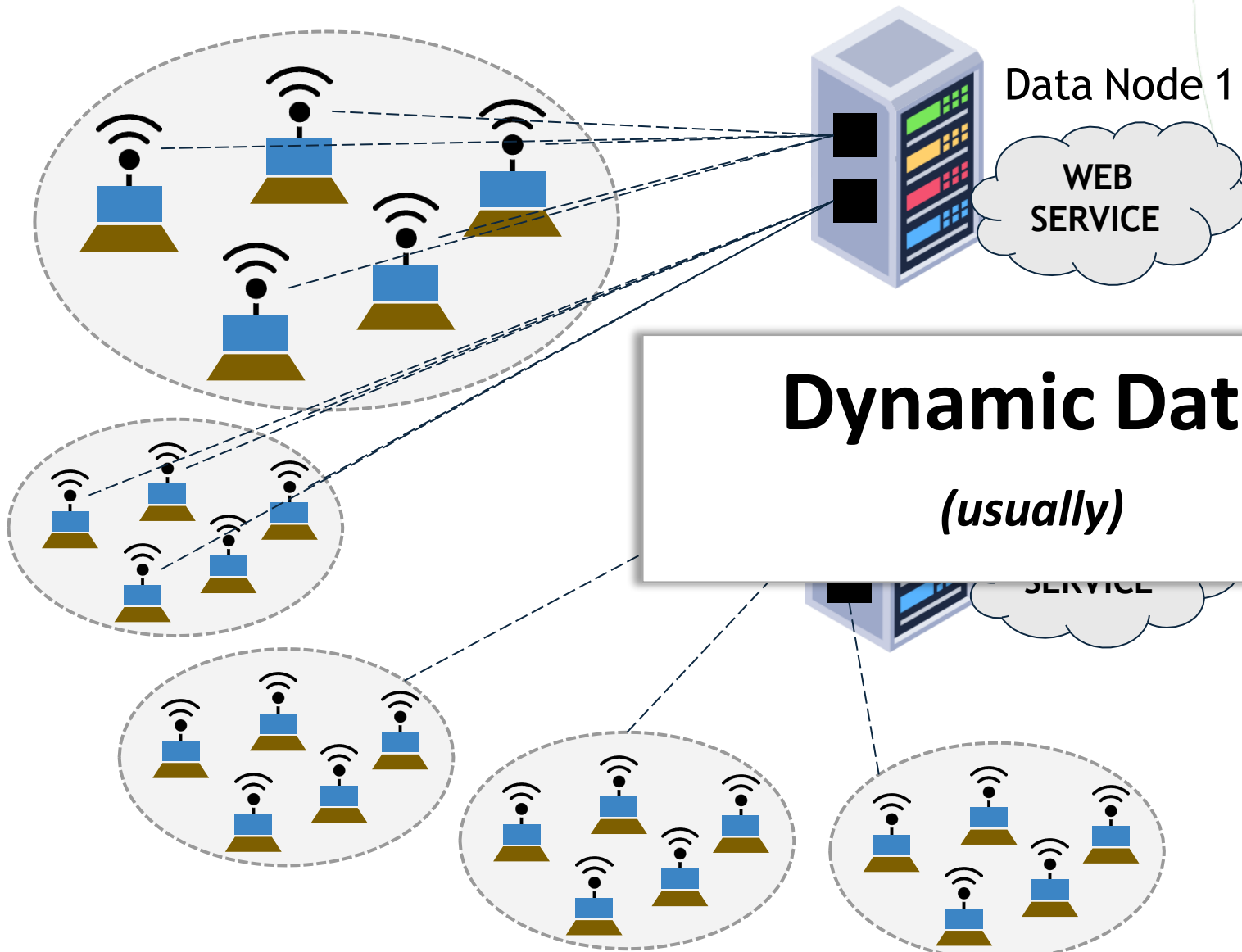
Data may be available in real-time or based on a periodic data collection, or manual reporting, causing delays.

Each relevant component should be identified and described.

Publications describing the network and the data workflow should be available and the relevant identifier to the data should be mentioned.



Data streams from networks

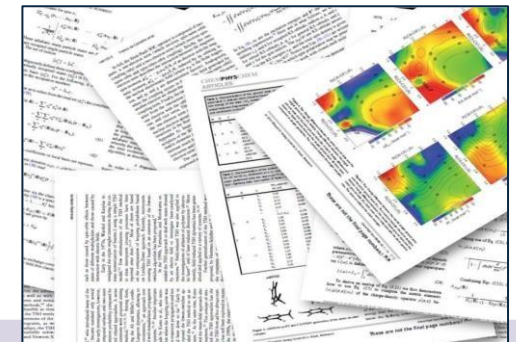


Environmental monitoring networks can be permanent or temporary.

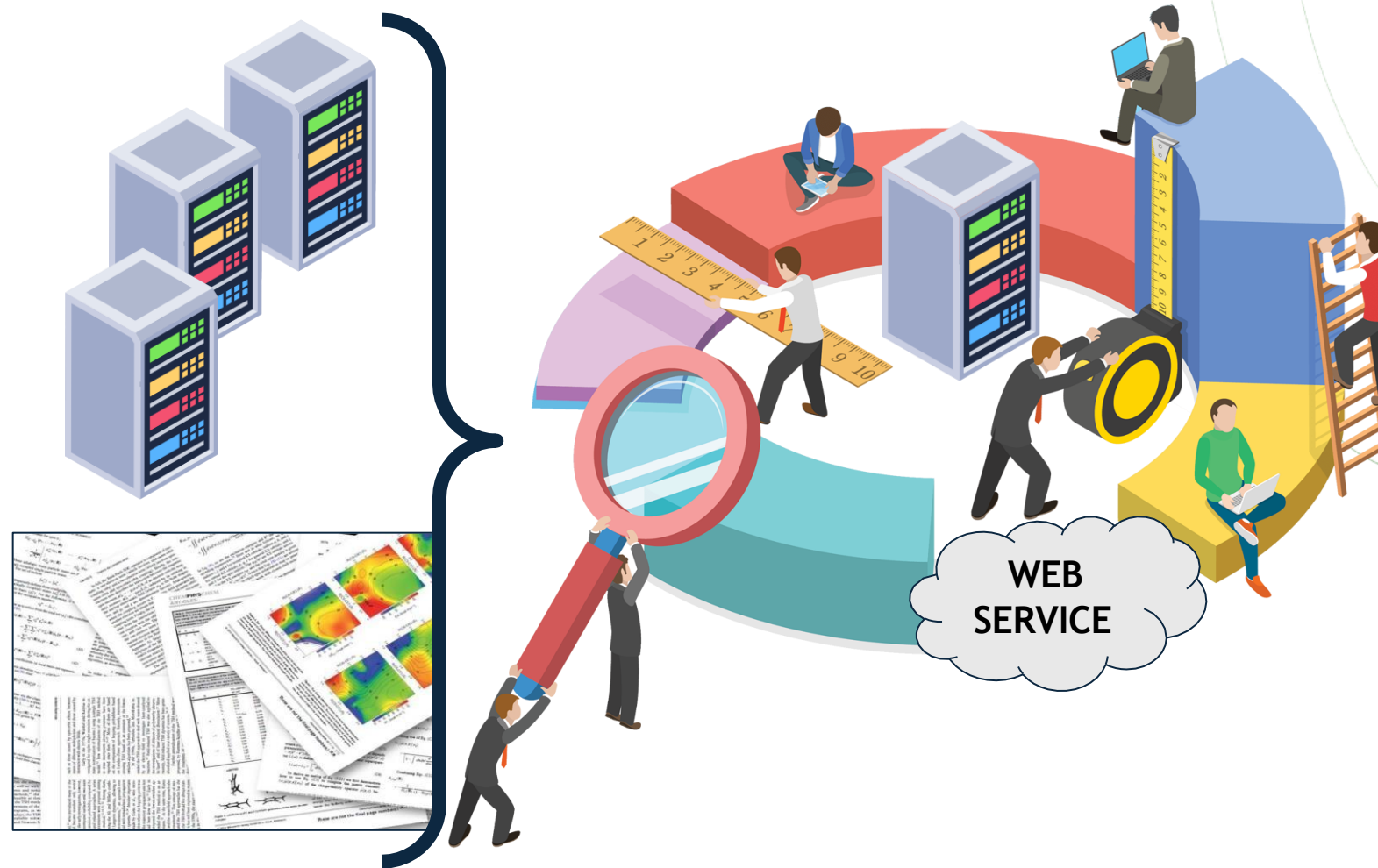
Data may be available in real-time or based on a periodic data collection, or manual reporting, causing delays.

Relevant component should be identified and described.

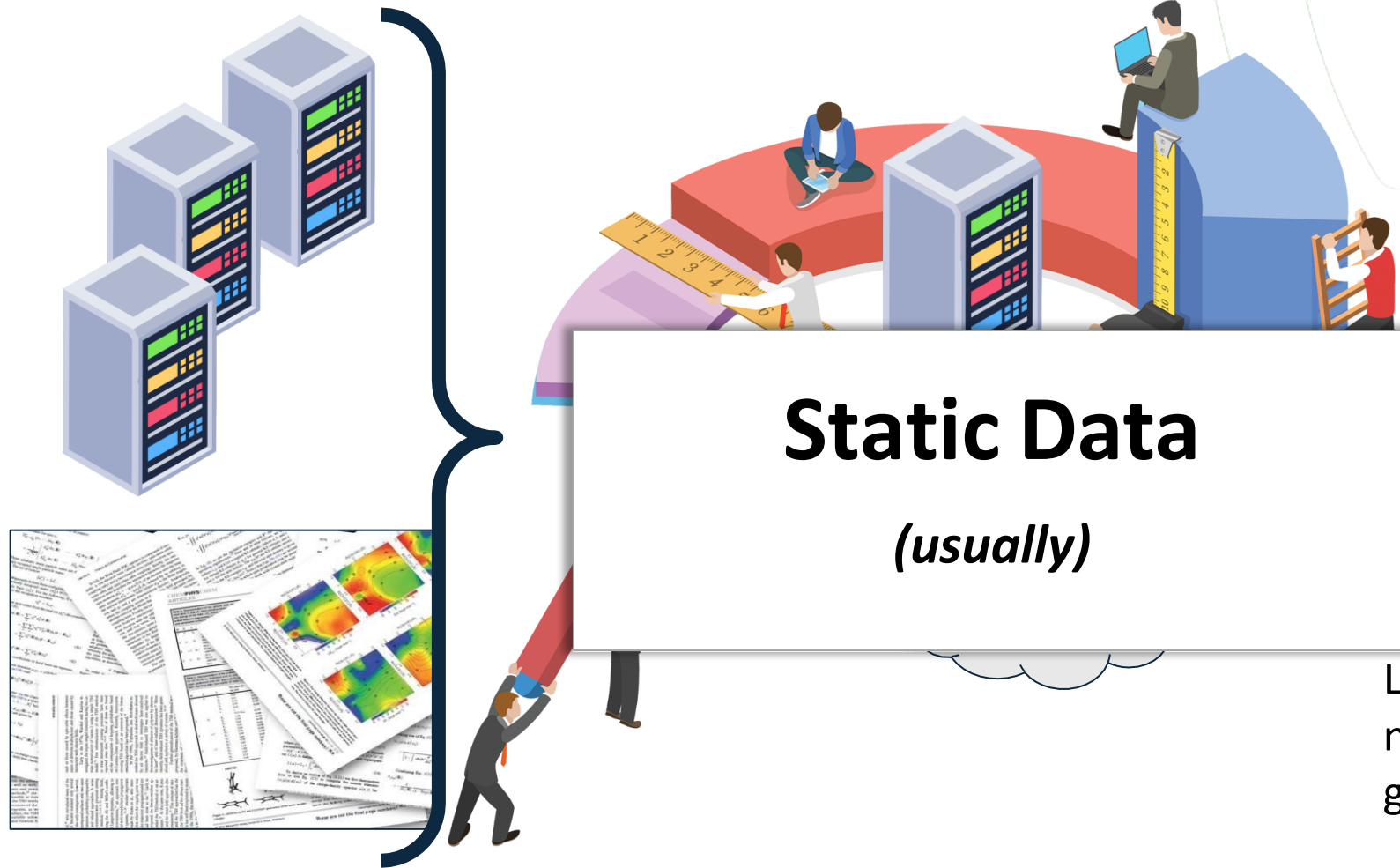
Metadata describing the network and a workflow should be available and the relevant identifier to the data should be mentioned.



Research Data Products



Based on the data coming from monitoring network(s) and from Literature sources, new data and publications are generated and publicly shared.



based on the data coming from monitoring network(s) and from Literature sources, new data and publications are generated and publicly shared.

Relationship between identifiers and metadata

Persistent identifiers allow elements to be uniquely identified, metadata describe them.

Some persistent identifier do not provide any description of what they identify (no metadata).



Some persistent identifiers are managed by organizations that also associate metadata.



Some metadata standard allows multiple persistent identifier standards.



What to identify? A question of granularity

Research products are by their nature complex, especially when it comes to data.

Deciding which elements to associate with identifiers is not easy; there is no single recipe.

Putting yourself in the shoes of potential users can help you understand the priorities and identify the most relevant elements unambiguously.

It is important to remember that research places reproducibility among its cornerstones, do clearly indicate to users all elements that have contributed to obtaining your results.

The same element can be:

- identified by multiple types of different identifiers
- described by multiple metadata standards, **mind that standards only partially overlap.**

Most common persistent identifiers you may encounter

Identifier	Extended name	Scope	Year	Homepage	Metadata schema
ISBN	International Standard Book Number	Books	1970	https://www.isbn.it/	ISBN
ISSN	International Standard Serial Number	Serials (e.g. magazines)	1976	https://www.issn.org/	ISSN
Handle	Handle System	Any online digital object	1994	https://www.handle.net/	No metadata
ISNI	International Standard Name Identifier	Who is involved in media creation	2012	https://isni.org/	ISNI
DOI	Digital object identifier	Many types of online digital object	2000	https://www.doi.org/	Crossref , DataCite , mEDRA , re3data
ARK	Archival Resource Key	Information objects of any type	2001	https://arks.org/	
IGSN	International Generic Sample Number	Physical samples	2007	https://www.igsn.org/	DataCite WIP
ORCID	Open Researcher and Contributor Identifier	Researchers	2009	https://orcid.org/	ORCID
RRID	Research Resource Identifiers	Research activities	2014	https://www.rrids.org/	
OFR	Open Funder Registry	Funder	2016	https://www.crossref.org/services/funder-registry/	Crossref
RAiD	Research Activity Identifier	Research activities	2017	https://www.raid.org.au/	DataCite
ROR	Research Organization Registry	Research organisations	2019	https://ror.org/	ROR
PIDINST	Instruments	Persistent Identification of Instruments	2020	https://www.rd-alliance.org/groups/persistent-identification-instruments-wg	WIP

URL, Uniform Resource Locator

A non persistent identifier

Invented by Tim Berners-Lee when he created the World Wide Web (WWW) in 1991.

It was standardized in 1994 by mutual agreement with the URI Working Group of the Internet Engineering Task Force (IETF).



The web server receives the request to a predefined port that is specific to each protocol, which can be omitted.

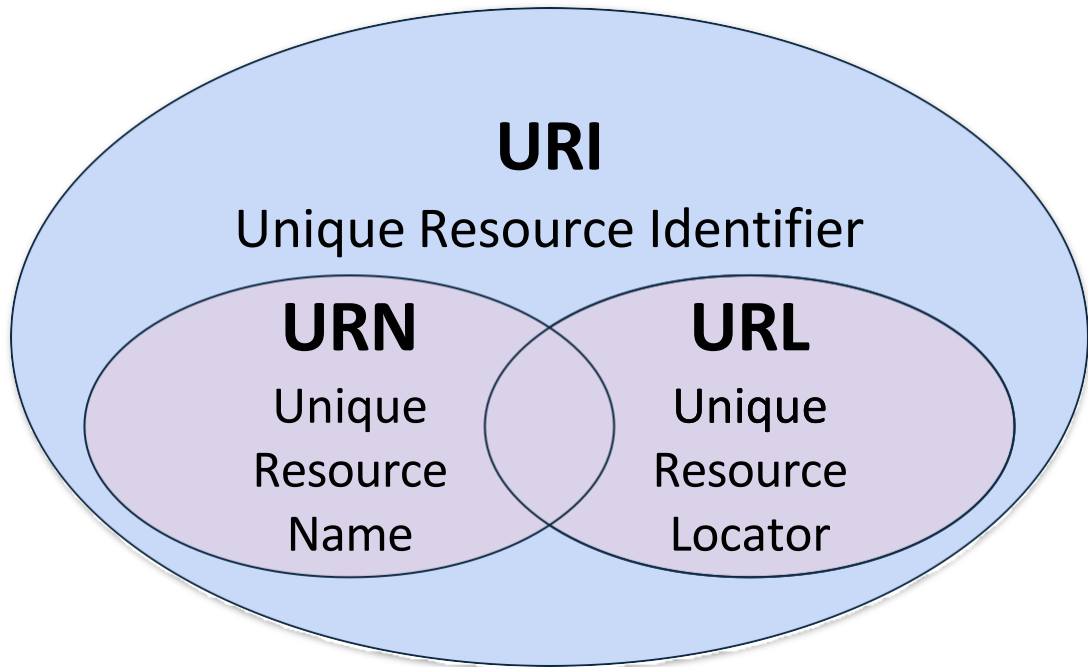
The default port for https (Hypertext Transfer Protocol Secure) is 443, so you can also write

<https://www.google.com:443/search?q=pnrr+itineris>

The old http protocol (Hypertext Transfer Protocol), now not recommended, uses the default port 80, so you can write

<http://www.google.com:80/search?q=meet+ipses>

Difference between URI, URN and URL



URI is a sequence of characters that universally identifies a resource within a schema [defined by IANA](#):

schema:element

Examples of schemes:

- https or http, for resources on the World Wide Web
- ftp, for resources available via File Transfer Protocol
- mailto, for email addresses
- files, to indicate files available on your computer

URN identifies a resource within the context of a “namespace” [defined by IANA](#), example using ISBN:

urn:isbn:978-88-6198-027-3

urn isbn 978-88-6198-027-3
URN namespace identifier

URN does not provide an indication of where the resource is located.

DCMI - Dublin Core

The Dublin Core metadata standard was created in 1995 and is managed by the Dublin Core Metadata Initiative (DCMI).

Created to allow the description of content published on the World Wide Web, the formalization in RDF dates to 2008.

The main use of the standard is to enrich metadata inserted directly into the source code of HTML pages.

It is used by search engines to retrieve useful information on the contents of web pages.

The metadata schema is available at <https://www.dublincore.org/schemas/>

Examples of “meta” tags in the “head” section of a HTML source code

```
<meta name="dc.type" content="Dataset" />
<meta name="dc.title" content="{title}" />
<meta name="dc.creator" content="{surname, name}" />    ← This element must be repeated for each author
<meta name="dc.publisher" content="Istituto Nazionale di Geofisica e Vulcanologia (INGV)">
<meta name="dc.identifier" scheme="DCTERMS.URI" content="https://doi.org/{doi}" />
<meta name="dc.rights" scheme="DCTERMS.URI" content="info:eu-repo/semantics/openAccess" />
<meta name="dc.rights.rightsHolder" content="Istituto Nazionale di Geofisica e Vulcanologia (INGV)">
<meta name="dc.license" content="Creative Commons Attribution 4.0 (CC BY 4.0)">
<meta name="dc.bibliographicCitation" content="<![CDATA[{here goes your bibliographic citation}]]>" />
```

Other meta standards designed for web pages

In addition to the terms defined in the original version of Dublin Core, in 2008 further elements were added described here <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

Example of "meta" tags in standard "dcterms" inserted in the "head" section of the HTML source

```
<meta name="dcterms.type" content="Dataset" />
<meta name="dcterms.title" content="{title}" />
<meta name="dcterms.creator" content="{surname, name}"> ← This element must be repeated for each author
<meta name="dcterms.publisher" content="Istituto Nazionale di Geofisica e Vulcanologia (INGV)">
<meta name="dcterms.identifier" content="https://doi.org/{doi}" />
<meta name="dcterms.license" scheme="dcterms.uri" content="https://creativecommons.org/licenses/by/4.0/" />
<meta name="dcterms.bibliographicCitation" content="<![CDATA[{testo della citazione bibliografica}]]>" />
```

In its "Scholar" platform, [Google support](#) in addition a standard initially developed by the publisher Highwire Press.

Example of metadata in Highwire Press format inserted in the "head" section of the HTML source

```
<meta name="citation_title" content="XXX TITLE XXX">
<meta name="citation_author" content="XXX SURNAME XXX, XXX NAME XXX">
<meta name="citation_publisher" content="Istituto Nazionale di Geofisica e Vulcanologia (INGV)">
<meta name="citation_keywords" content="XXX SUBJECTS XXX" />
<meta name="citation_doi" content="XXX DOI XXX" />
```

Handle

After the birth of the World Wide Web (WWW), the Defense Advanced Research Projects Agency (DARPA) required the construction in 1994 of a system to keep track of online resources that frequently changed addresses on the web (“link rot”).

The Handle System is operated by the Corporation for National Research Initiatives (CNRI), based in Virginia, USA.

For reasons of internet neutrality, the governance of the Handle System passed in 2014 to the DONA foundation in Switzerland.

The Handle identifier remains fixed while the associated URL can change over time, introducing persistence. The action of translating the Handle code into its corresponding URL is called "resolution" of a Handle.

The Code Handle is composed of **prefix:suffix**

To “resolve” a Handle, you compose this URL **https://hdl.handle.net/ + prefix:suffix**

Handle System only manages the link with a destination URL, it does not adopt a specific metadata standard.

Earth-Prints, the institutional archive of INGV established in 2005, assigns a Handle identifier to all archived elements.

Handle

Anyone can create and assign Handle codes,
To do this you need to install the «HANDLE NET» software
It is necessary to register as a “Handle Service Provider”
with CNRI and DONA, after validation of the request.

Each provider is assigned an identifier prefix.

Handle record

The most used function associated with the Handle code is the resolution to the corresponding URL.

but they are also available to REST APIs to interact programmatically, this is the related documentation

https://www.handle.net/proxy_servlet.html

APIs let users analyse the Handle Record as it contains a series of elements such as:

- **index**, an integer, usually progressive
- **type**, a string of text, there are usually at least two values in a record:
 - the “URL” type that indicates the URL to which the Handle identifier is associated
 - the “HS_ADMIN” type that indicates who is the owner who manages the record
- **data**, one or more elements defined by the pair of sub-elements:
 - format, which can take the following values: admin, string, hex, base64, vlist, site
 - value, for example contains the URL string when “type” is “URL”
- **ttl**, time-to-live, number of seconds that the record is considered valid, after which Handle updates the record
- **timestamp**, date the record was created or last updated

Example: <https://hdl.handle.net/api/handles/11304/a85d08eb-d771-4934-916c-c967fa4f1b7c>

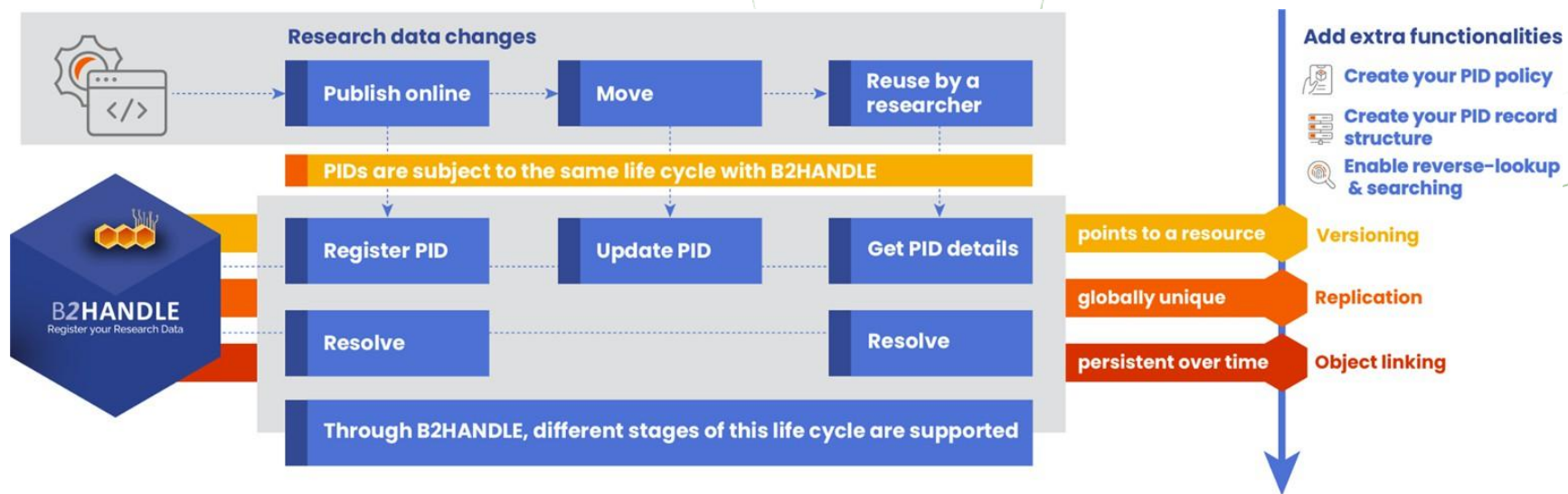
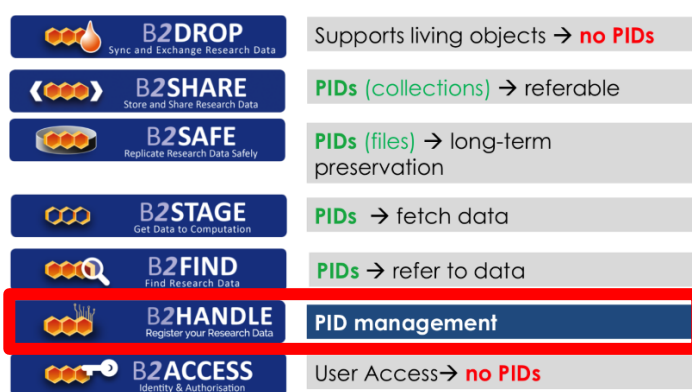
EUDAT



The Finland-based company EUDAT Collaborative Data Infrastructure (EUDAT CDI) was founded in 2018 thanks to two projects 1) FP7 [EUropean DATa EUDAT](#) (2011-2015) and 2) H2020 [EUropean DATa - EUDAT2020](#) (2015-2018).

EUDAT offers services like data replication, storage, sharing, and identity management. Some services are free, most services require a fee when used extensively.

Among their services offered for a fee, B2HANDLE allows the management of PID and is based on Handle System.



B2HANDLE does not adopt a specific metadata standard, the management of which is delegated to users.

DOI - Digital Object Identifier



ISO 26324 v1, 2012 → v2, 2022

The DOI Foundation ("International DOI Foundation", IDF) was established in 1997.

IDF is the largest "Handle Service Provider" in existence and began assigning DOIs in 2000.

To assign DOIs, IDF uses [authorised Registration Agencies](#).

DOI identifiers are based on the Handle system by adding various elements such as:

- a standardization of the types of elements associated with the record
- rules for the nomenclature and management of countless cases
- a minimal metadata model to ensure interoperability ([DOI kernel schema](#))
- larger metadata models are managed by individual registration agencies
- REST APIs with more parameters for records management

The management of DOIs is described in the [DOI Handbook](#).

Each registration agency operates in specific areas, the best known are:

- **Crossref**, founded in 1999, a non-profit association founded by a series of publishers active in the scientific field;
- **DataCite**, founded in 2009, is a non-profit association founded by organizations interested in the exchange of data.

DOI - API REST



The DOI foundation provides a **REST API that forward the requests to the Handle System.**

Documentation is available at <https://www.doi.org/doi-handbook/HTML/rest-api-request-format.html>

Each API requests follow this structure

`https://doi.org/api/handles/<doi>?<param_1>&<param_2>&..`

Resolving a DOI can be achieved with the call

<https://doi.org/api/handles/10.13127/epica.1.1>

The same result can be achieved using the Handle System service with the call

<https://hdl.handle.net/api/handles/10.13127/epica.1.1>

To extract only a specific type of item, such as “URL”

<https://doi.org/api/handles/10.13127/epica.1.1?type=URL>

DOI - Parameters accepted by the resolution system



List of the main parameters accepted by the [DOI Resolution Service](#):

- To get the entire record associated with a DOI code, pass the “**noredirect**” parameter
Example: <https://doi.org/api/handles/10.13127/epica.1.1?noredirect>
- To request a specific type of item contained in the DOI record, you pass the “**type**” parameter followed by the type
Example to get the "URL" type element in JSON format
<https://doi.org/api/handles/10.4401/ag-8756?type=URL>
- To pass parameters to the destination server that manages the URL associated with the DOI, use “**urlappend**”
For example, DOI “10.13127/asmi” resolves to <https://emidius.mi.ingv.it/ASMI/>
Example on how to add “/query_study/?filter_year=2022” suffix to the resolved destination URL
https://doi.org/10.13127/asmi?urlappend=/query_study/?filter_year=2022
The resolved URL call is
https://emidius.mi.ingv.it/ASMI/query_study/?filter_year=2022
This is a practical solution for achieving “**Data slicing**”, i.e. providing sub-sets of data.

N.B. A suffix may be added after the DOI using “#” symbol but in this way the suffix cannot be read by the destination server, so for example https://doi.org/10.13127/asmi#/query_study/?filter_year=2022 will not work

DOI - Content negotiation



Handle System, and therefore also DOI, allows you to tell the server that hosts the identified resource in which format you want to receive the answer, to do so you specify the requested [“media type”](#) (aka [“MIME type”](#)).

Assuming that the user is operating from a web browser, the "text/html" format is set as the default “media type”.

If you run the following command from a terminal (Linux Bash, macOS, Windows PowerShell)

```
curl https://doi.org/10.4401/ag-8756
```

the HTML web page associated with the DOI is returned

```
curl -LH "Accept: text/html" https://doi.org/10.4401/ag-8756
```

The server receiving the request may support additional formats (additional “media types”), for example:

- support to RDF-XML output

```
curl -LH "Accept: application/rdf+xml" https://doi.org/10.4401/ag-8756
```

- support to JSON output

```
curl -LH "Accept: application/json" https://doi.org/10.4401/ag-8756
```

Example bash code to check for supported media formats



Requirement: any linux bash terminal

Steps

1. create a text file with the code at the right
`nano test_mediatypes.sh`
2. make the file executable
`chmod +x test_mediatypes.sh`
3. launch the test
`./test_mediatypes.sh`
4. decide whether you want to check your identifier using either “handle” or “doi”
5. enter the identifier you want to check
6. see which media types listed are supported

```
#!/bin/bash
printf "\n"
read -p "Check using Handle or DOI resolution mechanism? [enter either 'handle' or 'doi'] " api_type
if [ $api_type == 'doi' ]; then
    url="https://doi.org/"
else
    if [ $api_type == 'handle' ]; then
        url="https://hdl.handle.net/"
    else
        printf "unknown answer\n"
        exit 0
    fi
fi
read -p "Enter the identifier: " identifier
url=$url$identifier
media_types=(
    "text/plain"
    "text/html"
    "text/xml"
    "text/csv"
    "text/turtle"
    "application/pdf"
    "application/json"
    "application/xml"
    "application/rdf+xml"
    "application/vnd.ms-excel"
    "application/vnd.openxmlformats-officedocument.spreadsheetml.sheet"
    "application/vnd.oasis.opendocument.spreadsheet"
    "application/x-sqlite3"
    "application/vnd.opengeospatial.geopackage+sqlite3")
printf "\nEndpoint that will be checked: $url\n\n"
for media_type in "${media_types[@]}; do
    printf "Checking support for '$media_type': "
    response=$(curl -s -o /dev/null -w "%{http_code}" -L -H "Accept: $media_type" "$url")
    if [ "$response" -eq 200 ]; then
        printf "\e[32mSUPPORTED\e[0m\n"
    else
        printf "\e[31mNOT SUPPORTED\e[0m\n"
    fi
done
printf "\n"
```

DOI - Kernel metadata schema



The DOI Foundation maintains a [minimal metadata schema](#) common to all agencies:

- **doiName**, contains the DOI code
- **primaryReferentType**, contains the creator of the DOI
- **registrationAgencyDoiName**, contains the Registration Agency
- **issueDate**, contains the date of the DOI creation
- **issueNumber**, contains the number of the record revision

followed by one or more of the following elements

- **referentCreation**, with the list of all authors of the item identified by the DOI code
- **referentParty**, with the list of details related to the creator of the DOI code
- **referentPlace** [optional], list of places associated with the authors and the creators of the DOI

All DOI authorised Registration Agencies must include all these elements in their extended metadata schemas.

DOI - Crossref

Non-profit organization founded in 1999 to meet the needs of publishers active in the online scientific field.

The most famous companies that founded Crossref:

- Academic Press
- American Institute of Physics
- American Physical Society
- Blackwell Science
- Elsevier
- Institute of Electrical and Electronics Engineers (IEEE)
- John Wiley & Sons
- Nature Publishing Group
- Oxford University Press
- Springer-Verlag



DOI - Crossref - Schema di metadati



Crossref

The metadata schema is at version 5.3.1 and is available at this address

<https://data.crossref.org/schemas/crossref5.3.1.xsd>

The documentation is available at this address

https://data.crossref.org/reports/help/schema_doc/5.3.1/index.html

The metadata schema is designed to describe textual objects published by publishers, for example:

- book
- conference
- dissertation
- report-paper
- database
- peer review
- pending publication

DOI - Crossref - API REST



The REST API documentation for querying Crossref can be found at <https://api.crossref.org/>

To get metadata in JSON format, the call is
`https://api.crossref.org/works/{doi}`

Example

<https://api.crossref.org/works/10.4401/ag-8756>

The APIs provide information only for DOIs assigned through Crossref services.

If you ask for a DOI code registered by a different agency, the system returns an HTTP 404 "Not found" error.

DOI - DataCite



A non-profit organization founded in 2009 to meet the needs of publishers active in the online scientific field.

The most famous organisations that founded DataCite :

- Academic Press
- American Institute of Physics
- American Physical Society
- Blackwell Science
- Elsevier
- Institute of Electrical and Electronics Engineers (IEEE)
- John Wiley & Sons
- Nature Publishing Group
- Oxford University Press
- Springer-Verlag

DOI - DataCite Commons

DataCite Commons is a web interface that allows users to explore research outputs such as datasets, publications, software.



It provides a comprehensive overview of connections between these outputs, their authors and their organisations, making it easier to discover and reuse research

The screenshot shows the profile page for Mario Locati on DataCite Commons. The profile includes a search bar, navigation tabs for Works, People, Organizations, and Repositories, and a share section with email, Twitter, and Facebook options. The main content area displays the researcher's name, ORCID ID, and various identifiers. It also lists employment at the Istituto Nazionale di Geofisica e Vulcanologia Sezione di Milano. A section for 'Aggregated Citations, Views and Downloads' shows 226 citations and accessibility achievements. At the bottom, there are filter sections for 'Filter Works' and 'Co-Authors'.

The screenshot shows the organization page for Istituto Nazionale di Geofisica e Vulcanologia on DataCite Commons. The page features a search bar, navigation tabs, and a 'Download Reports' section with statistics: 1,569 Works, 999 Citations, 375 Views, and 29 Downloads. It also includes a 'Share' section with social media links, 'Other Identifiers', and 'Geolocation' information. A 'Filter Works' section is present, along with a 'Creators & Contributors' list. The bottom of the page contains three charts: 'Publication Year' (a bar chart showing growth from 2015 to 2025), 'Work Types' (a stacked bar chart showing 54% Text), and 'Licenses' (a stacked bar chart showing 65% CC-BY-4.0).

DOI DataCite - Schema di metadati

The metadata schema is at version 4.6 and is available at <https://schema.datacite.org/meta/kernel-4.6/metadata.xsd>

The documentation is available at <https://datacite-metadata-schema.readthedocs.io/en/4.6/>

The metadata schema is designed to describe a broad spectrum of research outputs, such as:

- Audiovisual
- Award
- Book
- BookChapter
- Collection
- ComputationalNotebook
- ConferencePaper
- ConferenceProceeding
- DataPaper
- Dataset
- Dissertation
- Event
- Image
- InteractiveResource
- Instrument
- Journal
- JournalArticle
- Model
- OutputManagementPlan
- PeerReview
- PhysicalObject
- Preprint
- Project
- Report
- Service
- Software
- Sound
- Standard
- StudyRegistration
- Text
- Workflow



DOI DataCite - Schema di metadati - Dettaglio



The scheme allows you to include:

- DOI code
- URL associated with the DOI
- main title and alternative titles, for each title the language can be specified
- version of the resource identified
- year of publication
- list of lead authors, with first and last name, ORCID code (researcher) and ROR code (organization)
- list of contributors, as above but with the addition of the role
- general description (multilingual), description of the methods used and technical description
- geographical coverage, either in term of a generic text, a geographical polygon(s) or list of points
- time coverage, with a start date, and if any, the end date (→ useful for dynamic data)
- format(s) used for data encoding
- keyword(s)
- license associated with the identified object
- organization(s) that own the resource
- funding projects
- relations to other digital online resources. Each relationship can be associated with a relationship type

DOI DataCite - Relations to other online resources

Description of the relationship between the currently selected data (A) and the related object (B).

Relation	Definition
IsCitedBy	indicates that B includes A in a citation
IsCompiledBy	indicates B is used to compile or create A
IsContinuedBy	indicates A is continued by the work B
IsDescribedBy	indicates A is described by B
IsDocumentedBy	indicates B is documentation about/ explaining A; e.g. points to software documentation
IsMetadataFor	indicates additional metadata A for a resource B
IsNewVersionOf	indicates A is a new edition of B, where the new edition has been modified or updated
IsObsoletedBy	Indicates A is replaced by B
IsPartOf	indicates A is a portion of B; may be used for elements of a series
IsReferencedBy	indicates A is used as a source of information by B
IsRequiredBy	Indicates A is required by B
IsReviewedBy	indicates that A is reviewed by B
IsSourceOf	indicates A is a source upon which B is based
IsSupplementTo	indicates that A is a supplement to B
IsVariantFormOf	indicates A is a variant or different form of B
IsVersionOf	indicates A is a version of B
IsIdenticalTo	indicates that A is identical to B, for use when there is a need to register two separate instances of the same resource

Relation	Definition
Cites	indicates that A includes B in a citation
Compiles	indicates B is the result of a compile or creation event using A
Continues	indicates A is a continuation of the work B
Describes	indicates A describes B
Documents	indicates A is documentation about B; e.g. points to software documentation
HasMetadata	indicates resource A has additional metadata B
IsPreviousVersionOf	indicates A is a previous edition of B
Obsoletes	Indicates A replaces B
HasPart	indicates A includes the part B
References	indicates B is used as a source of information for A
Requires	Indicates A requires B
Reviews	indicates that A is a review of B
IsDerivedFrom	indicates B is a source upon which A is based
IsSupplementedBy	indicates that B is a supplement to A
IsOriginalFormOf	indicates A is the original form of B
HasVersion	indicates A has a version (B)

DOI DataCite - Limitations

As of today, the DataCite metadata schema does not provide information for a fully machine-actionable access to data.

For example:

- dedicated tags directly pointing to downloadable file
(more than one encoding data format might be available)
- dedicated tags directly pointing to web services (aka web APIs)
(more than one web service standard might be supported)
- it is not possible to state the period of activity for each involved person which is especially important for very long-term dynamic data (multiple years)

DOI DataCite - API REST



Documentation of the REST APIs to query the DataCite metadata registry at <https://support.datacite.org/docs/api>

To get metadata in JSON format, the call is
`https://api.datacite.org/doi/{doi}`

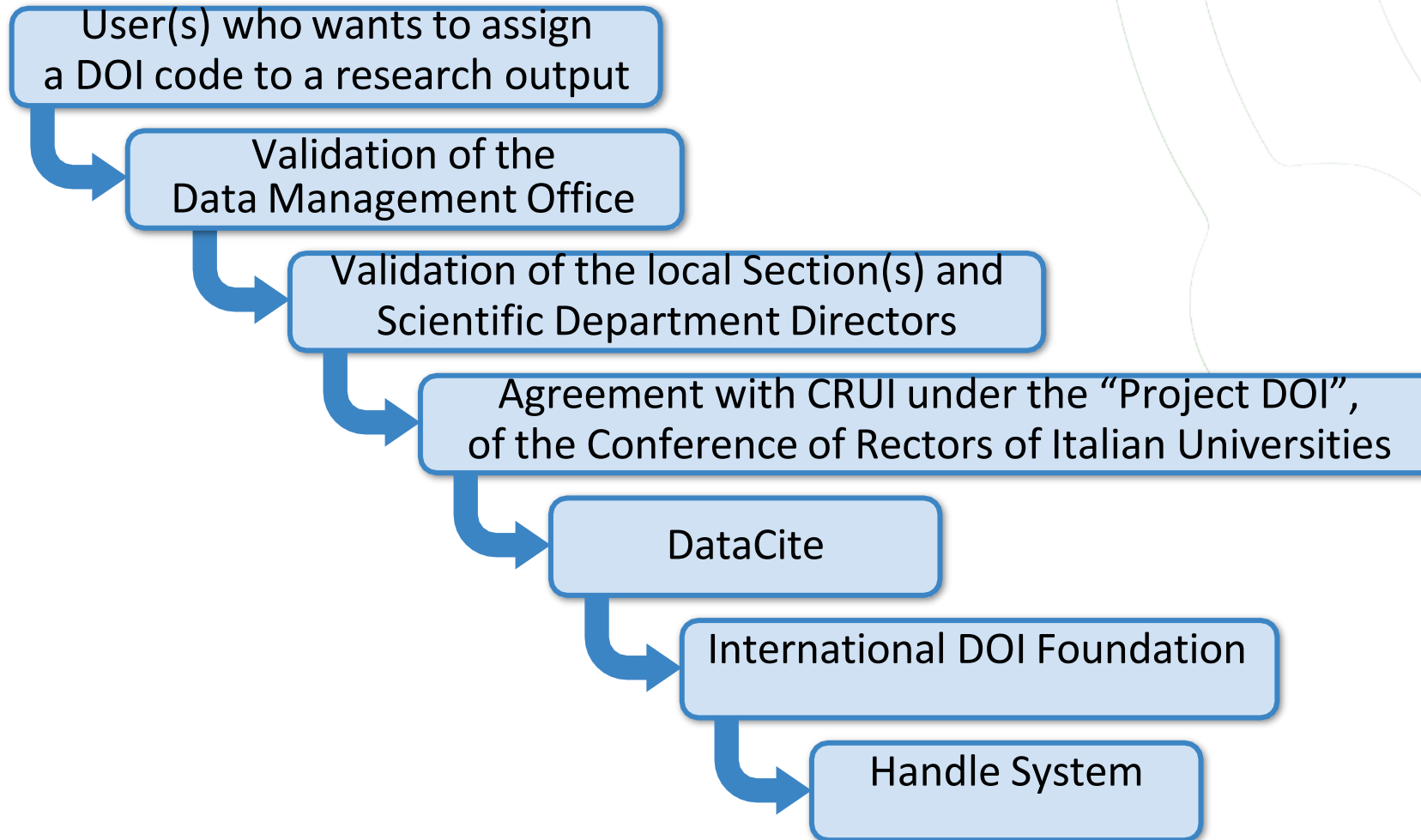
For example
<https://api.datacite.org/doi/10.13127/epica.1.1>

The APIs provide information only for DOIs assigned through DataCite services.

If you ask for a DOI code registered by a different agency, the system returns an HTTP 404 “Not found” error.

How INGV enables its own DOI minting service

List of subjects/organisations allowing a person affiliated to INGV to assign a DOI to a dataset



How to determine which agency has assigned the DOI

A REST APIs managed by the DOI Foundation let you identify which Agency has assigned the DOI code, the call is

`https://doi.org/ra/{DOI}`

Examples from different agencies:

- <https://doi.org/ra/10.13127/cpti/cpti15.4> → DataCite
- <https://doi.org/ra/10.1007/s10518-020-00818-y> → Crossref
- <https://doi.org/ra/10.26324/SIA1.PNS> → mEDRA
- <https://doi.org/ra/10.3030/101058527> → OI, Publications Office of the European Union (OP)

This service makes retrieving metadata easier for users.

Automatic bibliographic citation associated with a DOI



The DOI foundation provides a service to generate a bibliographic citation by exploiting the metadata available at <https://citation.crosscite.org/>

All DOIs created by authorised Registration Agencies Crossref (publications, global), DataCite (various products, global), mEDRA (Europe, multilingual DOIs), ChineseDOI (China), JapanLinkCenter, Korean DOI Center (Korea), AIRITI DOI (Taiwan).

In addition to the web interface, it is possible to make a call to the service indicating the [bibliographic style](#) and language.

Example

```
curl -LH "Accept: text/x-bibliography; style=harvard3; locale=en-GB" https://doi.org/10.13127/epica.1.1
```

outputs

```
Rovida, A., Locati, M., Antonucci, A., & Camassi, R. (2017). Archivio Storico Macrosismico Italiano (ASMI) (Version 1.0). Istituto Nazionale di Geofisica e Vulcanologia (INGV).  
https://doi.org/10.13127/ASMI
```

Tips for managing DOIs

Keep your metadata up to date:

- Add relations to publications as they become available
- Add relations to input or derived data
- Enter the geographical coverage to enable a geographical search of your data by users
- Enter the time coverage of your data to enable users to find your data by temporal search

The elements that make up the bibliographic citation should never change:

- list of main authors
- title
- year of publication

Make informed decisions about authors, especially for long-term datasets that may see turnovers.

Tip: In this case you may consider providing the generic name of a workgroup and dynamically define the list of contributors.

ARK - Archival Resource Key



The ARK persistent identifier system was established in 2001 and since 2009 it was registered as a Uniform Resource Identifier (URI) scheme.

California Digital Library began using ARKs in 2002, other early adopters of ARKs the Internet Archive, and the Bibliothèque Nationale de France.

`https://NMA/ark:/NAAN/Name [Qualifier]`

NAAN: Name Assigning Authority Number, mandatory unique identifier of the organization that originally named the object

NMA: Name Mapping Authority, optional and replaceable hostname of an organization that provides service for the object

Qualifier: optional string that extends the base ARK to support access to individual hierarchical subcomponents of an object, and to variants (versions, languages, formats) of components.

The v34 is the latest version and it was published November 10, 2024.

The documentation is available at https://arks.org/assets/documents/2024/ark_spec_39.pdf

Example <https://n2t.net/ark:/53355/cl010066723>

ORCID - Open Researcher and Contributor ID

A non-profit organization founded in 2009 to meet the converging needs of universities and research organizations, scientific publishers and research funding organizations.

All these subjects have the common need to uniquely identify people, ORCID is an open alternative to proprietary systems such as:

- Scopus Author ID, managed by ELSEVIER
- ResearchrID, managed by Thomson Reuters
- Scholar Profile, managed by Google
- Loop Profile, managed by Frontiers
- LinkedIn, managed by Microsoft
- Academia
- ResearchGate



ORCID - Metadata Schema



The metadata schema is designed to describe a person who works in the world of Research and understands, it includes:

- Name and surname and their variants
- Email, address
- Biography
- Area of expertise
- List of affiliations
- List of published works
- list of activities carried out including participation in funded research projects

The metadata schema is at version 3.0 and is available at this address

https://github.com/ORCID/orcid-model/tree/master/src/main/resources/record_3.0

The documentation is available at this address

<https://support.orcid.org/hc/en-us/articles/360006971353-Metadata-in-the-Works-section>

ORCID - API REST



REST API documentation for querying ORCID can be found at <https://info.orcid.org/documentation/features/public-api/>

To get metadata in JSON format, the call is
`https://pub.orcid.org/v3.0/{orcid}`

For more advanced functions, it is necessary to register for the service and be assigned a personal token.

Free access example to an ORCID record without personal tokens, it returns XML by default
<https://pub.orcid.org/v3.0/0000-0003-2185-3267>

Example of terminal requests

- request output in JSON format

```
curl -LH "Accept: application/json" https://pub.orcid.org/v3.0/0000-0003-2185-3267
```

- request output in XML format

```
curl -LH "Accept: application/xml" https://pub.orcid.org/v3.0/0000-0003-2185-3267
```

IGSN - International Generic Sample Number

A non-profit organisation founded in 2011 and controlled by GFZ in Potsdam. IGSN was initially created as part of the Solid Earth Health Registry (SESAAR) funded by the U.S. National Science Foundation.

The aim is to uniquely identify and describe the physical samples handled in the laboratories.



Physical Sciences:

- Rocks, minerals, metals
- Radioactive elements
- Soil, sediments (core & individual)
- Mud
- Dust
- Petroleum, oil
- Condensate, water, ice cores
- Air, atmosphere, gas
- Environmental DNA
- Astromaterials
- Other chemical samples

Life Sciences:

- Biofluids
- Tissues
- Cells
- Genomic
- Animal specimens
- Insect specimens
- Plant specimens
- Fungi, spores
- Microorganisms, biofilms
- Vegetation
- Botanical, herbarium samples

Social Sciences:

- Human, animal, insect, plants
- Fossils
- Artefacts: tools, weapons, utensils, machines, ornaments, art, written records, religious images, clothing,...

Applied Sciences:

- Seed accessions
- Crop & livestock samples
- Soil testing
- Synthetic materials
- Prototypes

IGSN - Metadata schema and API access



At the beginning IGSN was based on Handles.

Starting in 2021, the organization has started a close cooperation with DataCite.

IGSN metadata schema is modular:

- the **core schema** is based on the DataCite schema;
- the **domain specific schema** is developed on a community-based initiative.

In the DataCite schema the “resourceTypeGeneral” property must be set as “PhysicalObject”.

Samples are related to one another using the DataCite relations descriptors.

Access to metadata via API also leverages DataCite's services.

Recommendations are available for compiling metadata:

- <https://support.datacite.org/docs/igsn-id-metadata-recommendations>
- <https://docs.google.com/document/d/16GTTvcnYvTDIYvypCq9dimNhDFK-fCipayO6YUngc90/edit>
- https://igsn.uni-kiel.de/documentation/en/specifications/metadata_schema.html

ROR - Research Organization Registry

The registry was established in 2019 and is managed in collaboration between the California Digital Library, Crossref and DataCite.

It replaced the previous Global Research Identifier Database (GRID) system that was in operation between 2015 and 2021.

ROR allows you to describe the following elements of a research organization:

- Official name and its variants such as the English translation
- Address and geographical location
- Homepage URL
- Type of organization
- List of other available identifiers
- Relationship with other organizations, allows, for example, to assign ROR identifiers to departments and the various branches of a main organization

ROR - Requests, metadata schema and API

Anyone can request additions or changes to the register.

There are two ways to submit requests to ROR:

- from a web form at <https://curation-request.ror.org/>
- creating an issue at <https://github.com/ror-community/ror-updates/issues>

The metadata schema is in its 2nd version and is available at <https://ror.readme.io/v2/docs/data-structure>

Access via API is documented at <https://ror.readme.io/v2/docs/rest-api>

Example of a call that provides JSON output by default
<https://api.ror.org/organizations/00qps9a02>

OFR - Open Funder Registry

The register uniquely identifies the organizations that finance the Research activities, the register is managed by Crossref. Each funding organization is assigned a DOI identifier.

Consultation of the register is possible at <https://search.crossref.org/funding>

API documentation at

<https://www.crossref.org/documentation/retrieve-metadata/rest-api/>

<https://api.crossref.org/swagger-ui/>

Examples

<https://api.crossref.org/funders?query=ingv>

The log can be bulk downloaded as RDF or CSV formats at

https://gitlab.com/crossref/open_funder_registry/

<https://doi.crossref.org/funderNames?mode=list>



RRID - Research Resource Identifiers



The RRID was established in 2014 by FORCE11, a Nonprofit organization based in the USA made of a community of scholars, librarians, archivists, publishers, and research funders dedicated to improving scholarly communication and e-scholarship.

The metadata schema is designed to cover various types of research resources:

- plasmids
- cells
- organisms
(*e.g. rodents, invertebrates, fish, amphibians, high mammals, single cell*)
- biosamples
- antibodies
- tools such as software and databases

Documentation at <https://rrid.site/>

RAiD - Research Activity Identifier Service



The [RAiD identifier](#) was created in 2018 and is managed by the [Australian Research Data Commons](#) (ARDC) and follows the ISO 23527 standard, of which there is only one version published in 2022.

RAiD assigns a DOI to research activities, tracks and links together all the components of a research project:

- Research Projects, describes the context and objectives of the project.
- Dataset, connects the data collected and used during the search.
- Publish, associate articles, reports, and other scholarly outputs with the project.
- Funding, track funding sources and associated grants.
- Collaborators, identifies the researchers and institutions involved.

RAiD is based on DataCite's DOI identifier infrastructure and uses both the metadata schema and API access services.

The [re3data registry](#) was established in 2012 and its infrastructure is managed by Karlsruhe Institute of Technology (KIT) in collaboration with:

- Berlin School of Library and Information Science, Humboldt-Universität zu Berlin
- Helmholtz Open Science Office at GFZ German Research Centre for Geosciences
- Libraries of the Purdue University, Indiana, USA

re3data allows the description of the following aspects of a data repository:

- DOI DataCite
- name, additional name
- URL
- repository identifier
- description
- contact
- type
- size
- date (start, end)
- related repository
- language
- subjects
- keywords
- type of provider
- institution
(name, country, type, url, identifier)
- policy
- funding
- access
- license
- data upload
- software
- versioning
- API (type, url, documentation)
- PID system
- citation
- metrics
- quality management (status, measure)
- certificate (name, url)

re3data - Metadata schema and API

The metadata schema is in its 4th version and is described at this address

<https://doi.org/10.48440/re3.014>

The XML schema of v4 in XSD format is available at

<https://doi.org/10.48440/re3.015>

Access via API is documented at

<https://www.re3data.org/api/doc>

The standard API call follows this structure

<https://www.re3data.org/api/v1/repository/{DOI}>

Example of a browser call that returns XML output

<https://www.re3data.org/api/v1/repository/r3d100014389>

INSPIRE - ISO 19115/19139

The ISO 19115 standard is the European reference for INSPIRE.

It is the standard adopted by the Italian node of INSPIRE called **“Repertorio Nazionale dei Dati Territoriali”** (RNDT) for the metadata description of data and related access services.

The corresponding XML encoding is defined by the ISO 19139 standard.

Standardization is supervised by the ISO TC211 committee.

RNDT provides practical guidance for data declaration

<https://geodati.gov.it/geoportale/documenti>



geodati.gov.it

Repertorio Nazionale dei Dati Territoriali

Data Catalog Vocabulary Application profile for data portals

It is a specification based on the vocabulary of the Data Catalogue (DCAT) for the description of public sector datasets.

It is designed to enable the search of datasets on cross-referenced data portals and make public sector data more searchable across borders and sectors by facilitating the exchange of dataset descriptions between data portals.

Latest version 3.0, August 29, 2024

<https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe>

Description

<https://www.w3.org/TR/vocab-dcat-3/>

UML Class Diagram

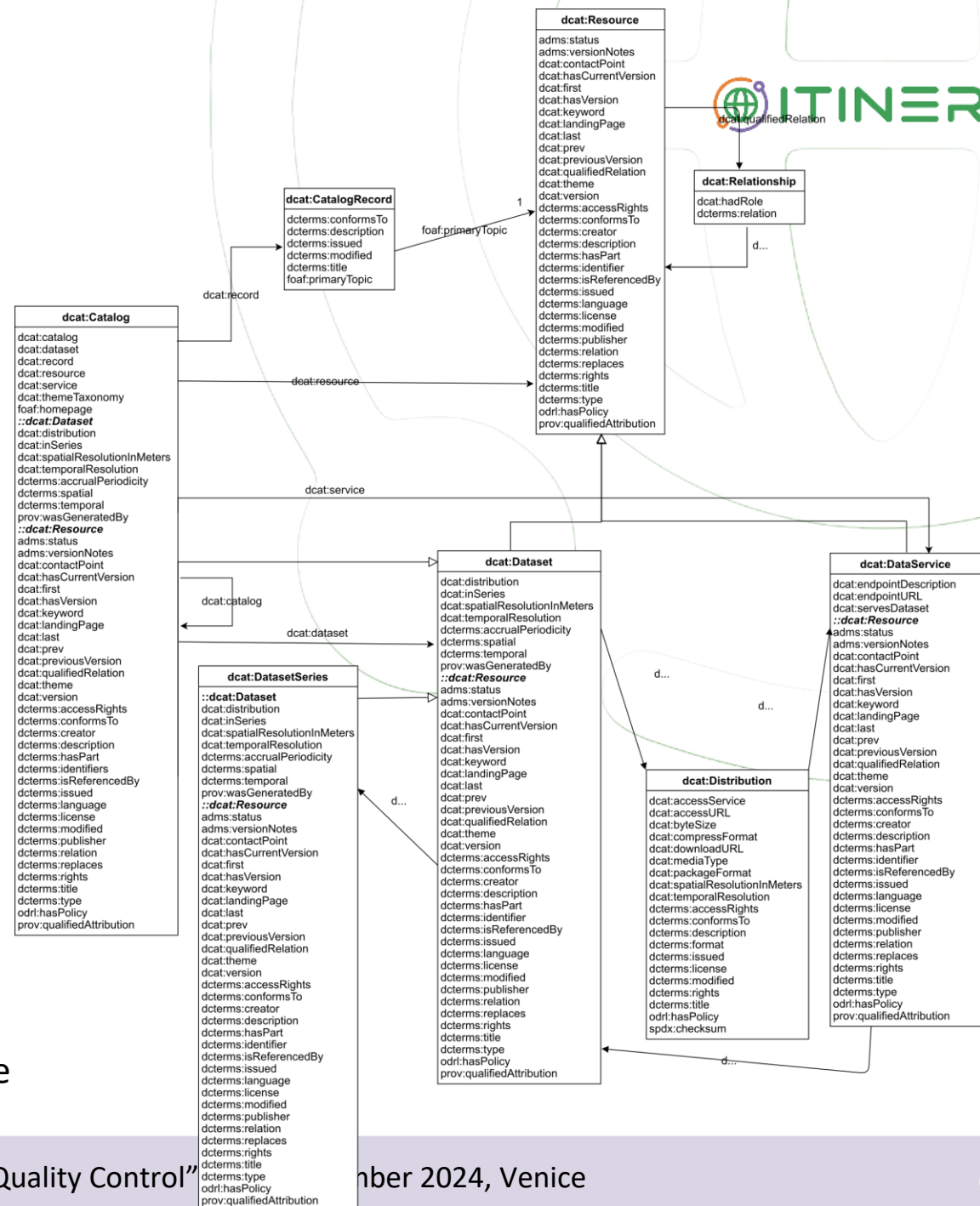
<https://www.w3.org/TR/vocab-dcat-3/images/dcat-all-attributes.svg>

DCAT-AP - Metadata schema



Items identified and described:

- **Resource**
Resource published or curated by a single agent.
- **Catalog**
A curated collection of metadata about resources.
- **Dataset**
A collection of data, published or curated by a single agent, and available for access or download in one or more representations.
- **Dataset Series**
A collection of datasets that are published separately, but share some characteristics that group them.
- **Distribution**
A representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles.
- **Data Service**
A collection of operations that provides access to one or more datasets or data processing functions



GeoDCAT-AP is an extension of DCAT-AP for describing geospatial datasets, dataset series and services

It provides an RDF syntax binding for the union of metadata elements defined in the core profile of ISO 19115:2003 and those defined in the framework of the INSPIRE Directive.

Its basic use case is to make spatial datasets, data series and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors..

For example, it adds things like:

- Geographic coverage of data
- Adoption of standardized vocabularies at INSPIRE level
- Support for the description of data series by allowing data to be grouped into a single entity
- Better mapping of programmatic geographic data access services (APIs)

Latest version 3.0, October 4, 2024

<https://joinup.ec.europa.eu/collection/semic-support-centre/solution/geodcat-application-profile-data-portals-europe>

Description

<https://semiceu.github.io/GeoDCAT-AP/releases/>

EPOS-DCAT-AP

extension to the DCAT-AP standard by EPOS ERIC

The EPOS-DCAT-AP is an extension of the DCAT Application Profile (DCAT-AP) specifically designed for the European Plate Observing System (EPOS).

It aims to represent the diverse and heterogeneous assets in the solid-Earth sciences domain.

It adds elements such as:

- Service
- Equipments
- Facilities
- Publications
- Organisations
- People
- Software

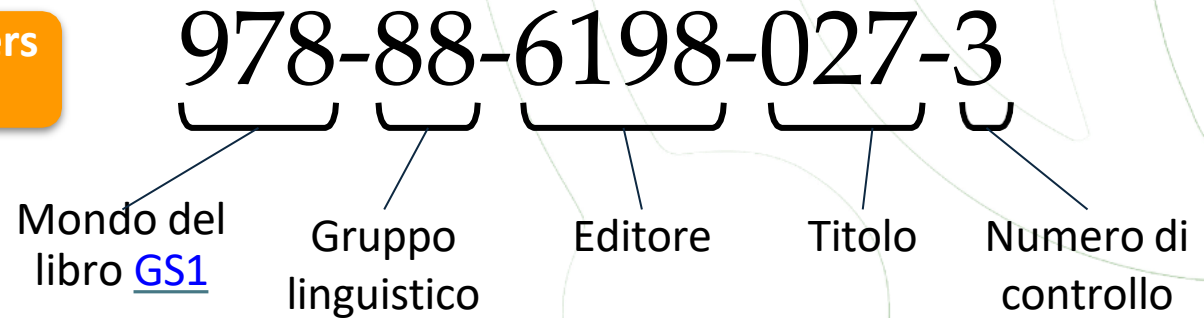


ISBN - International Standard Book Number

ISO 2108 v1, 1970 → v5, 2017

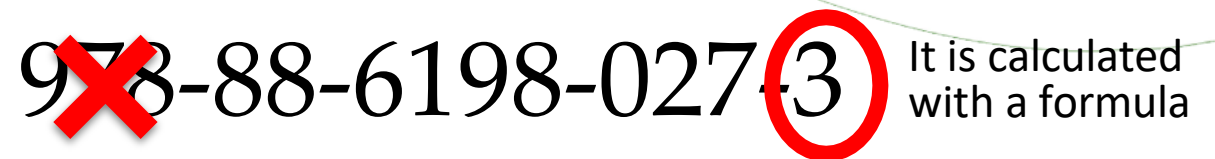
The only organizations that can assign an ISBN are publishers.

13 characters standard



The position of the separators can vary increasing the possible combinations (es. 1 number per linguistic group, 4 numbers for publishers, 4 numbers for titles, 1 check number)

10 characters standard



API documentation to get information: <https://openlibrary.org/dev/docs/api/books>

Examples:

<https://openlibrary.org/isbn/9788877943644.json> (simplified)

<https://openlibrary.org/api/books?jscmd=details format=json&bibkeys=ISBN:9788877943644> (extended)

ISSN - International Standard Serial Number

ISO 3297 v1, 1975 → v7, 2022

Codes [can be purchased](#) upon payment and after verification.

Each type of media type has a dedicated code:

- for printed version (print ISSN, **p-ISSN**)
- for the online version (electronic ISSN, **e-ISSN**).

Each series is assigned a unique code that binds all media (linking ISSN, **I-ISSN**).

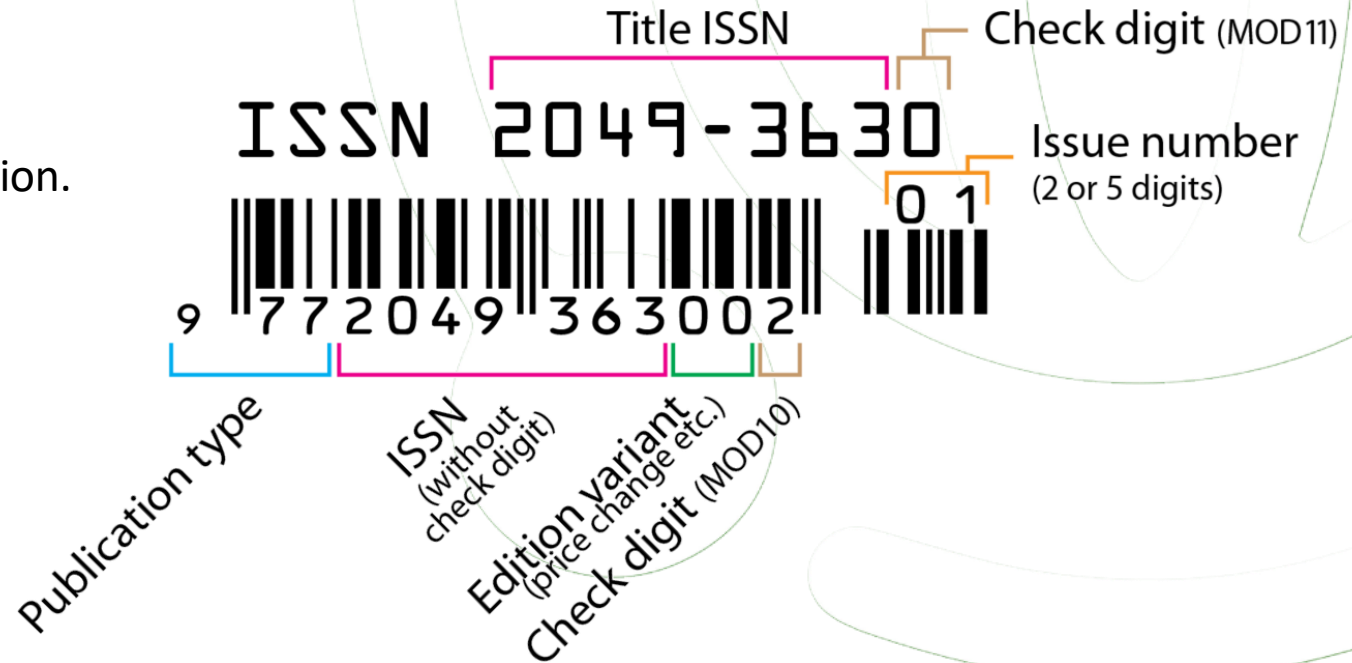
Example: journal “Quaderni di Geofisica” by INGV:
p-ISSN 1590-2595, e-ISSN 2421-5570, I-ISSN 1590-2595.

API description: <https://portal.issn.org/node/169> , output in XML, JSON, JSON-LD, RDF turtle

<https://portal.issn.org/resource/ISSN/1590-2595?format=json> → p-ISSN

<https://portal.issn.org/resource/ISSN/2421-5570?format=json> → e-ISSN

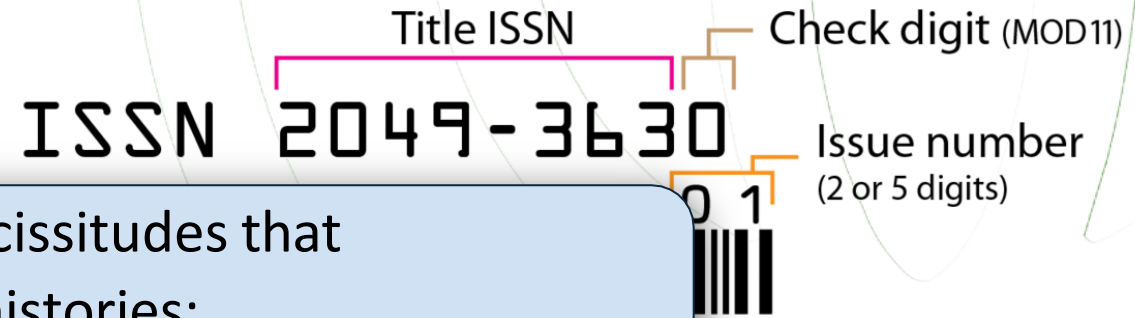
<https://portal.issn.org/resource/ISSN-L/1590-2595?format=json> → I-ISSN



ISSN - International Standard Serial Number

ISO 3297 v1, 1975 → v7, 2022

Codes [can be purchased](#) upon payment and after verification



Long-running journals often have vicissitudes that

can cause a wide spectrum of case histories:

Each type of media type has a dedicated code:

- for printed version (print ISSN, p-ISSN)
- for the online version (electronic ISSN, e-ISSN)

Each series is assigned a unique code that binds all media (linking ISSN, I-ISSN)

- More or less significant change in the title
- Suspension and resumption of issues
- Numbering change
- Change of commercial or intellectual publisher
- Physical Media Change

Example: journal “Quaderni di Geofisica” by INGV.

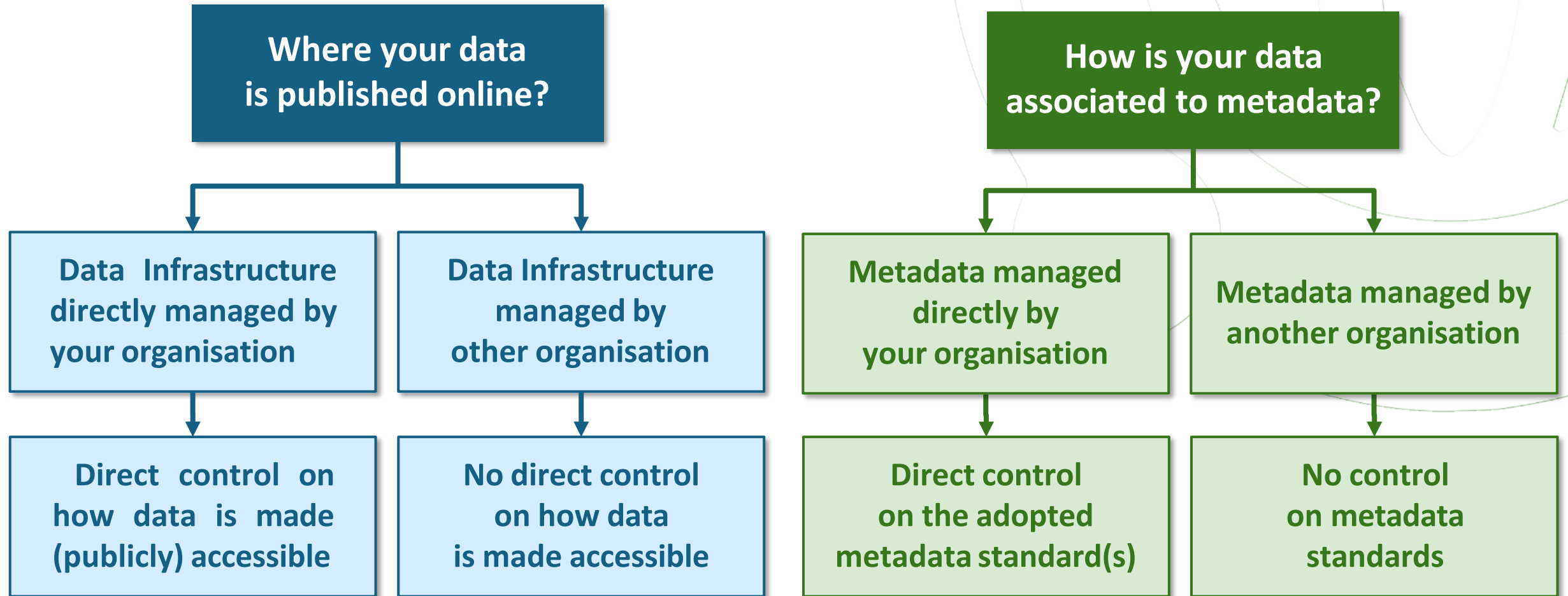
p-ISSN 1590-2595, e-ISSN 2421-5570

API description: <https://portal.issn.org/resource/ISSN/1590-2595?format=json> → p-ISSN

<https://portal.issn.org/resource/ISSN/2421-5570?format=json> → e-ISSN

<https://portal.issn.org/resource/ISSN-L/1590-2595?format=json> → I-ISSN

<https://portal.issn.org/resource/ISSN-L/1590-2595?format=json> → I-ISSN



Archives for Research Outputs

An open-access repository or an open archive is a digital platform that holds research output and provides free, immediate and permanent access to research results for anyone to use, download and distribute.

Thematic or disciplinary repositories

Designed for specific contents, curated by specific international communities (i.e. ArXiv, bioarXiv, Pangaea).

Institutional or national repositories

Managed by single institutions/countries.

Only authors of that institution/country can deposit, everyone can access.

Literature Repositories

Reserved to text deposit (i.e. articles, reports, books, abstracts). Metadata reflect the repository contents.

Data repositories (aka data archives)

Designed to deposit data. Often disciplinary and have specific metadata to describe data they preserve.

Resources: re3data, OpenDOAR e FAIRsharing

Catch-all repositories

All research products can be deposited (i.e. data, literature, presentations, poster, images, software).

Open Data Portals to inform users that data exists

- Web-based interfaces enabling users to find reliable and reusable data
- Catalyst for publication and better-quality data
- Open by definition to public administrations at EU, national and local level
- For administrations, the advantage is to provide public access without the need to reply individual requests for accessing data
- Relies heavily on metadata pointing to external archives where data is stored

Examples

Non-thematic

EU Data Portal, <https://data.europa.eu/>

French Data Portal, <https://www.data.gouv.fr/>

Thematic

EPOS ERIC Data Portal, <https://www.ics-c.epos-eu.org/>

INGV Open Data Portal, <https://data.ingv.it/>

More info at <https://digital-strategy.ec.europa.eu/en/policies/open-data-portals>

Image by Nuthawut Somsuk at <https://www.vecteezy.com/>



How to establish how FAIR is a dataset?

FAIR data principles are only ... principles.

How to quantitatively measure how FAIR is a dataset is still debated.

Any scientific domain put in place its own policies, procedures and tools, and the European Union is trying to set up a common framework.

There are multiple attempts defining their own set of Key Performance Indicators (KPIs).

Multiple tools providing an evaluation of the level of FAIRness of a dataset are being developed, no consensus, yet.

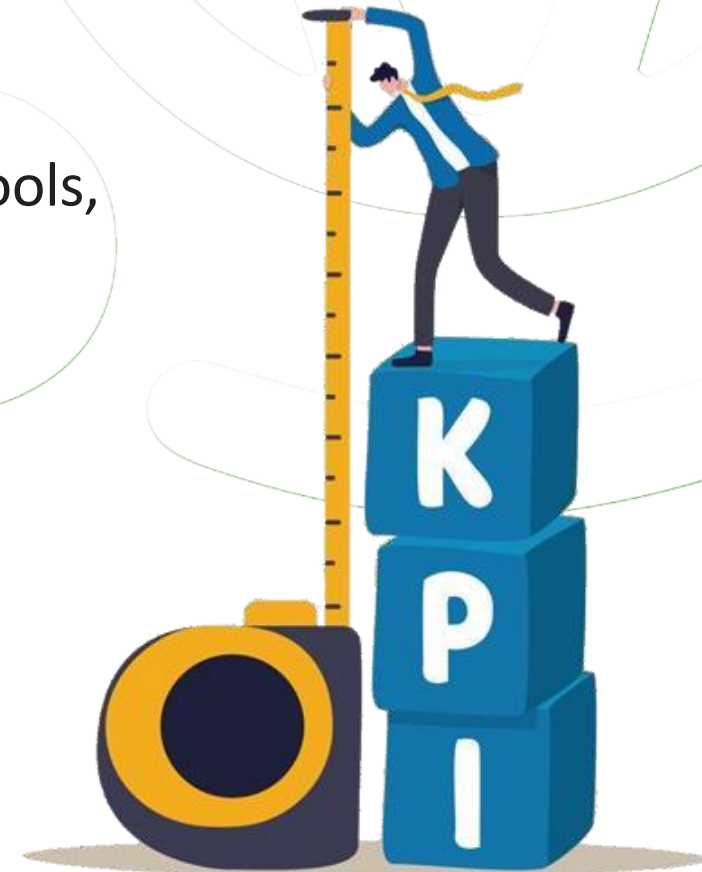
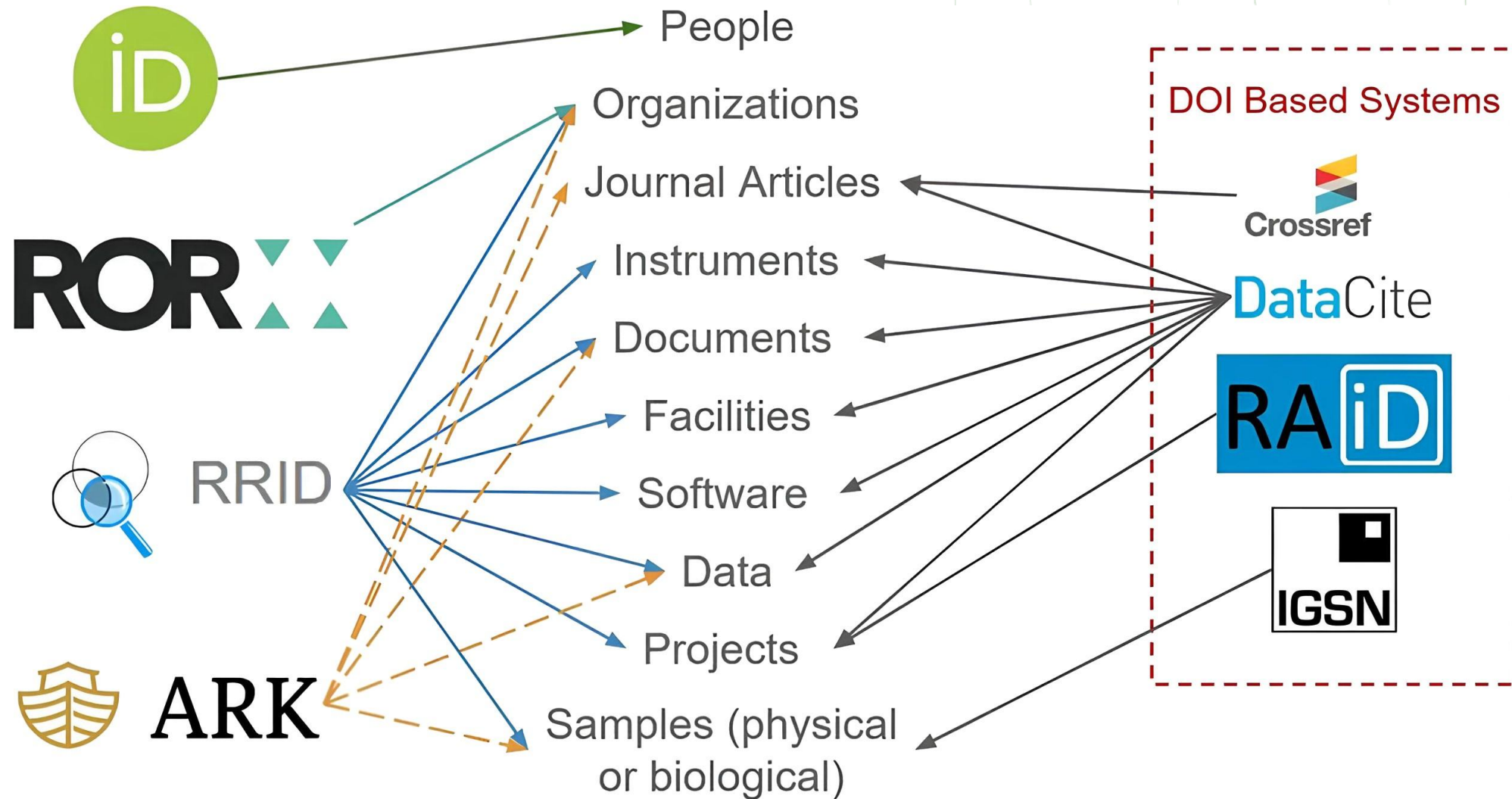


Image by Nuthawut Somsuk at <https://www.vecteezy.com/>

Summary of the PID landscape



F-UJI Automated FAIR Data Assessment

F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics.

<https://www.f-uji.net/>



FAIR assessment

F-UJI is a web service to programmatically assess FAIRness of research data objects (aka data sets) based on metrics developed by the [FAIRsFAIR](#) project.

Please use the form below to enter an identifier (e.g. DOI, URL) of the data set you wish to assess. Optionally you also can enter a metadata service (OAI-PMH, SPARQL, CSW) endpoint URI which F-UJI can use to identify additional information.

Research Data Object (URL/PID):*	Metric:
<input type="text" value="Enter a valid PID or URL of the dataset's landing page (e.g. a DOI)"/>	<input type="text" value="FsF Metrics v0.5"/>
<input type="button" value="▶ Start FAIR Assessment"/>	
Settings	

[About](#) [Feedback](#) [Privacy Policy](#) [Terms of Use](#) [Legal Notice](#)

F-UJI is a result of the [FAIRsFAIR](#) "Fostering FAIR Data Practices In Europe" project which received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 (grant agreement 831558).



Credits

Most slides of this presentation were prepared by me in the framework of

PNRR Research Infrastructure Project “MEET” - Monitoring Earth’s Evolution and Tectonics
Mission 4 “Education and Research”





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”



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca

