Metadata

• Dublin Core

This domain-agnostic metadata standard is one of the best known and most widely used metadata standards.

- DataCite
- CERIF
- DCAT

Metadata

- Dublin Core
- **DataCite**

The DataCite Metadata Schema is a set of mandatory, domain-agnostic metadata that must be registered when receiving a Digital Object Identifier (DOI).

- CERIF
- DCAT

Metadata

- Dublin Core
- DataCite
- **CERIF**

Common European Research Information Format is the standard recommended by the EU to its member states for recording information about research activities.

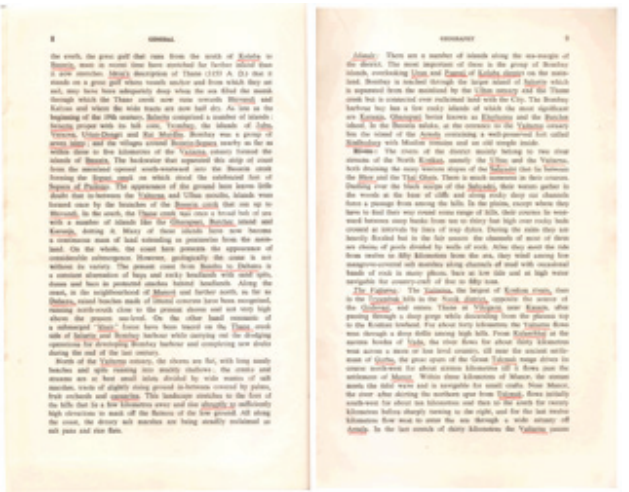
- DCAT

Metadata

- Dublin Core
- DataCite
- CERIF
- **DCAT**

Data Catalog Vocabulary is a metadata standard developed for the description of datasets on the internet which enhances the interoperability between data catalogs.

Knowledge Organisation Systems



• Gazetteer

A gazetteer is a book or a part of a book which lists and describes places. Usually gazetteers contain important information for data managers such as the official spelling of place names.

• Thesaurus

• Glossary

• Authority file

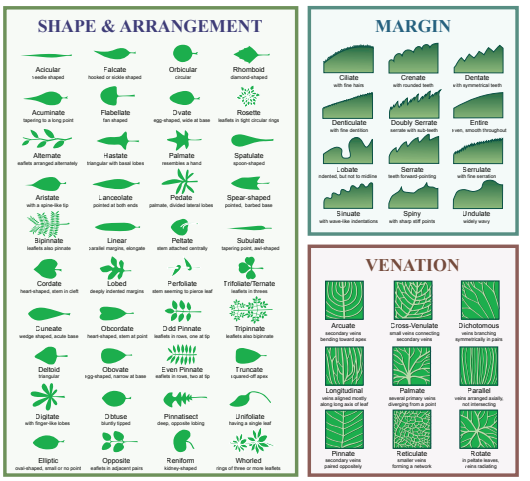
Knowledge Organisation Systems

- Gazetteer
- **Thesaurus**

A thesaurus is a controlled vocabulary including equivalent terms, interrelations and rules of application.

- Glossary
- Authority file

Knowledge Organisation Systems



• Gazetteer

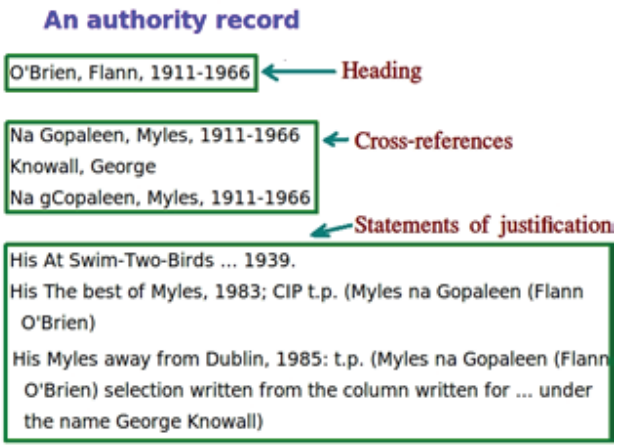
• Thesaurus

• **Glossary**

A glossary is a terminological dictionary which contains a list of designations from a subject field, together with equivalents in one or more languages.

• Authority file

Knowledge Organisation Systems



- Gazetteer
- Thesaurus
- Glossary
- **Authority file**

Authority files are lists of terms that are used to control the variant names of an entity or the domain value of a particular field.

Open Science milestones

- **Berlin Declaration (2003)**

International statement on Open Access and access to knowledge, which was the outcome of a conference hosted by the Max Planck Society.

- The Hague Declaration (2015)
- Amsterdam Call for Action on Open Access (2016)
- Vienna Declaration on EOSC (2018)

Open Science milestones

- Berlin Declaration (2003)
- **The Hague Declaration (2015)**

  Declaration written by 25 global experts to foster agreement on how to best enable access to facts, data and ideas for knowledge discovery in the Digital Age.
- Amsterdam Call for Action on Open Access (2016)
- Vienna Declaration on EOSC (2018)

Open Science milestones

- Berlin Declaration (2003)
- The Hague Declaration (2015)
- **Amsterdam Call for Action on Open Access (2016)**

  Main result of a Dutch Presidency conference on Open Science. It is a living document reflecting the current state of the Open Science evolution.
- Vienna Declaration on EOSC (2018)

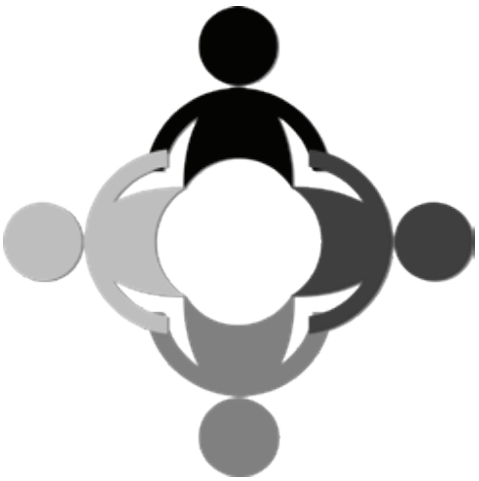
Open Science milestones

EUROPEAN OPEN SCIENCE CLOUD

- Berlin Declaration (2003)
- The Hague Declaration (2015)
- Amsterdam Call for Action on Open Access (2016)
- **Vienna Declaration on EOSC (2018)**

Declaration of ministers, delegates and other participants of the European Open Science Cloud (EOSC) launch event to work together towards realising the potential of the EOSC.

Sensitive personal data



- **Origin**

Data concerning people's racial or ethnic origin.

- Beliefs
- Biomedical data
- Sexual orientation

Sensitive personal data

- Origin
- **Beliefs**

Data concerning people's political, religious or philosophical opinions and beliefs.

- Biomedical data
- Sexual orientation

Sensitive personal data

- Origin
- Beliefs
- **Biomedical data**

  Data concerning personal genetic, biometric and health information.
- Sexual orientation

Sensitive personal data

- Origin
- Beliefs
- Biomedical data
- **Sexual orientation**

Data concerning people's sexual orientation.

RDM
in 4 steps

• Step 1: Establish permissions

Ask permission of those involved, such as participants, data suppliers and the ethics committee.

• Step 2: Document it

• Step 3: Organise storage and access

• Step 4: Up-to-date

RDM
in 4 steps

- Step 1: Establish permissions

- **Step 2: Document it**

Store raw and processed data, metadata and coding, and document the software used to ensure the data can be found, understood and linked to other resources.

- Step 3: Organise storage and access

- Step 4: Up-to-date

RDM
in 4 steps

- Step 1: Establish permissions
- Step 2: Document it
- **Step 3: Organise storage and access**

Archiving the data in a Trustworthy Digital Repository after completing your research ensures sustainable access, including a persistent identifier for citations.

- Step 4: Up-to-date

RDM
in 4 steps

- Step 1: Establish permissions
- Step 2: Document it
- Step 3: Organise storage and access
- **Step 4: Up-to-date**

Keep your data management plan up to date and relevant for everybody involved. Make sure that changes to your project are reflected in your plan.

Benefits of a Data Management Plan

• Future

A DMP helps you plan for the resources, tools and expertise that are required to store, handle and manage the types and volumes of data expected to be collected.

- Easy project management
- Makes data more FAIR
- Clarifies budget requirements

Benefits of a Data Management Plan

- Future

- **Easy project management**

Project management is easier if your DMP includes administrative information, data ownership information, registration numbers for funding, and ethics committee approvals.

- Makes data more FAIR

- Clarifies budget requirements

Benefits of a Data Management Plan

- Future
- Easy project management
- **Makes data more FAIR**

A DMP can help to make data more FAIR, e.g. by using documentation standards, preferred formats for data storage and licenses for reuse.

- Clarifies budget requirements

Benefits of a Data Management Plan

- Future
- Easy project management
- Makes data more FAIR
- **Clarifies budget requirements**

DMPs should define the time and resources needed for careful documentation, as well as for server space, backup solutions, and software.

GDPR Principles

• Lawfulness, fairness, transparency and accuracy

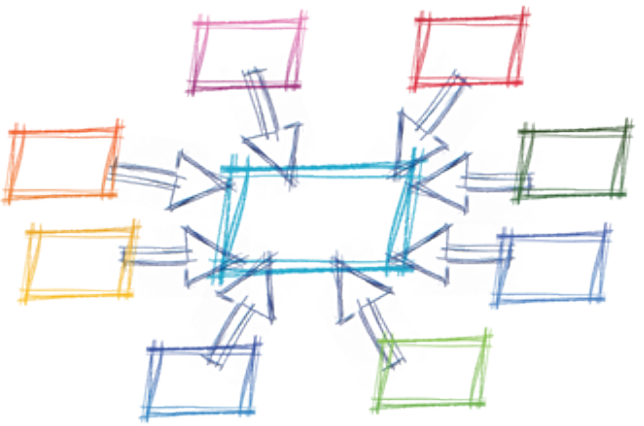
Personal data shall be processed in a lawful, fair and transparent manner and be kept up to date. There are 6 legal grounds for processing personal data lawfully. For science, a common legal ground is consent.

• Purpose limitation and data minimization

• Integrity and confidentiality

• Storage limitation

GDPR Principles



- Lawfulness, fairness, transparency and accuracy

- **Purpose limitation and data minimization**

  There must be a legitimate purpose; data may not be processed for other purposes. Data should be as minimal as possible. For science, further processing is allowed with appropriate safeguards.

- Integrity and confidentiality

- Storage limitation

GDPR Principles

- Lawfulness, fairness, transparency and accuracy
- Purpose limitation and data minimization
- **Integrity and confidentiality**

  Personal data requires appropriate security (e.g. against unauthorized processing) and technical and organisational measures (e.g. encryption and pseudonymization).
- Storage limitation

GDPR Principles

- Lawfulness, fairness, transparency and accuracy
- Purpose limitation and data minimization
- Integrity and confidentiality
- **Storage limitation**

Personal data shall only be stored as long as necessary for the initial purpose. For science, storing personal data for longer periods is allowed if appropriate safeguards are in place.

Persistent Identifiers (PIDs)

- **Digital Object Identifier (DOI)**

A DOI is a PID that ensures digital objects can be permanently found online, regardless of changes to their web address (URL).

- ORCID iD
- Research Organization Registry (ROR-ID)
- PID Graph

Persistent Identifiers (PIDs)

- Digital Object Identifier (DOI)
- **ORCID iD**

An ORCID iD provides a PID for researchers. This helps to bring together different formats of an author's name, and to distinguish between authors sharing the same name.

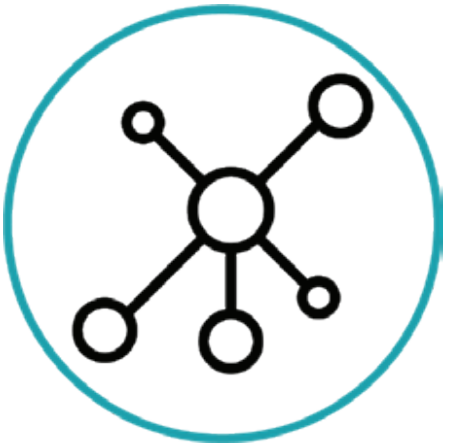
- Research Organization Registry (ROR-ID)
- PID Graph

Persistent Identifiers (PIDs)

- Digital Object Identifier (DOI)
- ORCID iD
- **Research Organization Registry (ROR-ID)**

  ROR provides a PID for organisations in the research community.
- PID Graph

Persistent Identifiers (PIDs)



- Digital Object Identifier (DOI)
- ORCID iD
- Research Organization Registry (ROR-ID)
- **PID Graph**

A PID Graph shows the connection between digital objects through their PIDs, linking two or more digital objects, even if the information is available in different sources.

DANS
Data Stations

• Social Sciences and Humanities

The Data Station SSH offers more than 7000 datasets from the Social Sciences and Humanities, ranging from survey results to archival materials, and from linguistic corpora to interview recordings.

ssh.datastations.nl

- Archaeology
- Life Sciences
- Physical and Technical Sciences

DANS
Data Stations

- Social Sciences and Humanities

- **Archaeology**

The Data Station Archaeology allows you to deposit and search for data in the field of Archaeology.

archaeology.datastations.nl

- Life Sciences

- Physical and Technical Sciences

DANS
Data Stations

- Social Sciences and Humanities
- Archaeology
- **Life Sciences**

The Data Station Life Sciences offers a secure environment to publish and archive data from the Medical, Health and Green Life Sciences.

lifesciences.datastations.nl

- Physical and Technical Sciences

DANS
Data Stations

- Social Sciences andHumanities
- Archaeology
- Life Sciences
- **Physical and Technical Sciences**

The Data Station for the Physical and Technical Sciences offers repository services for the Exact Sciences, in collaboration with 4TU.ResearchData.

phys-techsciences.datastations.nl

Why share data?

- **Accountability**

  Sharing data maximizes transparency and accountability.

- Visibility
- Reuse
- Reproducibility

Why share data?

- Accountability
- **Visibility**

  Sharing data increases the impact and visibility of research results.
- Reuse
- Reproducibility

Why share data?

- Accountability
- Visibility
- **Reuse**

  Sharing data encourages the reuse of data for new research questions.
- Reproducibility

Why share data?

- Accountability
- Visibility
- Reuse
- **Reproducibility**

Sharing data is needed for verification purposes and can enhance the reproducibility of results.

Causes of data loss

- **Hardware failure**

Malfunction of computers and other devices.

- Software malfunction
- Human error
- Malware or hacking

Causes of data loss

- Hardware failure
- **Software malfunction**

  The computer program cannot interpret the data anymore.
- Human error
- Malware or hacking

Causes of data loss

- Hardware failure
- Software malfunction
- **Human error**

The most common cause of data loss is human error, for instance the loss of passwords, unintentional data deletion or wrong usage of software.

- Malware or hacking

Causes of data loss

- Hardware failure
- Software malfunction
- Human error
- **Malware or hacking**

Damage and loss of data caused by malicious software and hackers.

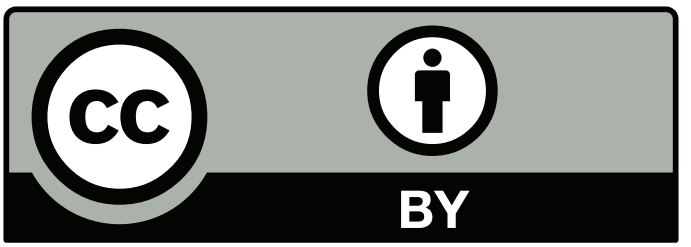
Creative Commons

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

Others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

- CC-BY (Attribution)
- CC-BY-SA (Attribution ShareAlike)
- CC-BY-NC (Attribution NonCommercial)

Creative Commons



- CC-0 (No rights reserved)

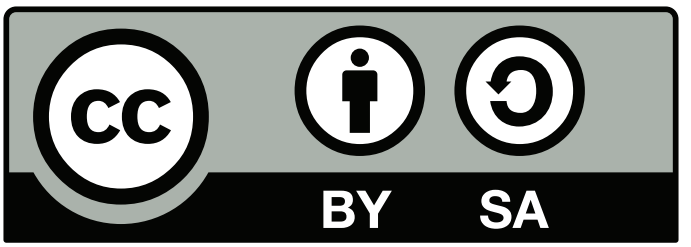
- **CC-BY (Attribution)**

Others may distribute, remix, tweak, and build upon someone's work, even commercially, as long as they credit the author.

- CC-BY-SA (Attribution ShareAlike)

- CC-BY-NC (Attribution NonCommercial)

Creative Commons

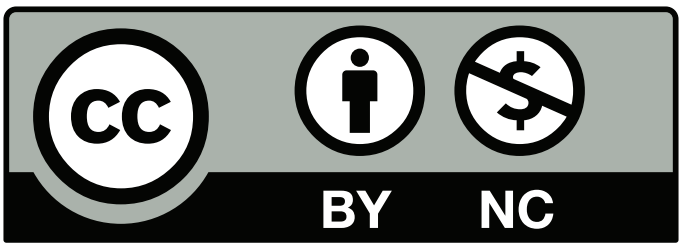


- CC-0 (No rights reserved)
- CC-BY (Attribution)
- **CC-BY-SA (Attribution ShareAlike)**

Others may distribute, remix, tweak, and build upon someone's work, even commercially, as long as they credit the author and distribute derivative works under an identical license.

- CC-BY-NC (Attribution NonCommercial)

Creative Commons



- CC-0 (No rights reserved)
- CC-BY (Attribution)
- CC-BY-SA (Attribution ShareAlike)
- **CC-BY-NC (Attribution NonCommercial)**

Others may remix, tweak, and build upon someone's work, but only non-commercially and as long as they credit the author.

FAIR

• Findable

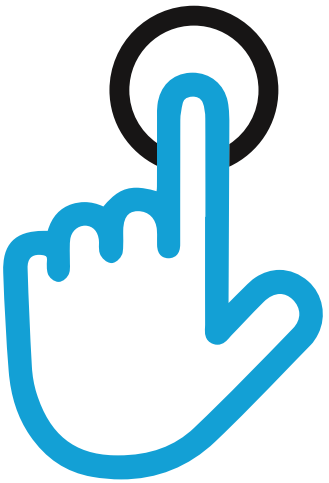
F1: (meta)data are assigned a globally unique and eternally persistent identifier. F2: data are described with rich metadata. F3: (meta)data are registered or indexed in a searchable resource. F4: metadata specify the data identifier.

• Accessible

• Interoperable

• Reuseable

FAIR

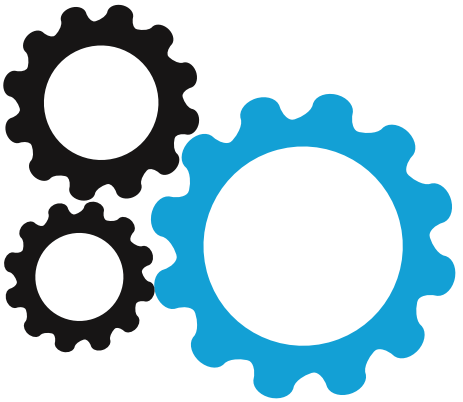


- Findable
- **Accessible**

A1: (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1: the protocol is open, free, and universally implementable. A1.2: the protocol allows for an authentication and authorization procedure. A2: metadata are accessible.

- Interoperable
- Reuseable

FAIR



- Findable
- Accessible
- **Interoperable**

I1: (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation. I2: (meta)data use vocabularies that follow FAIR principles. I3: (meta)data include qualified references to other (meta)data.

- Reuseable

FAIR

- Findable
- Accessible
- Interoperable
- **Reuseable**

R1: meta(data) have multiple accurate and relevant attributes. R1.1: (meta)data are released with a clear and accessible data usage license. R1.2: (meta)data are associated with their provenance. R1.3: (meta)data meet domain-relevant community standards.

History of CoreTrustSeal

- **Data Seal of Approval (DSA)**

The DSA was developed in 2008 by DANS to ensure archived data can be found, understood and used now and in the future.

- RDA Working Group
- CoreTrustSeal
- Sustainability

History of CoreTrustSeal

- Data Seal of Approval (DSA)

- **RDA Working Group**

In 2018, through an RDA Working Group, both the Data Seal of Approval and the ICSU World Data System certification procedures were merged into CoreTrustSeal.

- CoreTrustSeal

- Sustainability

History of CoreTrustSeal

- Data Seal of Approval (DSA)
- RDA Working Group
- **CoreTrustSeal**

CoreTrustSeal is an international quality mark for data repositories focused on sustainability, reliability and accessibility.

- Sustainability

History of CoreTrustSeal

- Data Seal of Approval (DSA)
- RDA Working Group
- CoreTrustSeal
- **Sustainability**

The CoreTrustSeal requirements describe the characteristics needed for certification and are updated every three years with community input.

DANS services

Data Stations

- **Data Stations**

The DANS Data Stations are domain-specific certified repositories where researchers can archive and publish data and find data for reuse.

dans.knaw.nl/en/data-stations/

- DataverseNL
- Data Vault
- Training & Consultancy

DANS services

- Data Stations

- **DataverseNL**

DataverseNL is a Dutch network of institutional data repositories to store, share and publish research data online, during the research period and thereafter.

dans.knaw.nl/en/data-services/dataversenl/

- Data Vault

- Training & Consultancy

DANS services

Data Vault

- Data Stations
- DataverseNL
- **Data Vault**

DANS offers its Data Vault, the facility for long-term preservation of all research data published with DANS, as a separate service to institutional clients.

dans.knaw.nl/en/data-services/data-vault/

- Training & Consultancy

DANS services

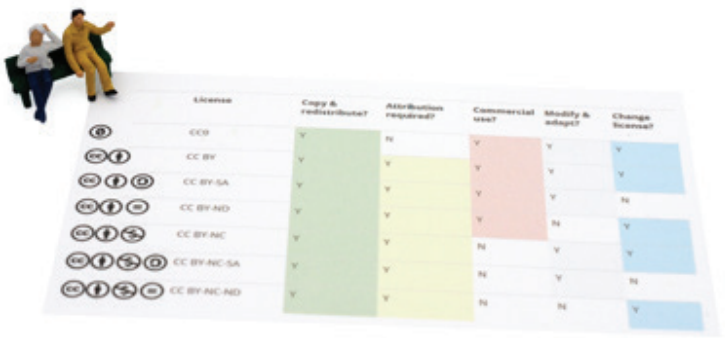
Training & Consultancy

- Data Stations
- DataverseNL
- Data Vault
- **Training & Consultancy**

DANS offers training courses and consultancy services for data professionals and researchers about all aspects of depositing, sharing and reusing research data.

dans.knaw.nl/en/training/

Aspects of sustainable data objects



- **Usage license**

Usage licenses make clear who can use the data under what conditions.

- Metadata
- Data format
- Sustainable storage

Aspects of sustainable data objects

- Usage license
- **Metadata**

Metadata contains essential information about the data object.

- Data format
- Sustainable storage

Aspects of sustainable data objects

- Usage license
- Metadata
- **Data format**

The bits of a file should be formatted according to an (open) future-proof standard in order to be sustainable.

- Sustainable storage

Aspects of sustainable data objects

- Usage licence
- Metadata
- Data format
- **Sustainable storage**

Data objects must be stored for the long-term on robust media.

Dutch networks and communities

- **Open Science Community Netherlands (OSCNL)**

Network of Open Science Communities across the Netherlands, a bottom-up initiative with local communities in university cities across country.

osc-international.com/open-science-community-the-netherlands/

- Reproducibility Network
- Citizen Science Network
- Research Software Community

Dutch networks and communities

- Open Science Community Netherlands (OSCNL)

- **Reproducibility Network**

The Netherlands Reproducibility Network (NLRN) is a national peer-led consortium with the goal to increase the quality and efficiency of research in the Netherlands.

reproducibilitynetwork.nl/

- Citizen Science Network

- Research Software Community

Dutch networks and communities

CS-NL
Citizen Science Nederland

- Open Science Community Netherlands (OSCNL)
- Reproducibility Network
- **Citizen Science Network**

The Dutch community of Citizen Scientists from all walks of life - citizens, community leaders, educators, researchers, funders, policy makers, and anyone involved in Citizen Science.

cs-nl.network/

- Research Software Community

Dutch networks and communities

- Open Science Community Netherlands (OSCNL)
- Reproducibility Network
- Citizen Science Network
- **Research Software Community**

NL-RSE brings together the community of people writing and contributing to research software from Dutch universities, knowledge institutes, companies and other organizations. **nl-rse.org/**

Open Archival Information System (OAIS) functions

- **Ingest**

Ingest function receives information from producers and packages it for storage.

- Preservation Planning / Archival Storage
- Management / Administration
- Access

Open Archival Information System (OAIS) functions

- Ingest

- **Preservation Planning / Archival Storage**

Preservation Planning supports tasks performed to keep archived materials accessible and understandable. Archival Storage concerns the storage, maintenance and dissemination of material.

- Management / Administration

- Access

Open Archival Information System (OAIS) functions

- Ingest
- Preservation Planning / Archival Storage
- **Management / Administration**

Management coordinates the descriptive information of the archive material and the system information that supports the archive. The Administration function manages the daily operation of the archive.

- Access

Open Archival Information System (OAIS) functions

- Ingest
- Preservation Planning / Archival Storage
- Management / Administration
- **Access**

Access includes the user interface and the provision of archive material to users.

European Open Science Cloud (EOSC)

- **Where**

The best place to start discovering the EOSC is through the EOSC Portal

eosc-portal.eu/

- What
- Why
- How

European Open Science Cloud (EOSC)

- Where
- **What**

EOSC aims to be a major European vehicle for helping transform individual research efforts into collective efforts to face global challenges.

- Why
- How

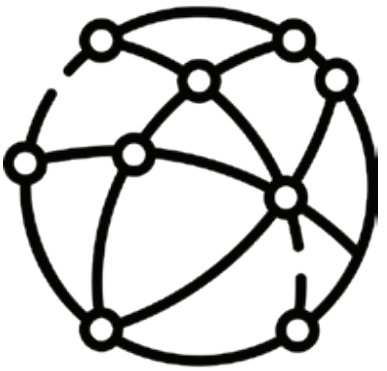
European Open Science Cloud (EOSC)

- Where
- What
- **Why**

EOSC will help deliver Europe’s contribution to the realisation of scientists’, and science’s, potential in the digital age, enhancing Europe’s leadership position in exploiting digital capabilities at the service of science.

- How

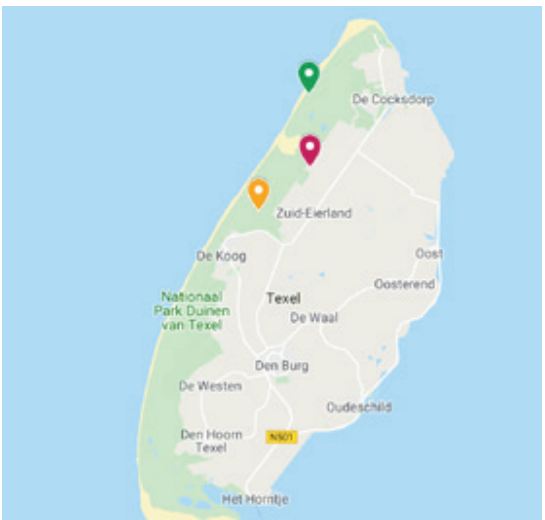
European Open Science Cloud (EOSC)



- Where
- What
- Why
- **How**

Providing an open and trusted environment for accessing and managing a wide range of publicly funded research data and related services and complementary commercial services.

Geographic data

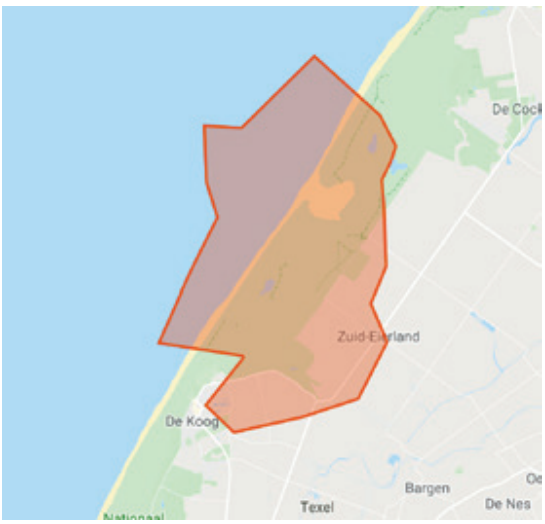


- **Point data**

XY coordinates in a geographic reference system. Point data carry no dimension and can be expressed in degrees, minutes, seconds and smaller units.

- Polygon
- Bounding Box
- Grid cell data

Geographic data



- Point data
- **Polygon**

Connected point data that enclose a certain area. Polygons are two-dimensional and often used to visualise geographical shapes such as city borders or scientific sample areas.

- Bounding Box
- Grid cell data

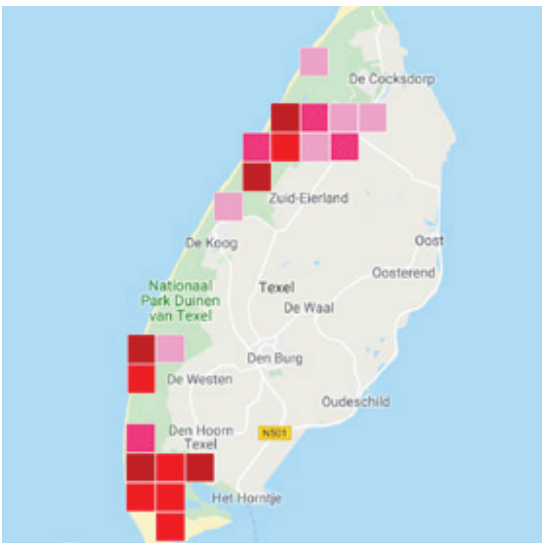
Geographic data

- Point data
- Polygon
- **Bounding Box**

A rectangular geographical area defined by two latitudes and two longitudes. Bounding boxes are easy to generate and often used to indicate the general position of data collections.

- Grid cell data

Geographic data



- Point data
- Polygon
- Bounding Box
- **Grid cell data**

Numbers, densities or concentrations expressed for geographical predefined regular reference structures (e.g. square kilometer). Usually based on a given (national) coordinate reference system.