Funded by
European Union

Grant Agreement 952111
H2020 WIDESPREAD-05-2020 (Twinning)

Standard course – Computer Science
Lesson SC2 – Metadata roles

Lluís Pesquer – CREAF

Co-authors: Ester Prat, Núria Julià (CREAF)

What is metadata?

What is metadata?

- Metadata could be very briefly defined as **“data about data”**.
- It provides information describing the **data**, in our case geographic information data.
- It should include all the **information** necessary to be able to **describe** the cartographic **base** in all its aspects: identification, content description, georeferencing and scope, quality, history, availability and conditions and other more technical information and characteristics.
- It answers the **what**, **when**, **where**, **who**, **how** and **why** questions about the **content** of the data and its **sources**.

Importance of metadata

Importance of metadata

- The importance of metadata is evident at different levels:
- By itself:
  - **Metadata complements** the **data** set → making it more complete and useful
- **Discovery** metadata (descriptive elements)
  - Facilitate the search for cartographic bases (particularly on the Internet)
    - They are the basic support material for data search engines on the Internet
  - They help to maintain and manage the data (efficient use)
    - When metadata is distributed, the layer is advertised, preventing someone from generating it again
- **Exploration** metadata (semantic elements)
  - They allow to know if data is adequate for the purpose pursued
  - They make it easy to choose from several alternative layers without having to examine the data
    - For example from the dates: base date and content date
- **Exploitation** metadata (technical elements of use)
  - They allow to determine how to access, use and store data
  - They help in the transfer of information between different users

Importance of metadata

- They allow searching in metadata catalogs
  - General catalogs: Google, Bing,… cannot index geospatial data directly but they can index metadata
  - Specific catalogs (geospatial): special search tools depending on the type of field:
    - Search by coordinates
    - Search by date (dataset versions)
- What can be searched? What metadata is there?
  - Helpful finders → They should show things completely and in the same way
  - We need standards for metadata:
    - What metadata to document?
      - Mandatory, applicability
    - What is meant by each concept within the metadata. Definitions
      - Allows semantic interoperability
      - E.g. What should the Alternative Title contain: a translation in other languages? A shorter title?

Importance of metadata

What is metadata for → to provide answers to…

What processes have been used?

What is the reference system?

What geographical area do they cover?

Who is the data provider?

What is the thematic quality of the data?

What types of variables, magnitudes, units?

What does the dataset describe?

What is positional accuracy?

How reliable is the data?

When is the information updated?

Main geospatial metadata standards

Metadata standards

CSDGM

Content Standard for Digital Geospatial Metadata: it is the standard by the Federal Geographic Data Committee US

ISO19115

Formal definitions from the main international standardization body for many sectors and purposes

INSPIRE

The directive that establishes the infrastructure for the geospatial information in Europe for supporting environmental policies

National profiles

National geographical agencies adapt their own profile to their Spatial Data Infrastructure (SDI): e.g. NEM at Spain, MGB at Brazil.

ISO metadata standards

- ISO TC211 standards are identified by numbers. There are two types of important metadata standards: conceptual standards and implementation standards.

- **ISO 19115:2003** Geographic information — Metadata
  - https://www.iso.org/standard/26020.html
- **ISO/TS 19139:2007** Geographic information — Metadata — XML schema implementation
  - https://www.iso.org/standard/32557.html

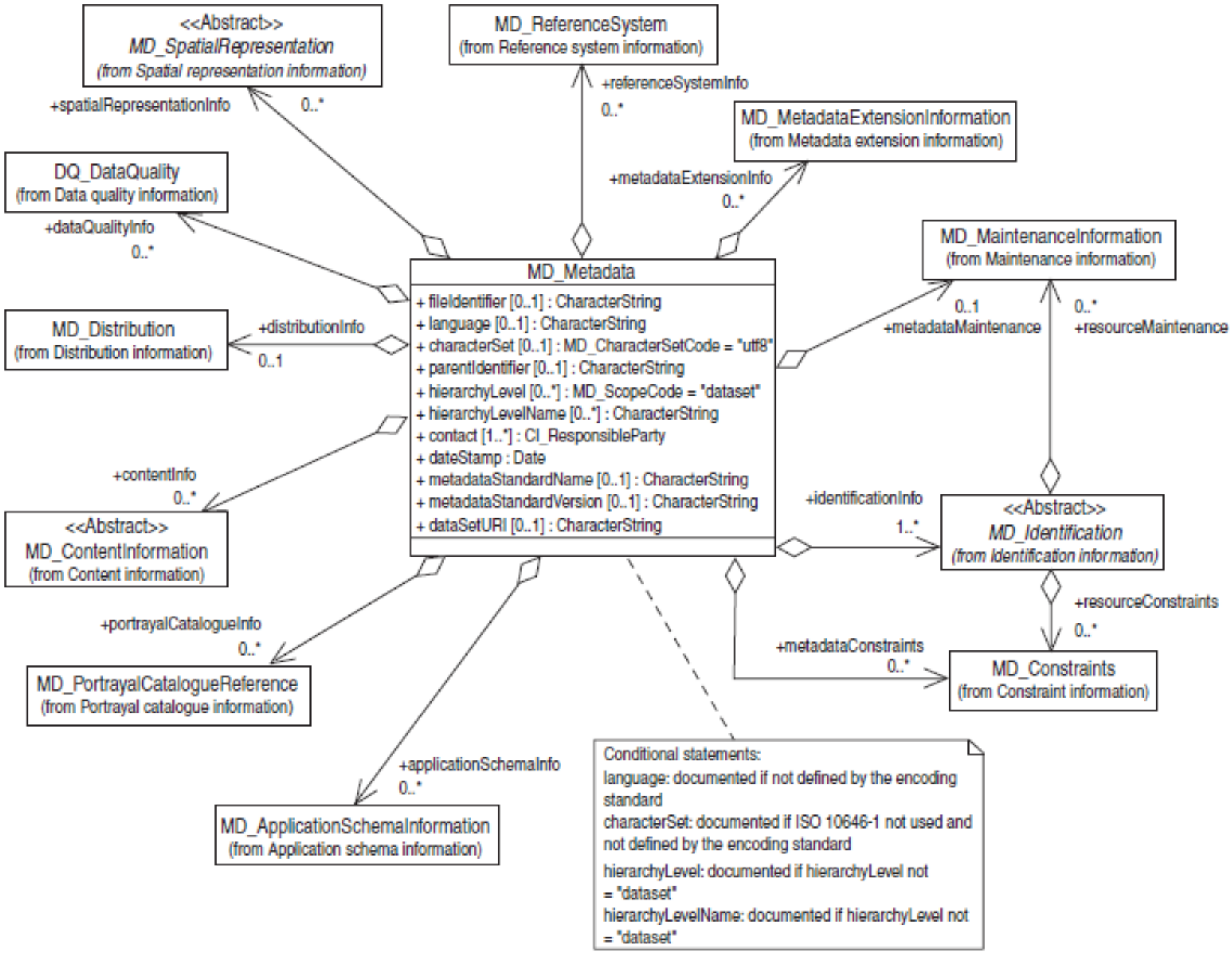
- **ISO 19115-1:2014** Geographic information — Metadata — Part 1: Fundamentals
  - https://www.iso.org/standard/53798.html
- **ISO 19115-2:2019** Geographic information — Metadata — Part 2: Extensions for acquisition and processing
  - https://www.iso.org/standard/67039.html
- **ISO/CD 19115-3** Geographic information — Metadata — Part 3: XML schema implementation for fundamental concepts
  - https://www.iso.org/standard/80874.html

- **ISO 19157:2013** Geographic information — Data quality
  - https://www.iso.org/standard/32575.html
- **ISO/TS 19157-2:2016** Geographic information — Data quality — Part 2: XML schema implementation
  - https://www.iso.org/standard/66197.html

ISO 19115:2003, Geographic information – Metadata

- Defines the **schema** required to **describe geographic information and services**. It provides information on the identification, extent, quality, spatial and temporal pattern, spatial reference, and distribution of digital geographic data.
- It is applicable to:
  - catalogs of cartographic data, distributed services to locate metadata and the complete description of the data sets.
  - the cartographic bases, cartographic series, individual geographical elements and the properties of these elements.
- Defines:
  - **required, conditional, and optional metadata** elements, entities, and sections
  - a minimum set of metadata so that applications using the metadata can function correctly (discovery, determination, access, transfer, use,… of the data) → ***ISO Metadata Core***.
  - optional elements allow a more extensive description of the data

Metadata schema according to ISO 19115, where more than 100 entries are defined



Dataset title (M) (MD_Metadata > MD_DataIdentification.citation > CI_Citation.title)	**Spatial representation type** (O) (MD_Metadata > MD_DataIdentification.spatialRepresentationType)
Dataset reference date (M) (MD_Metadata > MD_DataIdentification.citation > CI_Citation.date)	**Reference system** (O) (MD_Metadata > MD_ReferenceSystem)
Dataset responsible party (O) (MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty)	**Lineage** (O) (MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)
Geographic location of the dataset (by four coordinates or by geographic identifier) (C) (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or EX_GeographicDescription)	**On-line resource** (O) (MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Dataset language (M) (MD_Metadata > MD_DataIdentification.language)	**Metadata file identifier** (O) (MD_Metadata.fileIdentifier)
Dataset character set (C) (MD_Metadata > MD_DataIdentification.characterSet)	**Metadata standard name** (O) (MD_Metadata.metadataStandardName)
Dataset topic category (M) (MD_Metadata > MD_DataIdentification.topicCategory)	**Metadata standard version** (O) (MD_Metadata.metadataStandardVersion)
Spatial resolution of the dataset (O) (MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)	**Metadata language** (C) (MD_Metadata.language)
Abstract describing the dataset (M) (MD_Metadata > MD_DataIdentification.abstract)	**Metadata character set** (C) (MD_Metadata.characterSet)
Distribution format (O) (MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)	**Metadata point of contact** (M) (MD_Metadata.contact > CI_ResponsibleParty)
Additional extent information for the dataset (vertical and temporal) (O) (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent or EX_VerticalExtent)	**Metadata date stamp** (M) (MD_Metadata.dateStamp)

ISO Metadata Core minimum metadata (21 entries):

- Mandatory (M)
- Optional (O)
- Conditional (C): it is required if another entity or element has been documented,

Mandatory (M): The metadata entity or metadata element shall be documented
Conditional (C): The metadata entity or metadata element shall be documented if another entity or element has been documented, or if a condition is or isn't met elsewhere.
Optional (O): Provided to allow users to document their data more fully.

Dataset title (M) A unique title (within your metadata records) for your data.	**Spatial resolution of the dataset** (O) Scale or factor which provides a general understanding of the density of the spatial data in the dataset.	**On-line resource** (O)	**Metadata language** (C) Language used to document the metadata. You must supply the metadata language if it is not defined by the document encoding.
Dataset reference date (M)	**Abstract defining the dataset** (M) Brief narrative summary of the content of the resource.	**Metadata file identifier** (O) Unique identifier for this metadata file	**Metadata character set** (C) Full name of the character encoding used for the metadata set. You must supply this character set in your metadata if you are not using the ISO/IEC 10646-1 character set (http://en.wikipedia.org/wiki/Universal_Character_Set) AND if your character set is not defined by the document encoding. Note as most XML and HTML pages provide a character set as part of their own metadata, it is likely that you will not need to explicitly state this for your own layer metadata
Dataset responsible party (O)	**Distribution format** (O)		
Geographic location of the dataset (by four coordinates or by geographic identifier) (C) If the metadata applies to a data set which is spatially referenced this is required.	**Additional extent information for the dataset** (vertical and temporal) (O)		
Dataset language (M) Language(s) used within the dataset. Required even if the resource does not include any textual information; defaults to the Metadata language.	**Spatial representation type** (O) The method used to represent geographic information in the dataset. i.e., vector, grid, TIN etc.	**Metadata standard name** (O) Name of the metadata standard (including profile name) used	**Metadata point of contact** (M) Party responsible for the metadata information
Dataset character set (C) Full name of the character encoding used for the data set. You must supply this character set if you are not using the ISO/IEC 10646-1 character set and if your character set is not defined by the document encoding.	**Reference system** (O)		
Dataset topic category (M) Main theme(s) of the data set described using the most appropriate term defined in the standard.	**Lineage** (O)	**Metadata standard version** (O) Version (profile) of the metadata standard used	**Metadata date stamp** (M)

Source: Adapted from https://www.onegeology.org/wmscookbook/2_7.html

ISO/TS 19139:2007 Geographic information — Metadata — XML schema implementation

- ISO/TS 19139:2007 defines an XML geographic metadata encoding (gmd: Geographic MetaData XML) which is an XML implementation schema derived from the concepts and elements defined in ISO 19115.

ISO 19115 is **conceptual** and **ISO 19139** is about **implementation**

What is an SDI?

What is an SDI?

- In essence:
  - Spatial Data Infrastructure (SDI) is a set of data and **geographic information** configured and **structured** according to **rules and standards** in order to make it **compatible** for as many users as possible.

Source: Xavier Pons; Anna Arcalís. Diccionari terminològic de teledetecció.

  - The term Spatial Data Infrastructure is often used to denote the relevant base collection of **technologies, policies and institutional arrangements** that facilitate the **availability of and access to spatial data**. The SDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general.

Source: GSDI Cookbook (2012): http://gsdiassociation.org/index.php/publications/sdi-cookbooks.html.

What is an SDI?

- In a broader sense :
  - **Cooperation framework** that aims to facilitate the knowledge, access and use of **geographical information** available in an area through the Internet, by means of **geoportals** that offer data and metadata catalog services; visualization, through **map services**; location, by means of addresses and nomenclatures, and, eventually, other types of specific geoservices, in addition to download of data, documents or other geospatial information resources.

Source: Xavier Pons; Anna Arcalís. *Diccionari terminològic de teledetecció*.

  - A set of **technological and non-technological set-ups within and between organizations to facilitate access, sharing and use of spatial data**, thereby contributing to the enhanced performance of the business, policy making and service provision processes.

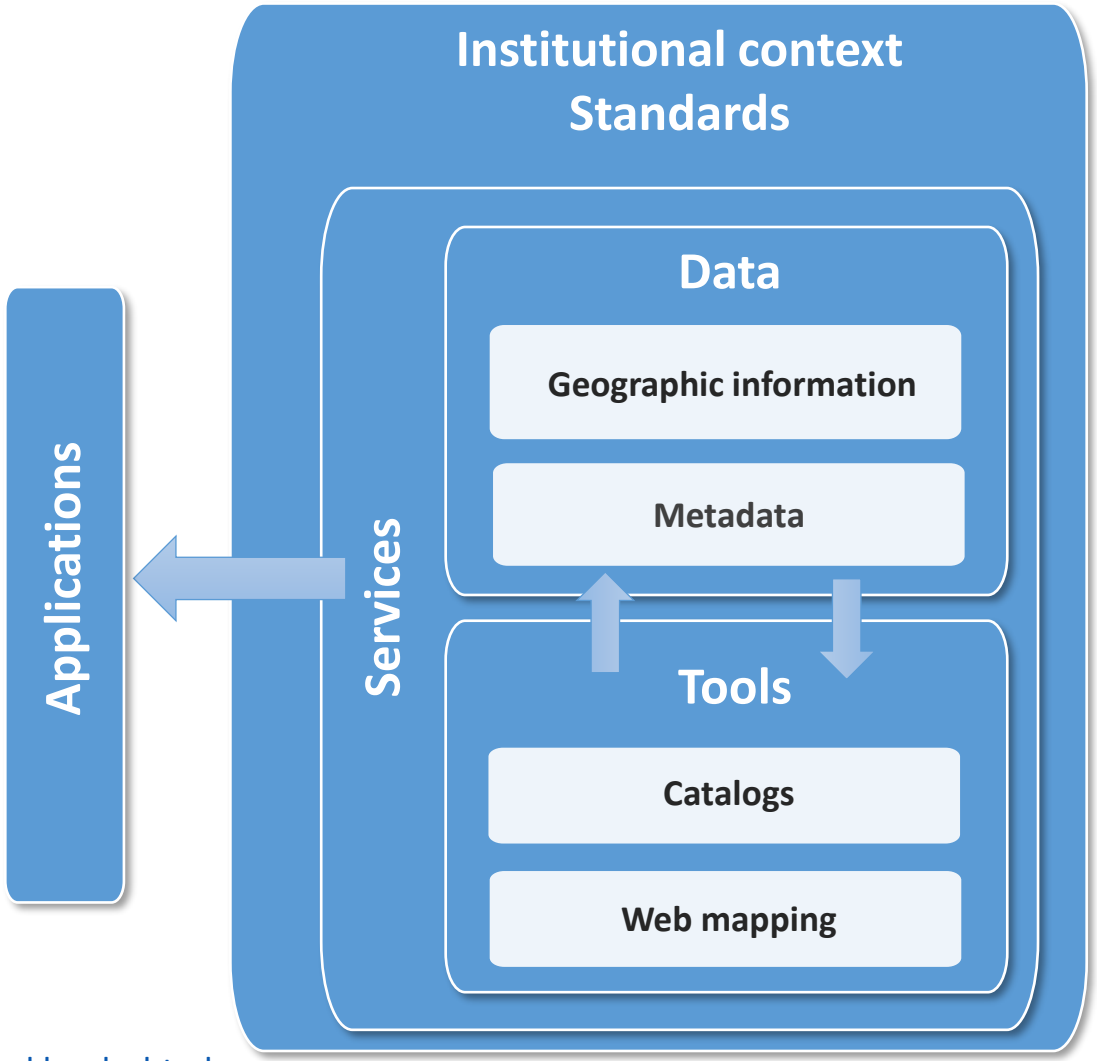
Source: Marc Olijslagers, Danny Vandenbroucke, Maria da Saudade De Brito Pontes, SADL – KU Leuven. Introduction to SDI Architecture and Components. EO4GEO Lecture. http://www.eo4geo.eu/training/introduction-to-sdi-architecture-and-components/

SDI components

An SDI must be more than a single data set or database; an SDI hosts **geographic data** and attributes, sufficient documentation (**metadata**), a means to discover, visualize, and evaluate the data (**catalogues and Web mapping**), and some method to provide access to the geographic data.

To make an SDI functional, it must also include the **organizational agreements** needed to coordinate and administer it on a local, regional, national, and or trans-national scale.

The infrastructure provides the ideal environment to connect applications to data – influencing both data collection and applications construction through minimal appropriate **standards and policies**.



Source: GSDI Cookbook (2012): http://gsdiassociation.org/index.php/publications/sdi-cookbooks.html

Examples of SDI

- Global Earth Observation System of Systems (GEOSS)
- Infrastructure for Spatial Information in the European Community (INSPIRE)
- Chilean Spatial Data Infrastructure
- The United Nations Spatial Data Infrastructure (UNSDI)
- Canadian Geospatial Data Infrastructure
- German National Spatial Data Infrastructure
- Dutch Geo register
- Spanish National Spatial Data Infrastructure (IDEE)

INSPIRE international framework

The INSPIRE European Directive

- The INSPIRE Directive aims to create a **European Union Spatial Data Infrastructure** for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. This European Spatial Data Infrastructure will enable the **sharing of environmental spatial information** among public sector organisations, facilitate public access to spatial information across Europe and assist in policy-making across boundaries.

Directive 2007/2/CE, *Infrastructure for Spatial Information in Europe*

https://inspire.ec.europa.eu/

INSPIRE Directive 2007/2/CE

- Defines a set of general and mandatory rules for the establishment of a Spatial Information Infrastructure in the European Union based on the Infrastructures of the Member States (SDI = Spatial Data Infrastructure)

- The Directive came into force on **15 May 2007** and has been implemented in various stages, with full implementation required by **2021**.

- INSPIRE is based on the infrastructures for spatial information established and operated by the **Member States** of the European Union. The Directive addresses **34 spatial data themes** needed for environmental applications.

INSPIRE Themes

ANNEX: 1

- Addresses
- Cadastral parcels
- Geographical grid systems
- Hydrography
- Transport networks
- Administrative units
- Coordinate reference systems
- Geographical names
- Protected sites

ANNEX: 2

- Elevation
- Land cover
- Geology
- Orthoimagery

ANNEX: 3

- Agricultural and aquaculture facilities
- Atmospheric conditions
- Buildings
- Environmental monitoring Facilities
- Human health and safety
- Meteorological geographical features
- Natural risk zones
- Population distribution and demography
- Sea regions
- Species distribution
- Utility and governmental services
- Area management / restriction / regulation zones & reporting units
- Bio-geographical regions
- Energy Resources
- Habitats and biotopes
- Land use
- Mineral Resources
- Oceanographic geographical features
- Production and industrial facilities
- Soil
- Statistical units

https://inspire-geoportal.ec.europa.eu/

- More than 150 thousand datasets are documented and increasingly made available (i.e., discoverable, viewable and downloadable) within the pan-European INSPIRE infrastructure.

Priority Data Sets Viewer

The application displays the availability and provides access to the selected priority data sets used for environmental reporting. It allows filtering by environmental domain, environmental legislation and country, as well as individual priority data sets.

Browse Priority Data Sets

INSPIRE Thematic Viewer

The application displays the availability and provides access to all EU MS data sets falling under the scope of INSPIRE Directive filtered by data themes and countries (i.e. Annex I, II and III).

Browse INSPIRE Thematic Data Sets

Metadata in INSPIRE

- To ensure that the SDI of the Member States are **compatible** and **interoperable** in a community and cross-border context, the Directive requires the adoption of **specific common Implementing Standards** for data, metadata and services.
- According to Article 5(1) of INSPIRE Directive 2007/2/EC, Member States shall ensure that metadata are created for the spatial data sets and services corresponding to the themes listed in Annexes I, II and III, and that those metadata are kept up to date.
- The Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007 set out the requirements for the creation and maintenance of this metadata.

Creation and edition of metadata

Editing metadata

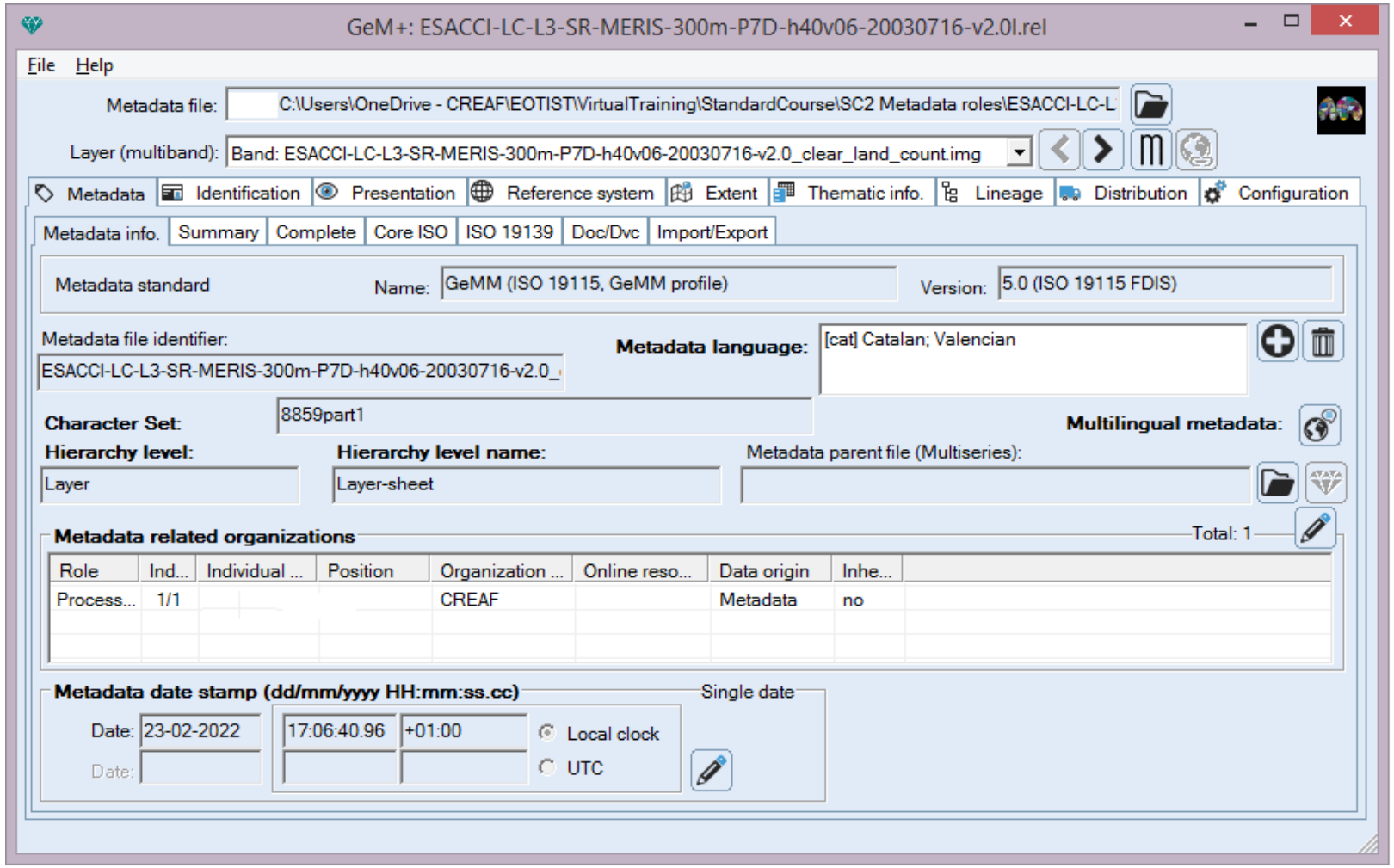
- And how do we edit metadata in XML ISO 19139?
- Using a text editor?

NDVI_LT0518802620050902_final.xml

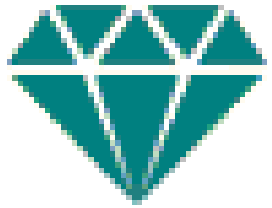
```
<?xml version="1.0" encoding="ISO-8859-1"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gmi="http://www.isotc211.org/2005/gmi" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    <gmd:fileIdentifier>
        <gco:CharacterString>NDVI_LT0518802620050902_finalI_54078</gco:CharacterString>
    </gmd:fileIdentifier>
    <!-- 2.2.11.3 Metadata Language -->
    <gmd:language>
        <gmd:LanguageCode codeList="http://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#LanguageCode" codeListValue="eng">English</gmd:LanguageC
    </gmd:language>
    <gmd:characterSet>
        <gmd:MD_CharacterSetCode codeList="http://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#MD_CharacterSetCode" codeListValue="8859part1">88
    </gmd:characterSet>
    <!-- 2.2.1.3 Resource type -->
    <gmd:hierarchyLevel>
        <gmd:MD_ScopeCode codeList="http://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#MD_ScopeCode" codeListValue="dataset">Layer</gmd:MD_Scop
    </gmd:hierarchyLevel>
<gmd:hierarchyLevelName>
        <gco:CharacterString>Layer-sheet</gco:CharacterString>
    </gmd:hierarchyLevelName>
    <!-- 2.2.11.1 Metadata Contact -->
    <gmd:contact>
        <gmd:CI_ResponsibleParty>
            <gmd:individualName>
                <gco:CharacterString>Alaitz Zabala</gco:CharacterString>
            </gmd:individualName>
            <gmd:organisationName>
                <gco:CharacterString>Geography Department - Universitat Autònoma de Barcelona</gco:CharacterString>
            </gmd:organisationName>
            <gmd:contactInfo>
                <gmd:CI_Contact>
                    <gmd:address>
                        <gmd:CI_Address>
                            <gmd:electronicMailAddress>
```

Editing metadata

- Or, using a specific tool?



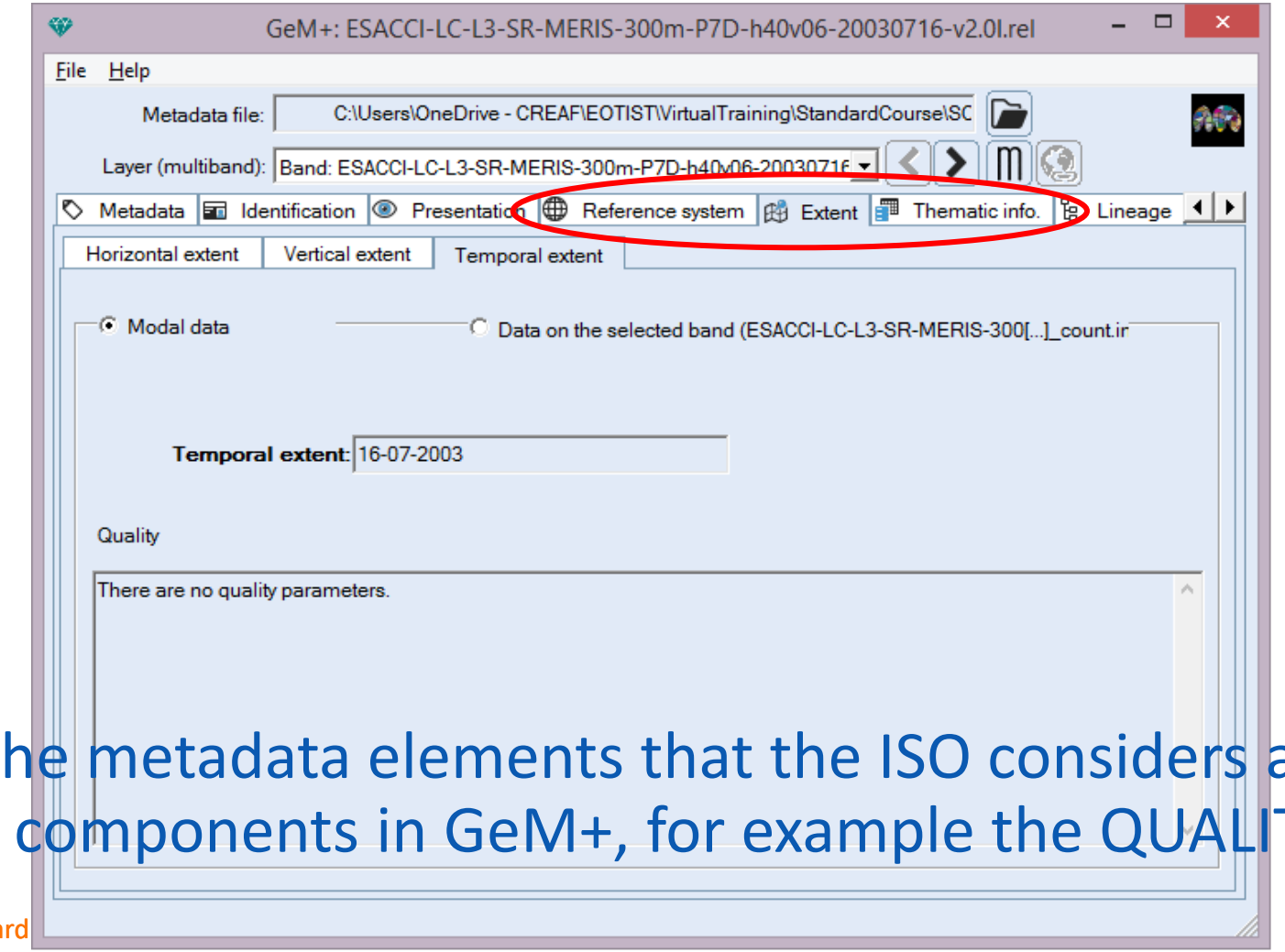
GeM+: Universal Geospatial Metadata Manager



- **Desktop** application for **Windows** platforms
- **Free** and freely to use
- It can be installed and run from MiraMon itself (Tools | Metadata Manager) or independently
- It allows to create, view, edit and export cartographic database metadata in various formats, including:
  - XML compliant with ISO 19115-19139
  - XML compliant with the INSPIRE Directive
- **Multilingual** editing of metadata
- It incorporates various keyword **dictionaries**, including GEMET, Themes INSPIRE,...
- Direct **connection** to the **data** and automatic extraction of metadata from the data
- https://www.miramon.cat/GeMPlus/ESP/index.htm

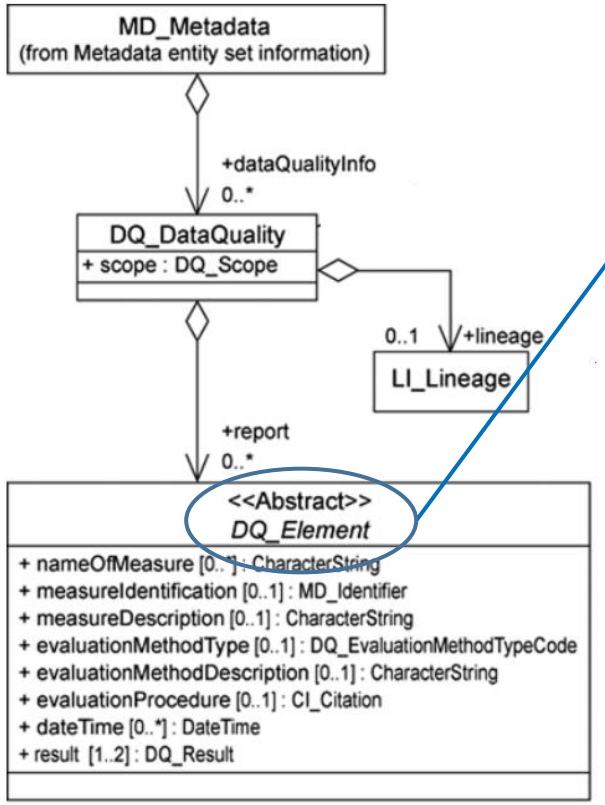
Organization of the information in GeM+

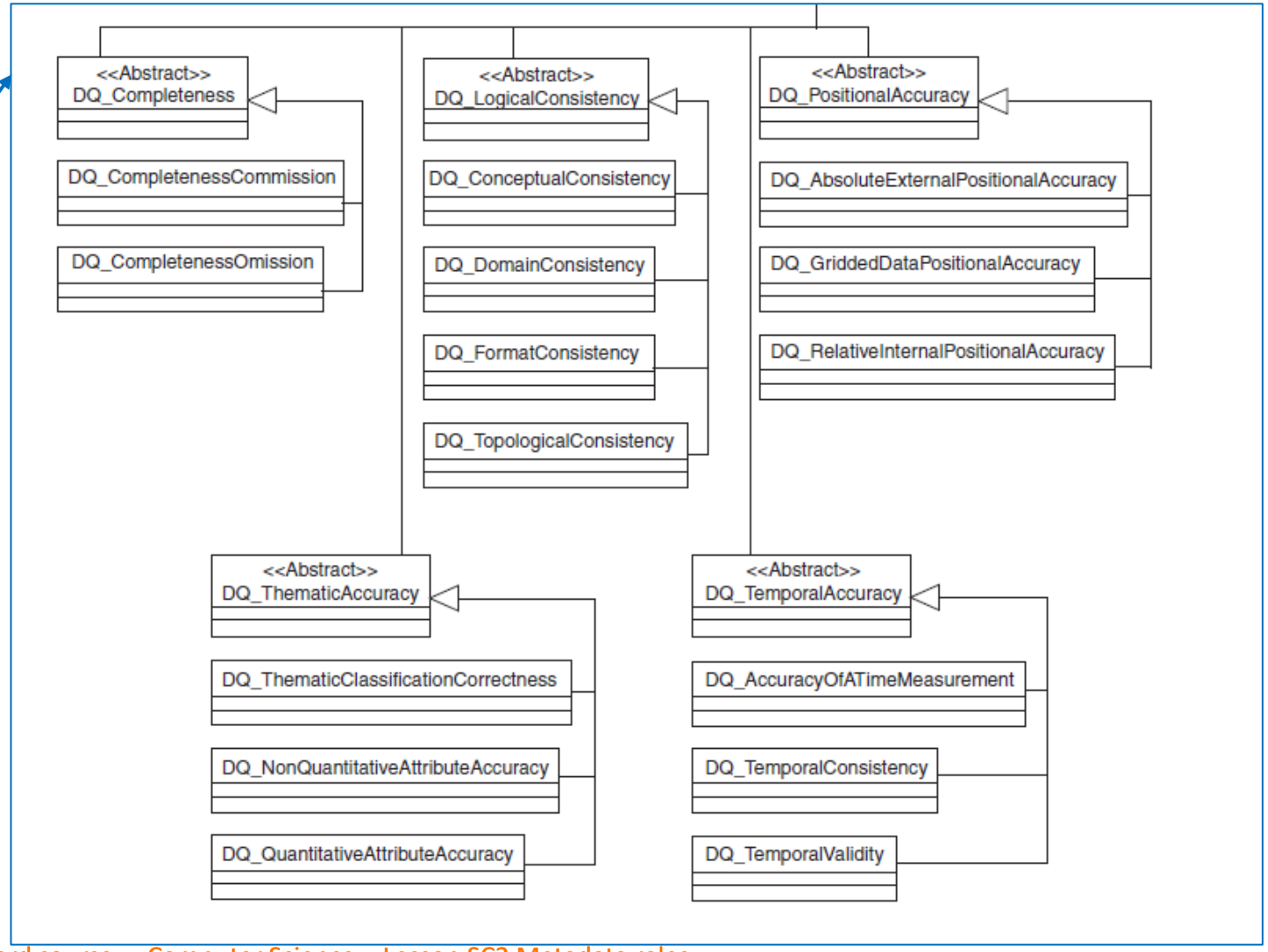
- It is based on the spatial information components:
  - Spatial Component
    - Reference system
    - Spatial extension
  - Temporal Component
    - Temporal extension
  - Thematic Component
    - Thematic information



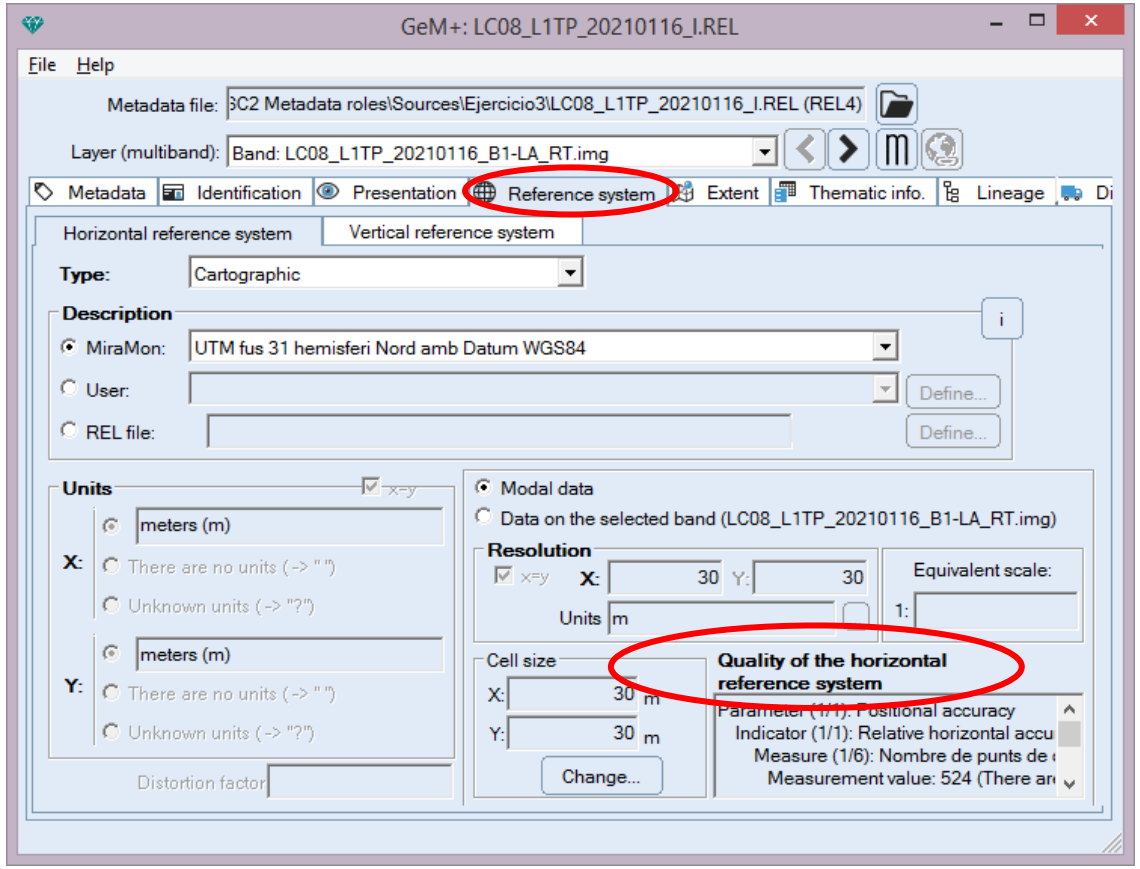
- For this reason, some of the metadata elements that the ISO considers a package are separated by components in GeM+, for example the QUALITY

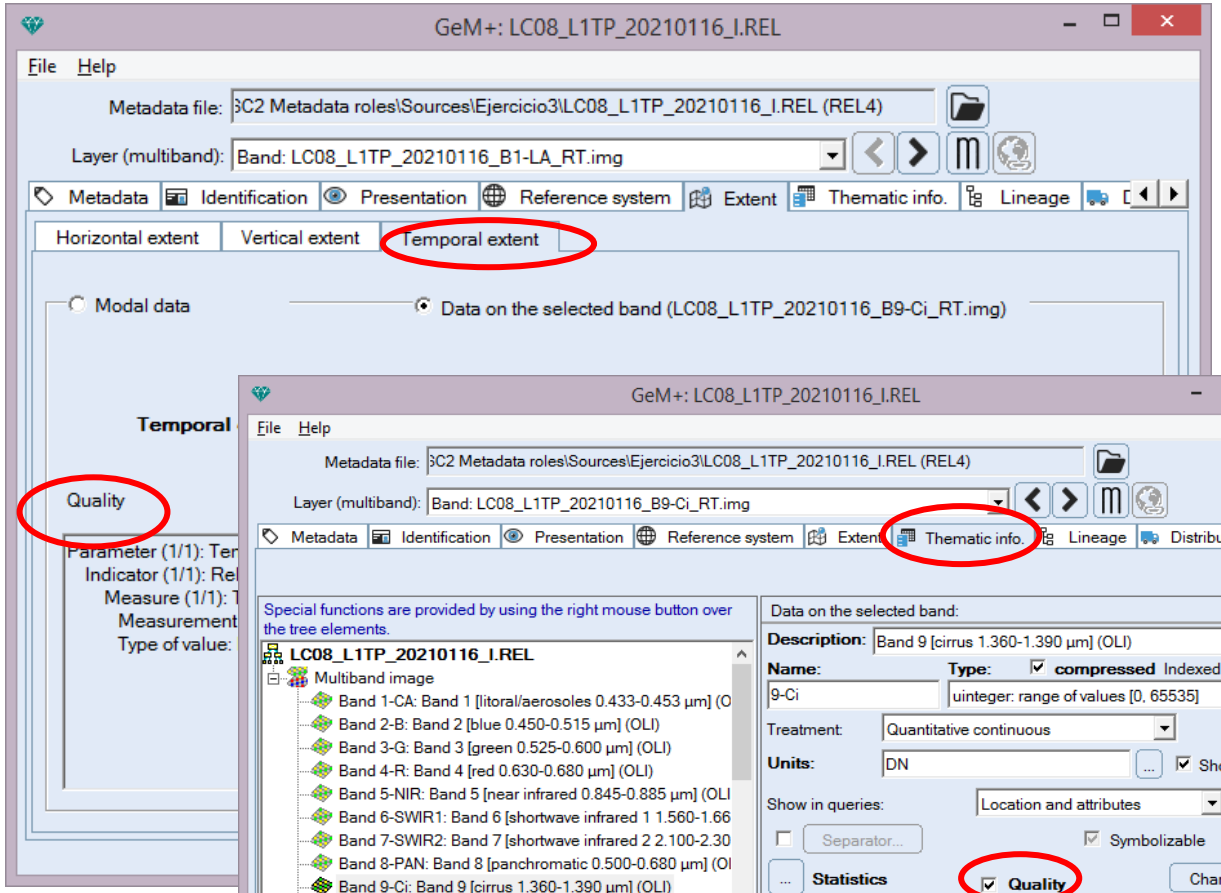
Quality elements according to ISO





Quality elements in GeM+





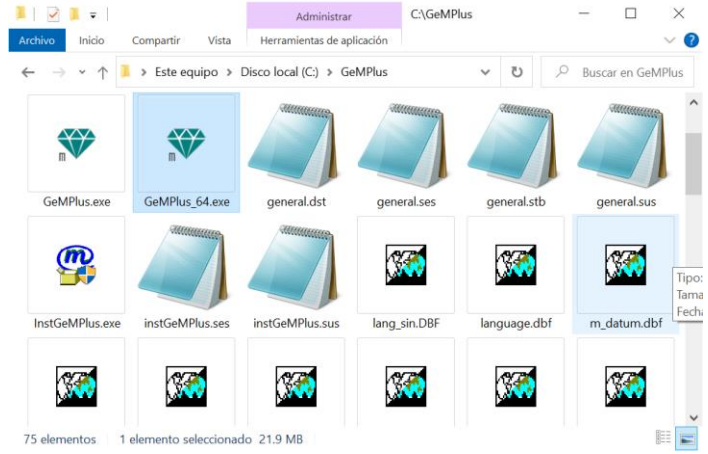
How to get GeM+

- Access the MiraMon website https://www.miramon.cat/Index_usa.htm or directly the GeM+ download website https://www.miramon.cat/USA/Prod-GeMPlus.htm.

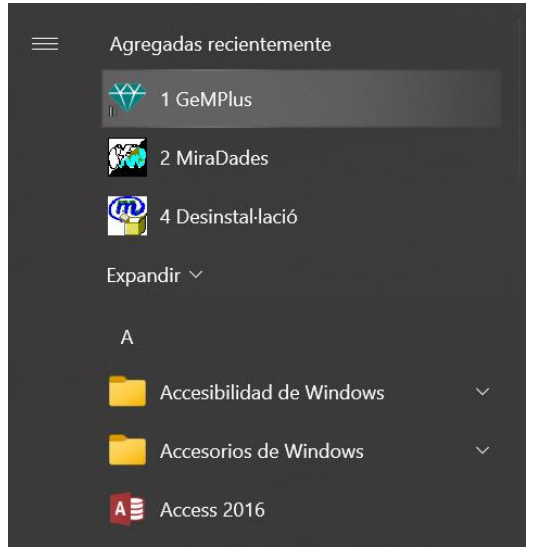
- Download the latest version:
  - From the installer: Download the InstGeMPlus.exe file in a local folder and run it following the steps indicated by the installer.
  - In ZIP format: Download the GeMPlus.zip file in a local folder and unzip it in a folder for example c:\GeMPlus. This option does not require installation and does not register extensions.

How to run GeM+

- Double click on the file GeMPlus_64.exe (or GeMPlus.exe on 32-bit systems) in the corresponding folder (typically c:\GeMplus)

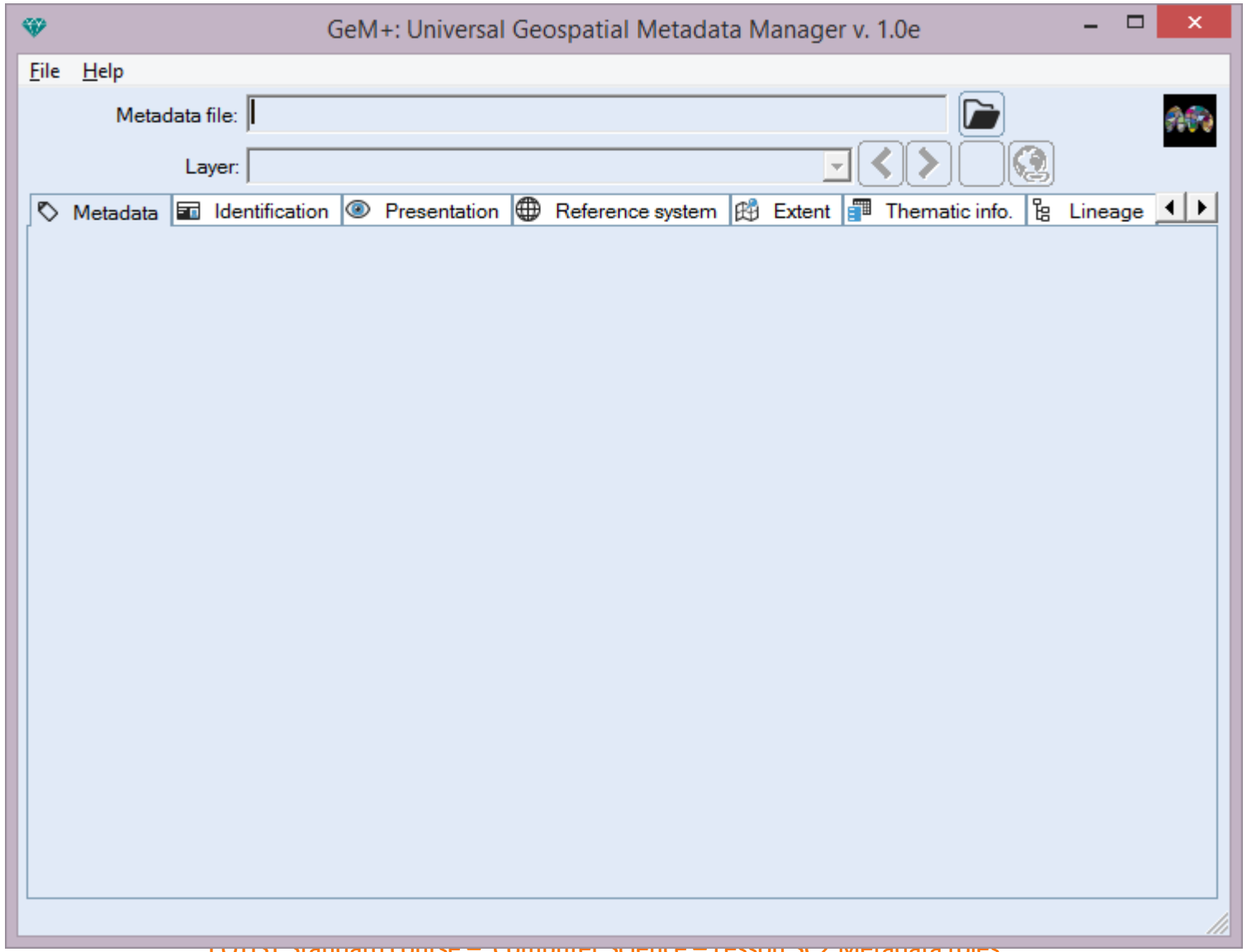


- Or click on GeMPlus from the menu



- Or double click on the shortcut to GeMPlus from the desktop

GeM+ first run



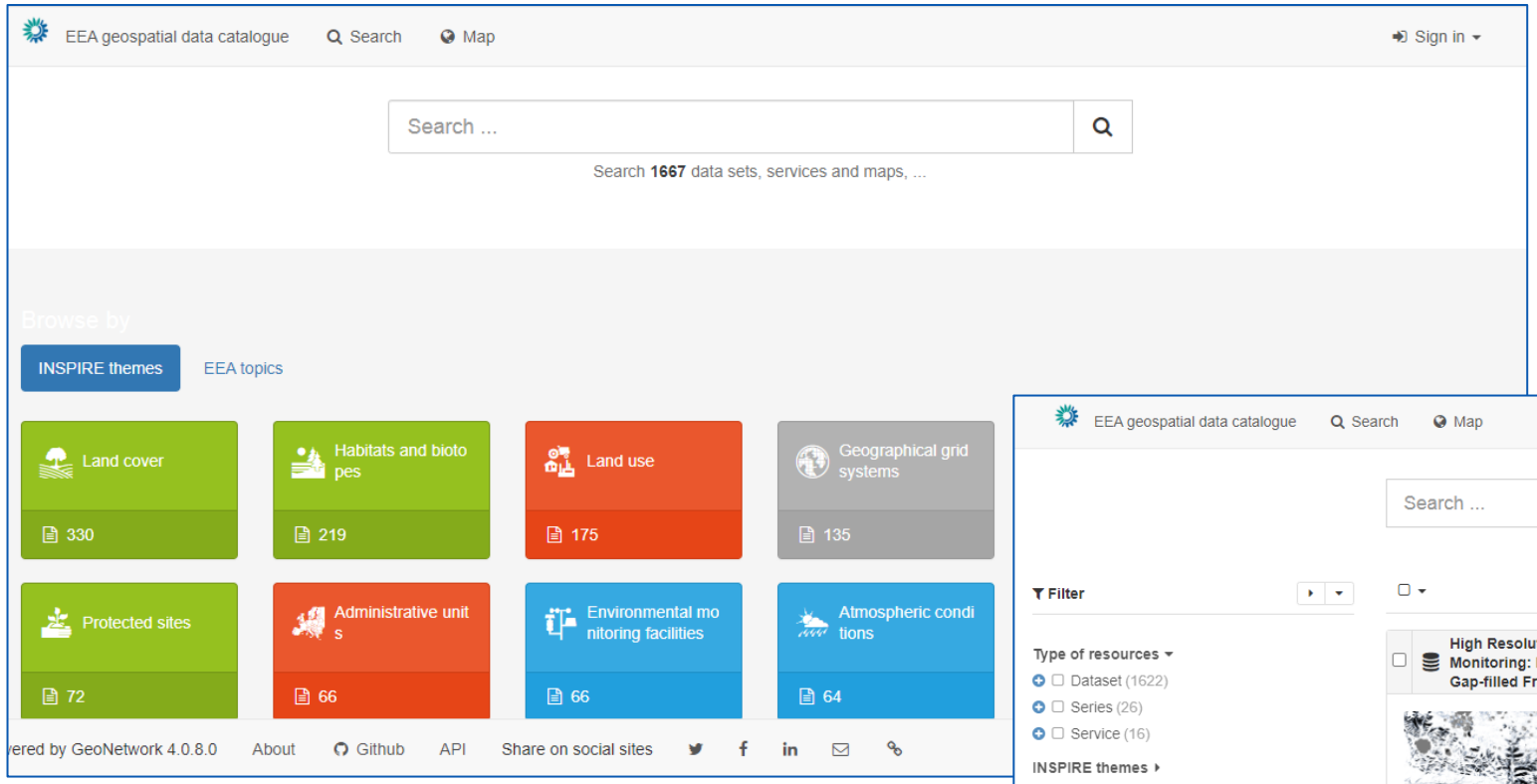
GeoNetwork

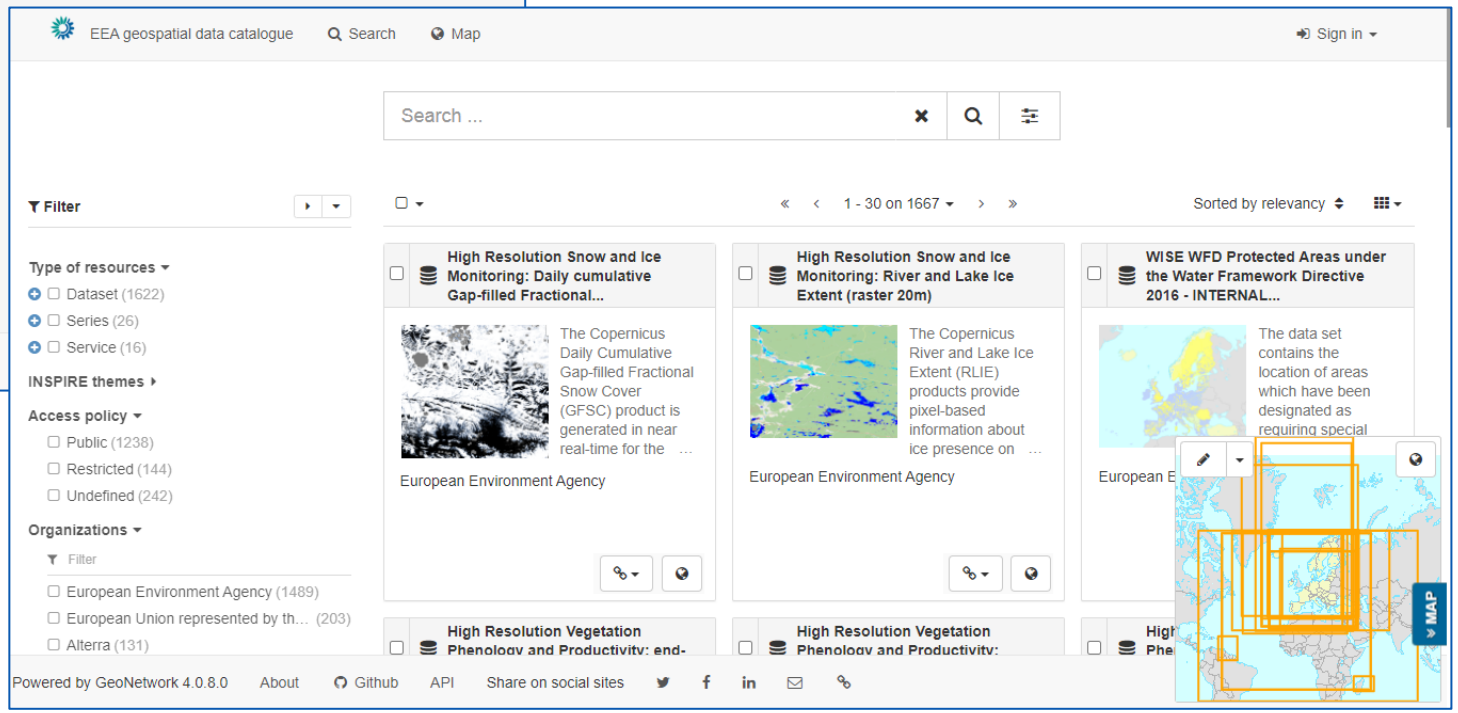
- GeoNetwork is a catalog application to **manage spatially referenced resources**. It provides powerful **metadata editing** and **search** functions as well as an interactive web map viewer. It is currently used in **numerous Spatial Data Infrastructure initiatives across the world**.
- It provides an easy to use **web interface to search geospatial data** across multiple catalogs. GeoSpatial layers, but also services, maps or even non geographic datasets can be described in the catalog.
- The **interactive map viewer based on OpenLayers** provides access to OGC services (WMS, WMTS) and standards (KML, OWS). User maps can be annotated and printed and shared with others.
- **Monitoring and reporting tools** provide summarized information about the content of the catalog and statistics on the search.

GeoNetwork metadata editor

- GeoNetwork provides an **online metadata editing tool** that supports ISO19115/119/110 standards used for spatial resources and also Dublin Core format usually used for opendata portals.
- Based on user profiles (e.g. reviewer, editor), a **dashboard** provides easy access to their information and tasks. Online editing of metadata is based on a powerful template system and directories of information (e.g. contacts, thesaurus).
- The editor provides uploading of data, graphics, documents, pdf files and any other content type. It supports among others:
  - multilingual metadata editing,
  - validation system,
  - suggestion to improve metadata quality
  - geopublication of layers to publish geodata layers in OGC services (e.g. GeoServer)

Example of a GeoNetwork catalog





European Environment Agency
SDI – Geospatial Data Catalogue
https://sdi.eea.europa.eu/catalogue

Metadata in GeoNetwork

About this resource

Categories	Datasets · Environment · Oceans
Other keywords	• GMOS OCEANOGRAPHIC CAMPAIGN • GMOS
Language	• English
Resource identifier	• http://sdi.iia.cnr.it/gos4mcat/srv/resources/554bd691-1787-4205-aa27-8a5619fe8a45
Contact for the resource	**CNR-Institute of Atmospheric Pollution Research** • Publisher: Nicola Pirrone • Originator: Francesca Sprovieri
Status	• Completed

Technical information

Update frequency	As needed
Representation type	• Vector
Lineage	The integrated Tekran speciation system was mounted on the top deck of the R/V Urania with the inlet at about 10 m above the sea

.3.0 About Github API

Spatial extent

Spatial extent

Temporal extent

Publication date
2021-04-15

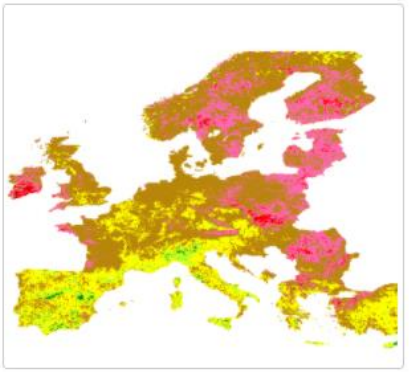
Period
Sat Jan 01 2011 01:00:00 GMT+0100 ⏩ Wed Jan 01 2020 01:00:00 GMT+0100

Provided by

Updated:

a year ago

Overview



No ratings ★

See all feedback · Add your review

Metadata information

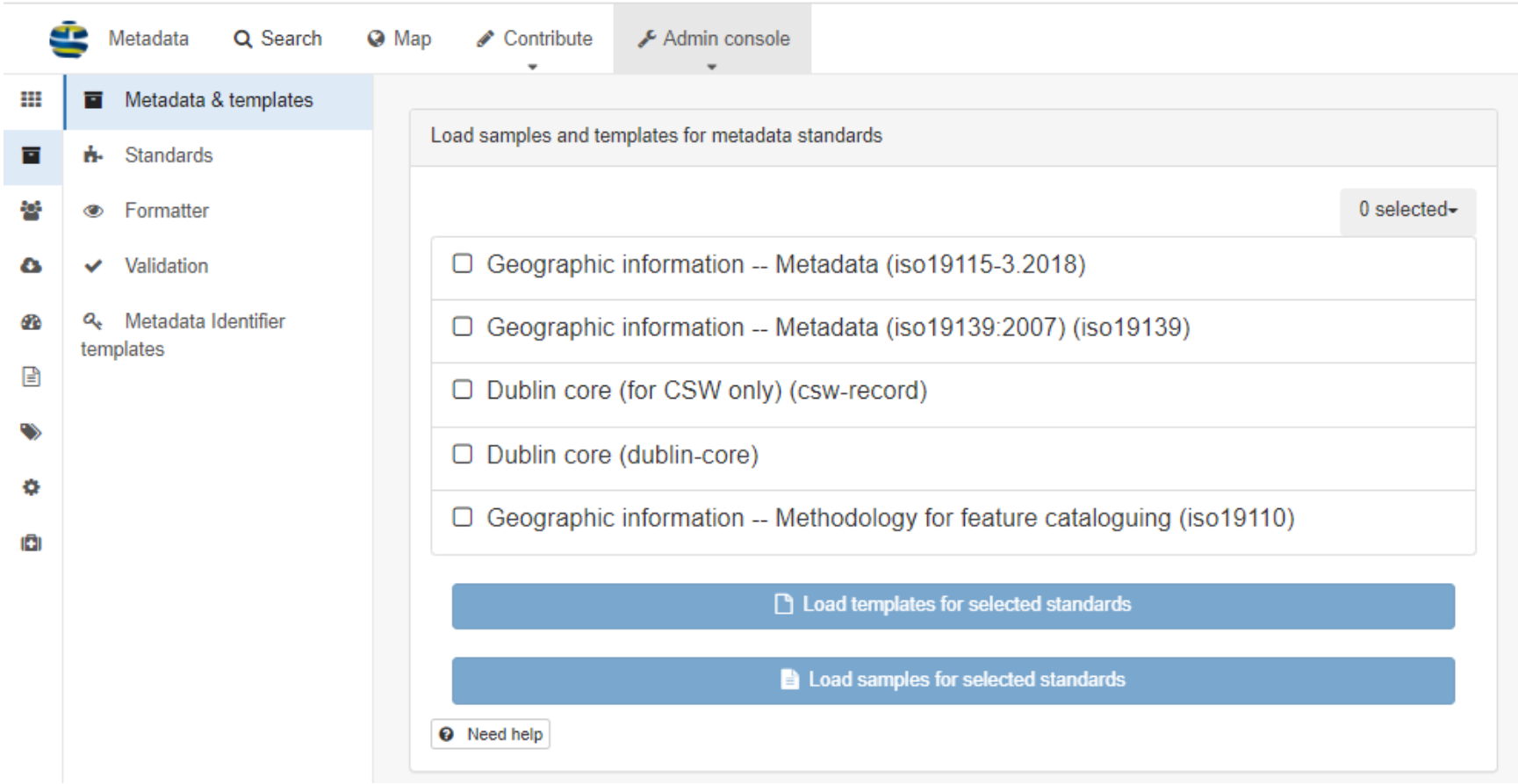
Download metadata

Metadata can be downloaded in XML format

Contact	**CNR-Institute of Atmospheric Pollution Research** • Point of contact: Francesco D'Amore
Metadata language	• English
Identifier	554bd691-1787-4205-aa27-8a5619fe8a45

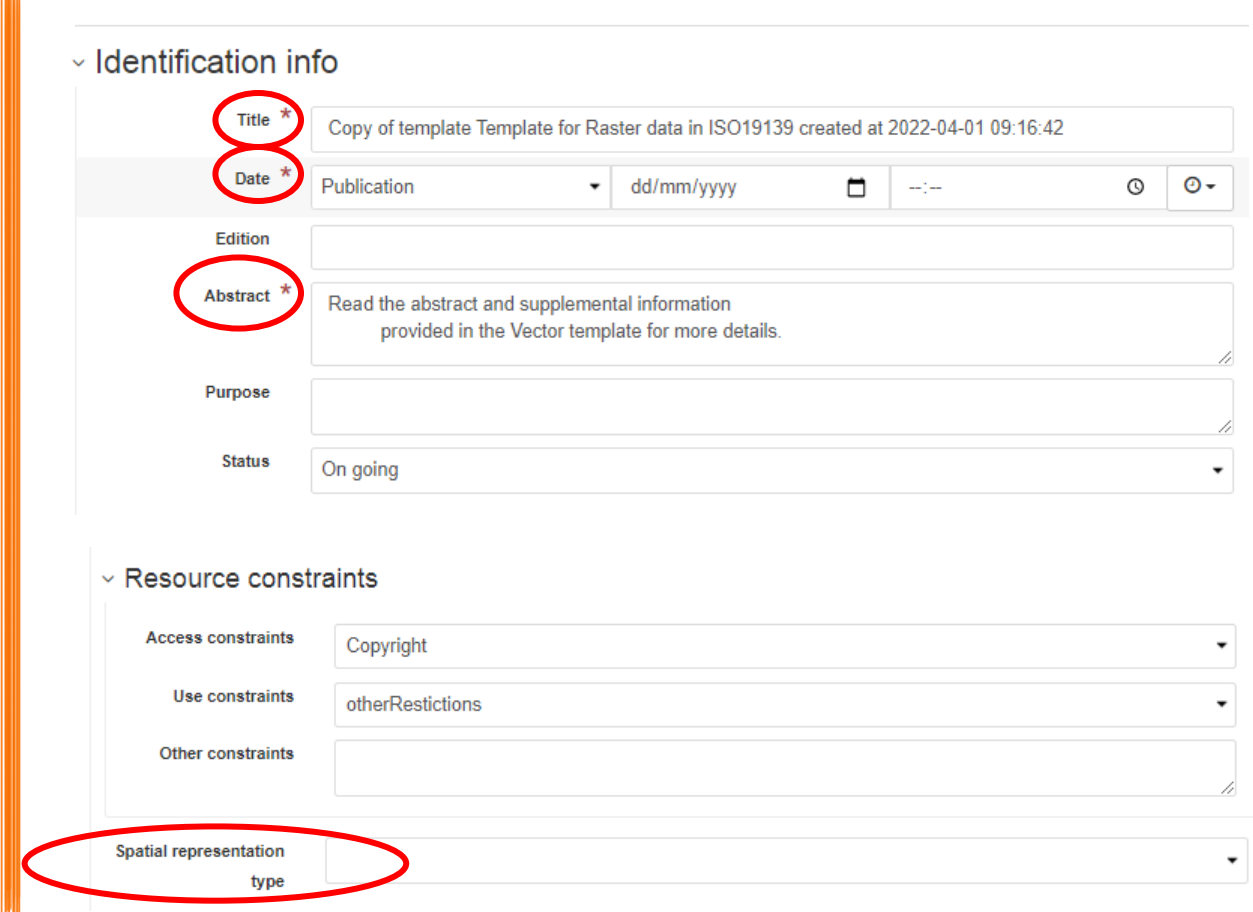
Metadata editing in GeoNetwork

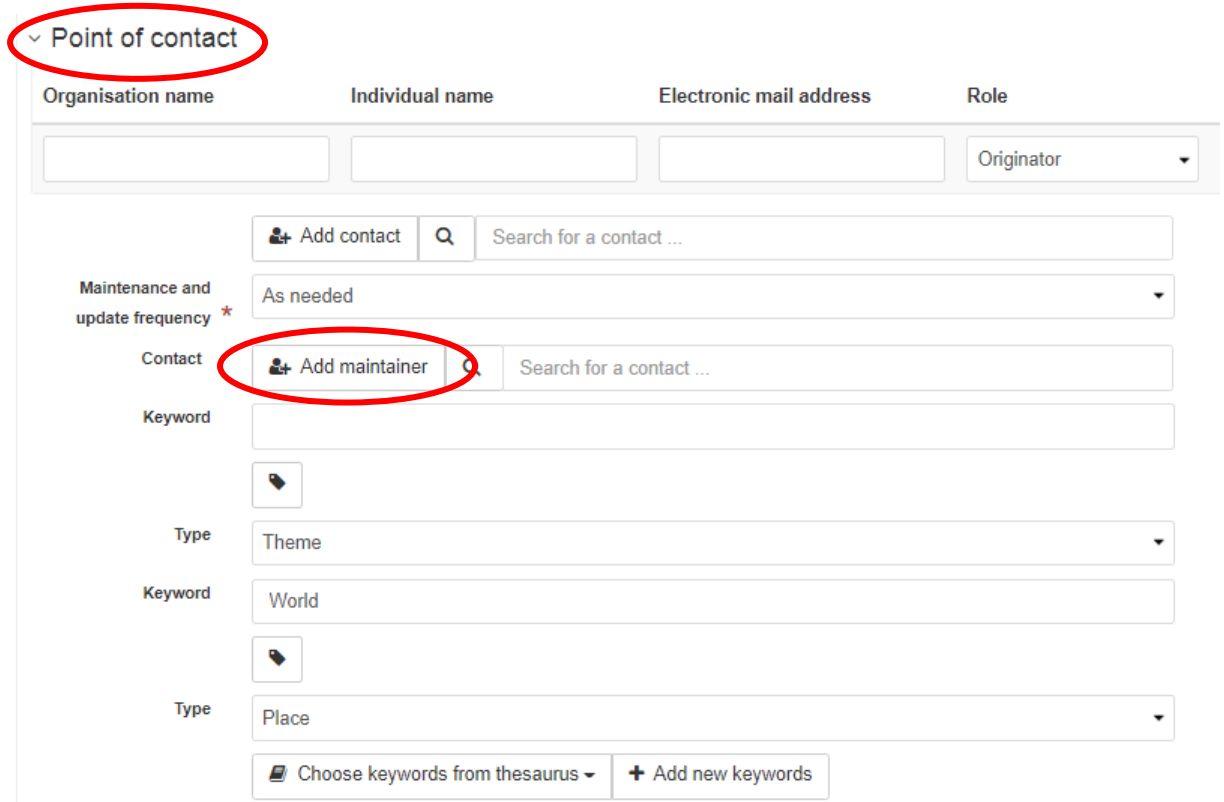
- Several standards templates/samples are provided for documenting metadata...



ISO Core elements in GeoNetwork

- ... that can be applied when creating new resources





ISO Core elements in GeoNetwork

Spatial resolution

Denominator *

Recommended values

+

Language

eng

Character set

UTF8

Topic category *

Boundaries x

Search ...

Extent

Temporal Extent

Begin date *

dd/mm/yyyy

End date *

dd/mm/yyyy

Extent

Geographic bounding box

Choose a region

Draw extent

WGS84 (EPSG:4326)

90 N

-180 W

180 E

-90 S

Supplemental Information

ISO Core elements in GeoNetwork

Reference System Information

Unique resource identifier * WGS 1984

Spatial representation info

Number of dimensions * 3

Axis Dimensions Properties

Dimension name * Row

Dimension size *

Resolution (Measure) Value Unit

Axis Dimensions Properties

Dimension name * Column

Dimension size *

Resolution (Measure) Value Unit

Axis Dimensions Properties

Dimension name * Vertical

Dimension size *

Resolution (Measure) Value Unit

Cell geometry * Area

Distribution Information

Distribution format +

Distributor +

Data quality info

Hierarchy level * Dataset

Lineage

Statement

File identifier 6704c524-a864-48a7-ab9b-0abcc0e4d9e3

Metadata language eng

Character set UTF8

Date stamp * 2022-04-01T11:42:13

Metadata standard name ISO 19115:2003/19139

Metadata standard version 1.0

Contact

Organisation name	Individual name	Electronic mail address	Role
			Point of contact

Other metadata editors

- ArcGIS Pro: https://pro.arcgis.com/en/pro-app/latest/help/metadata/create-inspire-metadata.htm
- Wx.metadata which was developed during Google Summer of Code 2014/2015 is currently available in GRASS 7 add-ons: https://github.com/OSGeo/grass-addons/tree/master/grass7/gui/wxpython/wx.metadata
- US Environmental Protection Agency (EPA) Metadata Editor: https://www.epa.gov/geospatial/epa-metadata-editor
- CatMDEdit: https://catmdedit.sourceforge.io/