Course title: EOTIST Standard course

Course subject: Computer Science

Teacher: Lluís Pesquer (CREAF)

Co-authors: Alaitz Zabala (UAB), Joan Masó (CREAF)

LESSON SC2 - METADATA ROLES
HOW TO CREATE ISO METADATA WITH GEM+ IN XML

EOTIST project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 952111

[H2020 WIDESPREAD-05-2020 (Twinning)]

TABLE OF CONTENTS

DOCUMENT METADATA

Project number	952111
Project title	Earth Observation Training in Science and Technology at the Space Research Centre of the Polish Academy of Sciences

Document title	LESSON SC2 - METADATA ROLES - GEM+ TUTORIAL
Document status	Final
Document version	1.3
Online access	-
Dissemination level	Restricted
Author(s)	Lluís Pesquer, Alaitz Zabala, Joan Masó (CREAF)
Editor(s)	-
Keywords	Metadata, Inspire, ISO 19115
Abstract	The main goal of this document is to guide the users for filling the metadata fields (as complete as possible) of the ISO 19115-1:2014 core by using the metadata editor GeM+.

HOW TO CREATE ISO METADATA WITH GEM+ IN XML

INTRODUCTION

The main goal of this document is to guide the users for filling the metadata fields (as complete as possible) of the ISO core (see Figure 1) for each dataset. These metadata fields are stored in an XML file, one XML for each database.

We are going to generate the XML file with the ***GeM+*** application. GeM+ is a free module of the MiraMon software (http://www.miramon.cat/Index_usa.htm) that runs in a Windows OS platform. GeM+ is used for research purposes in papers (Closa et al 2019) and projects (GeoViQua https://cordis.europa.eu/project/id/265178 FP7, ECOPotential http://www.ecopotential-project.eu/ H2020, NextGEOSS https://nextgeoss.eu/ H2020) also for training and management goals.

TECHNICAL STANDARDIZATION COMMENTS

Although the ISO 19115:2003 metadata document has already been withdrawn and replaced by the ISO 19115-1:2014, the last document does not modify the basic metadata that this guide proposes to encode (namely the "ISO Core" concept in 2003 that is similar to the "metadata for discovery" in 2014). Additionally, the XML encoding of the new version of the standard differs from the initial one (not only in the new metadata elements added, but also in the namespace used). Moreover, the "de facto" encoding used in most metadata distributed by official datasets producers is still the one for the first version (i.e. ISO 19139). For all these reasons this document describes the ISO Core metadata and the initial ISO 19139 XML encoding.

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)

Figure 1: Core metadata for geographic datasets. Source: ISO 19115:2003 Geographic information — Metadata

OBTAINING GEM+

The procedure to obtain GeM+ is quite easy, we just need the right located in any address, hard disk, USB, any folder we have access permissions. We should follow the next steps:

- Access https://www.miramon.cat/USA/Prod-GeMPlus.htm
- Download the last version in a **ZIP file**. We find a *GeMPlus.zip* file in our local downloads folder.
- Unzip the *GeMPlus.zip* file into a new folder, i.e. C:\GeMPlus, this name is not mandatory, any SO available pathname is allowed.

We can check the successful installation; we will find the GeMPlus.exe and the GeMPlus_64.exe files in the mentioned folder. You can use the one according to your operating system.

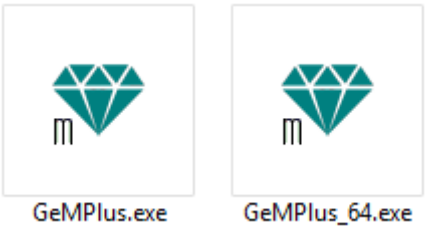

Figure 2: GeM+ icon identification

FIRST GEM+ EXECUTION

To execute the GeM+ application, double click in the file on the corresponding folder, (we can generate a link in the desktop).

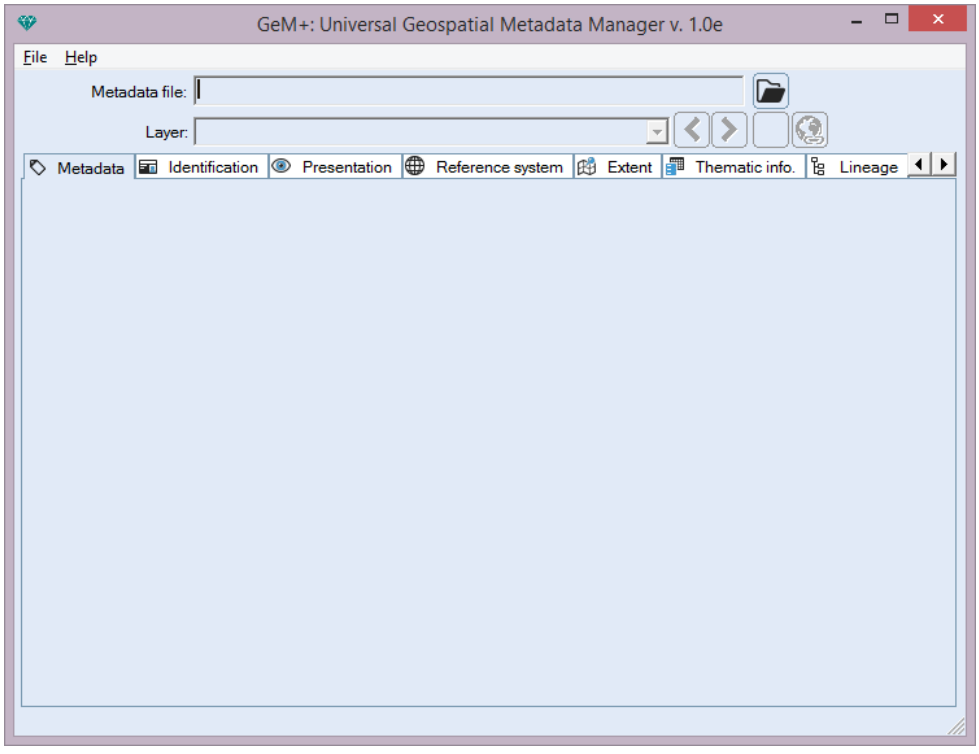

Figure 3: Default GeM+ interface

XML CREATION

We access to *New XML* option in the *File* menu:

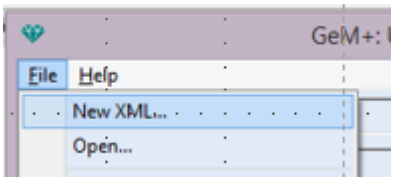

Figure 4: New XML menu access

We browse in our folder system and select the output folder and the name of the XML.

In Figure 5 we can see an example:

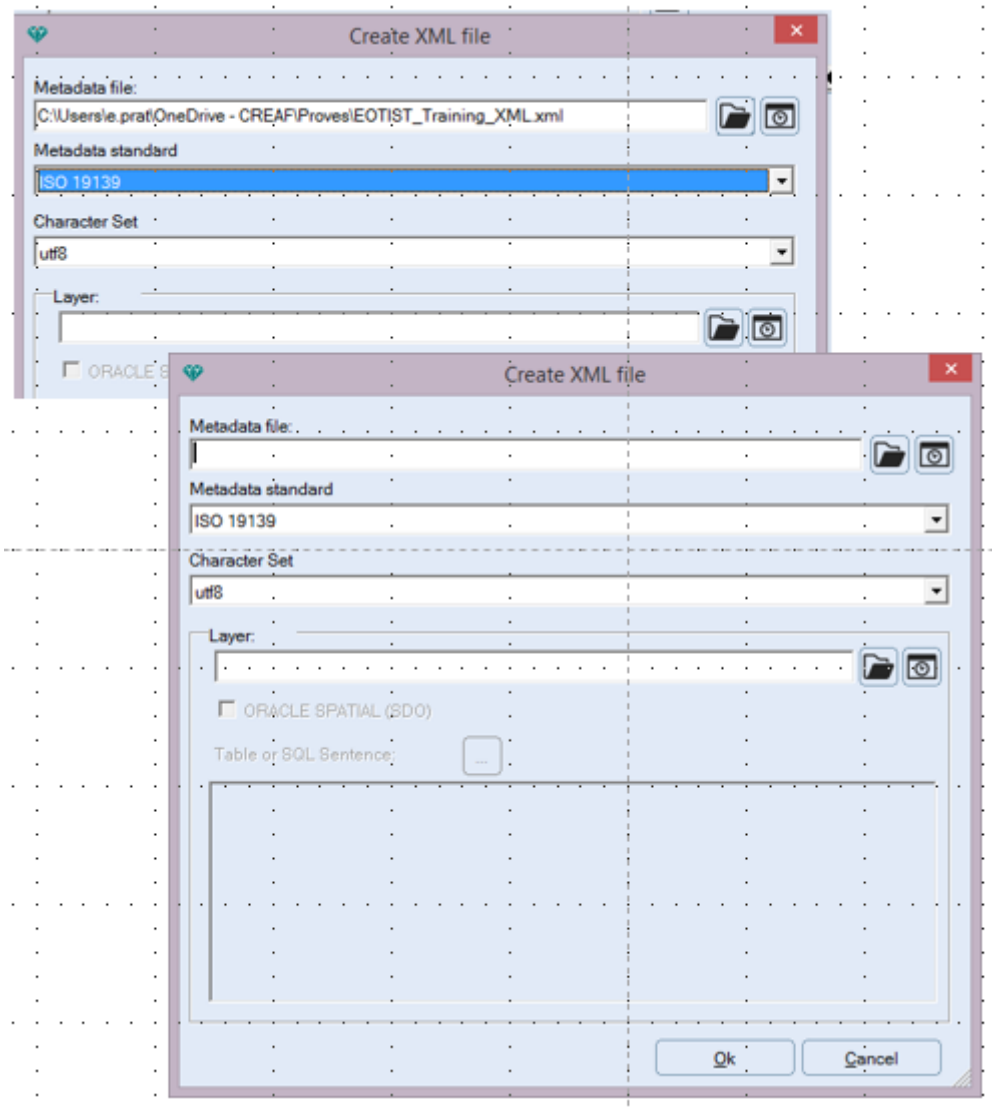

Figure 5: XML creation interface

We should maintain the default values and then, the **Metadata standard name**, **Metadata standard version** and **Dataset character set** are defined in this form. We can push the *OK* button for generating the preliminary XML file.

Additionally, the Metadata date stamp is generated, however it is updated at every Save action of the XML file.

XML CONTENT FILLING

DATASET TITLE (& MORE)

We access to Identification tab and generate a short Title. Additionally, we can fix the **Metadata language**.

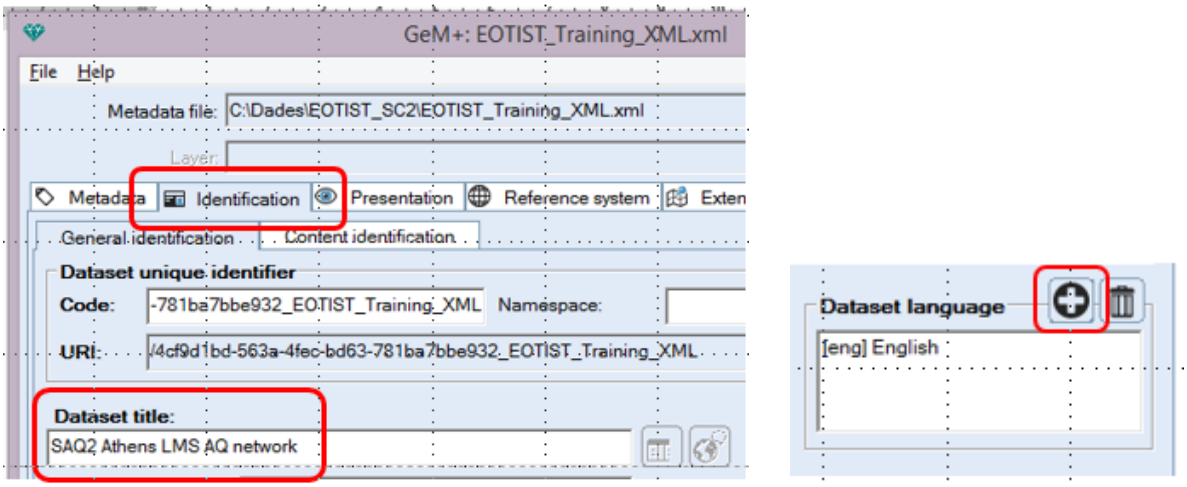

Figure 6: Title and language definition

With this action, **Metadata file identifier** field is automatically defined.

ABSTRACT DESCRIBING THE DATASET

We can also define the Summary and the Purpose field. Both metadata fields are in the *Presentation* | *General Presentation* tab (Figure 7).

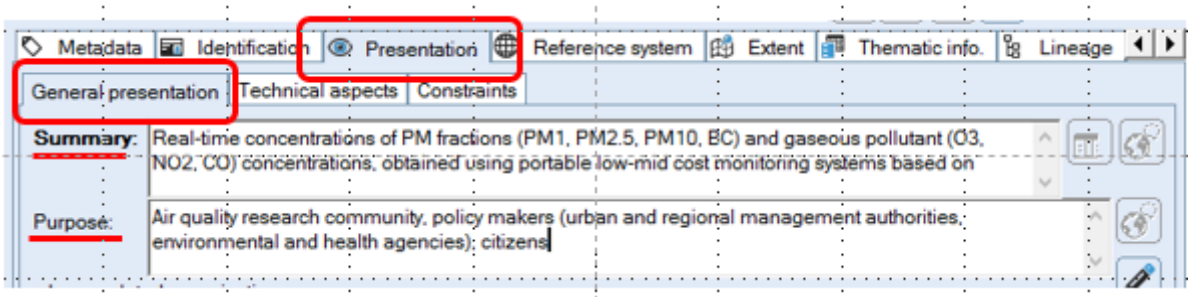

Figure 7: Filling the abstract and the purpose

DATASET LANGUAGE

In the mentioned *Identification* tab, we can define the language of the dataset, by clicking in the button  and selecting the corresponding language. This is the language of the content of the dataset, and it could be different from the metadata language.

We can save the current content of the XML and continue filling (see ***Save XML*** section in this document).

DATASET RESPONSIBLE PARTY (& MORE)

In the *Presentation* | *General Presentation* tab, there is a *Layer related organizations* section, and we can add the needed content thought the edit (pencil) button  in the right side of this section.

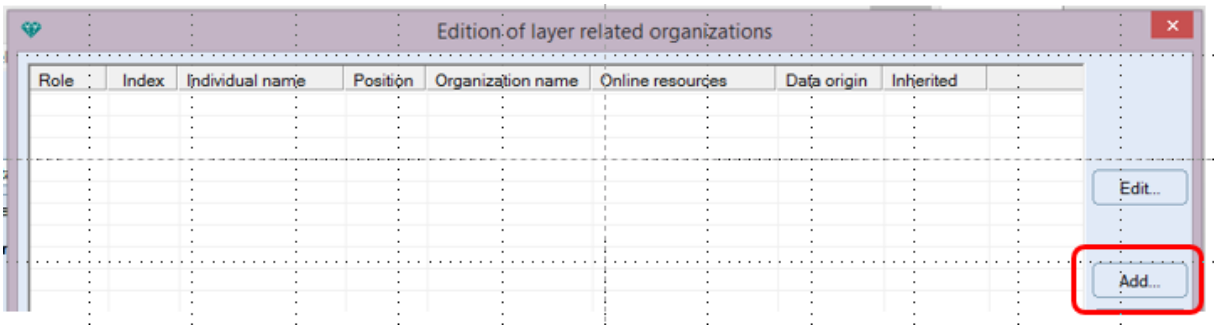

Figure 8: Interface for adding a layer role item

By filling in the different boxes of the interface in this section (see Figure 9), we will complete all the desired details of the responsible party.

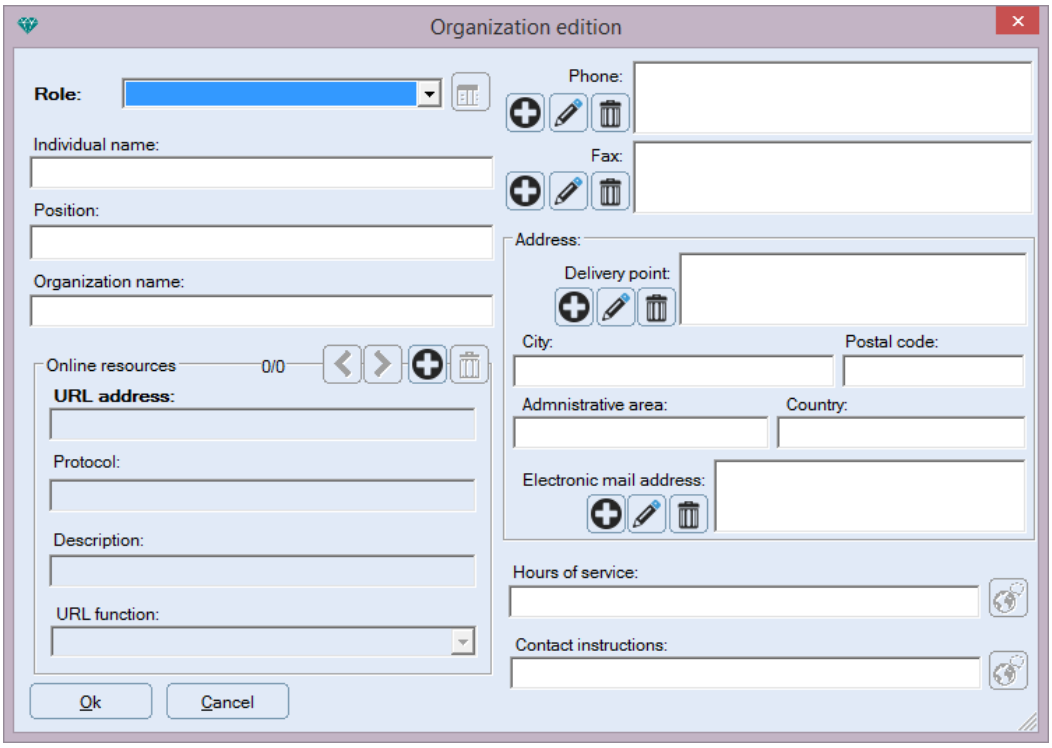

Figure 9: Organization edition interface

Same interface to edit the information for an organization (or responsible party) is used for defining the **Metadata point of contact.** This field is accessed from the *Metadata* | *Metadata info* tab.

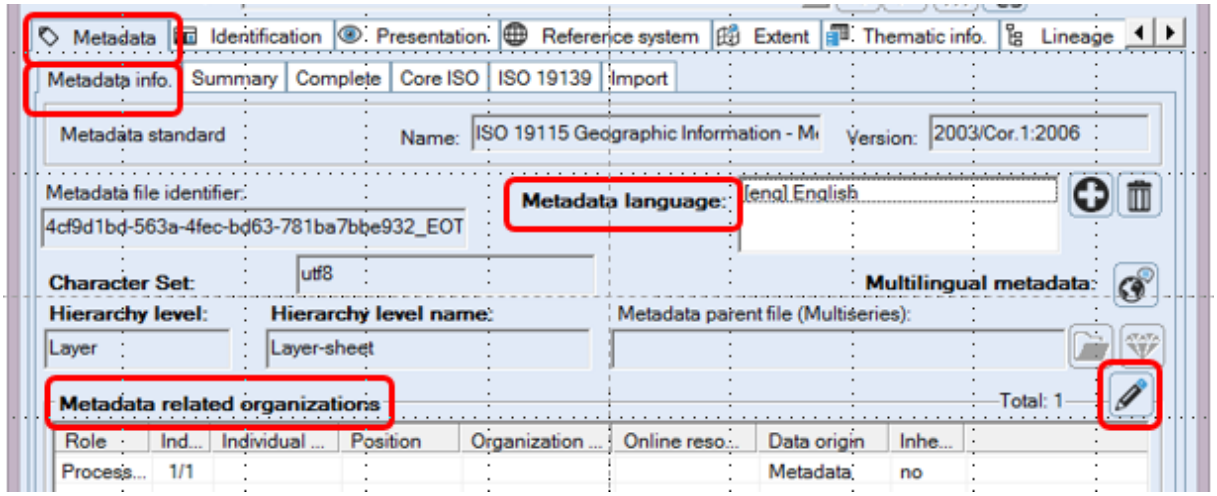

Figure 10: Interface of main fields of metadata information

Finally, again in this tab, we can define the **Metadata language**.

DATASET TOPIC CATEGORY

We can select one (or more) topic category from the proposed list and add some keywords related to the dataset domain. These topic categories and keywords will be helpful for the catalogs search utilities.

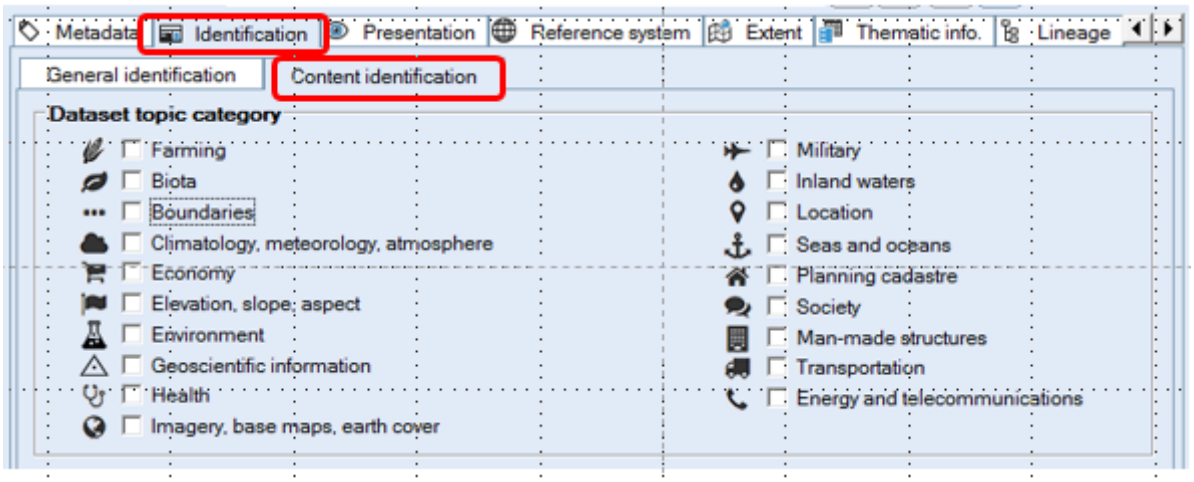

Figure 11: Choice of the topic categories

If keywords are described, it is advisable to select them among the ones provided by controlled vocabularies, such as Gemet. Having at least one keyword from the GEMET vocabulary is a requirement for INSPIRE metadata implementing rules, so one of these keywords should be selected at least by double clicking in it in the central window (see Figure 12).

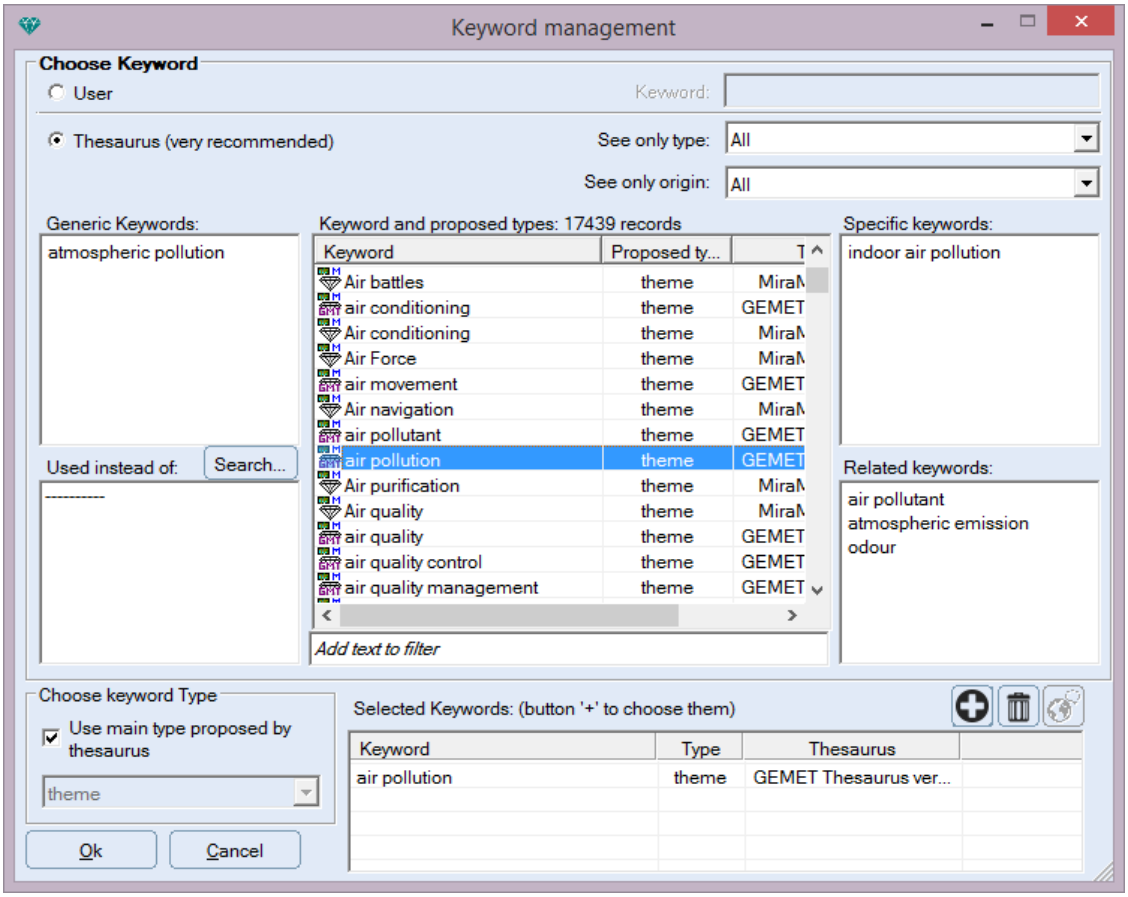

Figure 12: GEMET keywords selection

SPATIAL REPRESENTATION TYPE

We can select the data model in the *Presentation* | *Technical aspects* tab and select the corresponding option from the offered list.

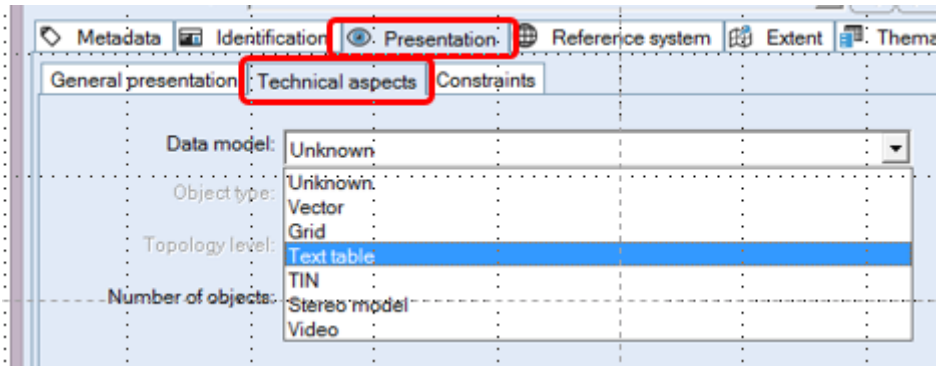

Figure 13: Choice of the data model

DISTRIBUTION FORMAT

In the *Distribution* tab (we can move between tabs with forward and reverse icons on the right side for accessing all tabs if we use a small size interface) the format of the dataset can be described.

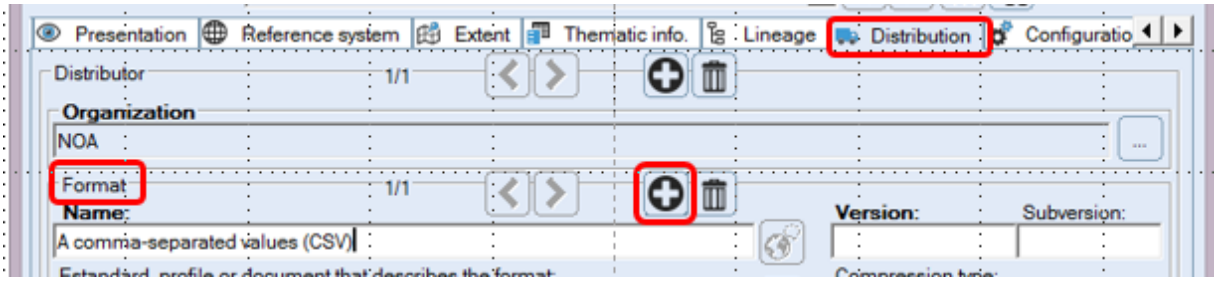

Figure 14: Choice of the data format

In case of informing about the Transfer options (they are optional), note that *the size of the distribution unit* is expressed in megabytes.

REFERENCE SYSTEM

The reference system of the dataset is defined in the *Spatial reference system* tab. GeM+ is a module of the MiraMon software and it follows its own design and codification for the complete reference system definition (more info in http://www.creaf.uab.es/miramon/help/eng/ and Pesquer et al 2007).

This system supports most of the EPSG codes (https://epsg.io/) and the most usual systems are included and easy to link it in the metadata. However, in case of some local reference systems, we will need to manually add all details of the reference system in the corresponding MiraMon geodesy tables.

As an example in geographic coordinates on WGS84 datum, we need to select a *Cartographic* type and the corresponding description item in the drop-down list.

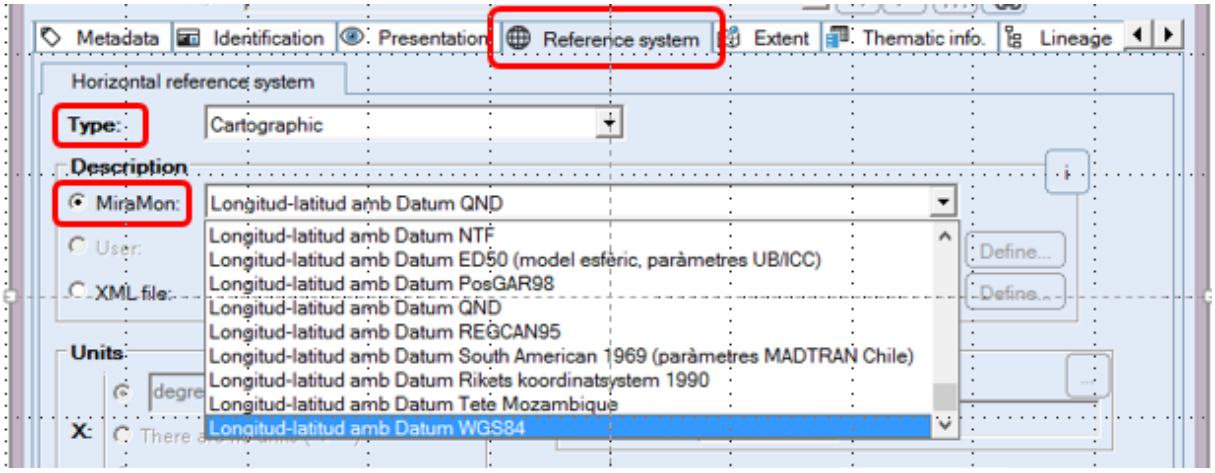

Figure 15: Reference system definition

GEOGRAPHIC LOCATION OF THE DATASET (& MORE)

The Core ISO propose the definition by four coordinates or by geographic identifier. In the *Extent* tab we can define these four coordinates as projected coordinates or in latitude/longitude system. In the *Description* tab, we can define it as a text (Toponym or longer text Description).

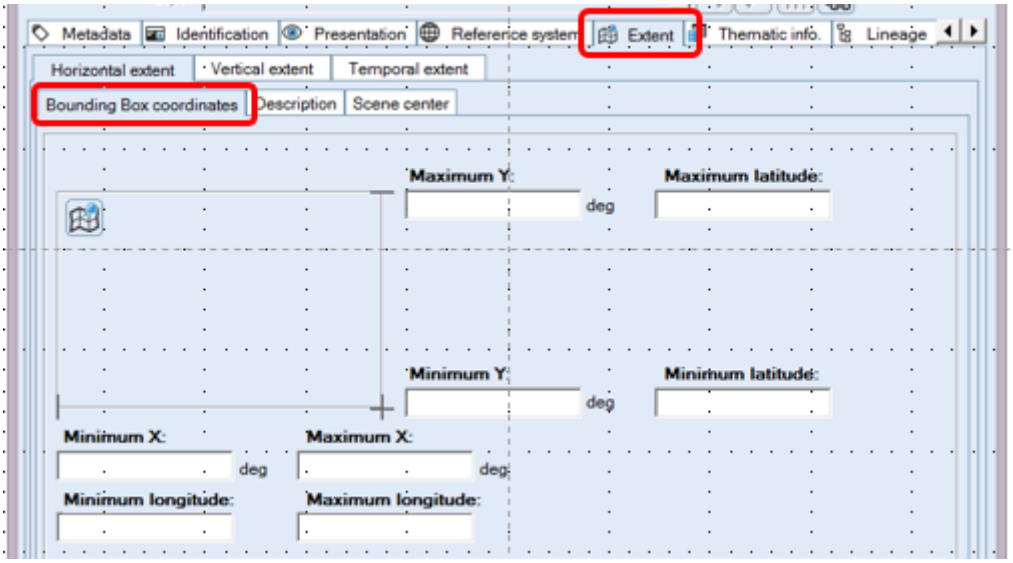

Figure 16: Geographic location by four coordinates

In this *Extent* tab, we can define the **Additional extent information for the dataset** metadata fields. In particular, for the datasets obtained from in-situ collection (and others too), the Temporal extent definition is very relevant. We will use the *Temporal extent* tab (see Figure 17).

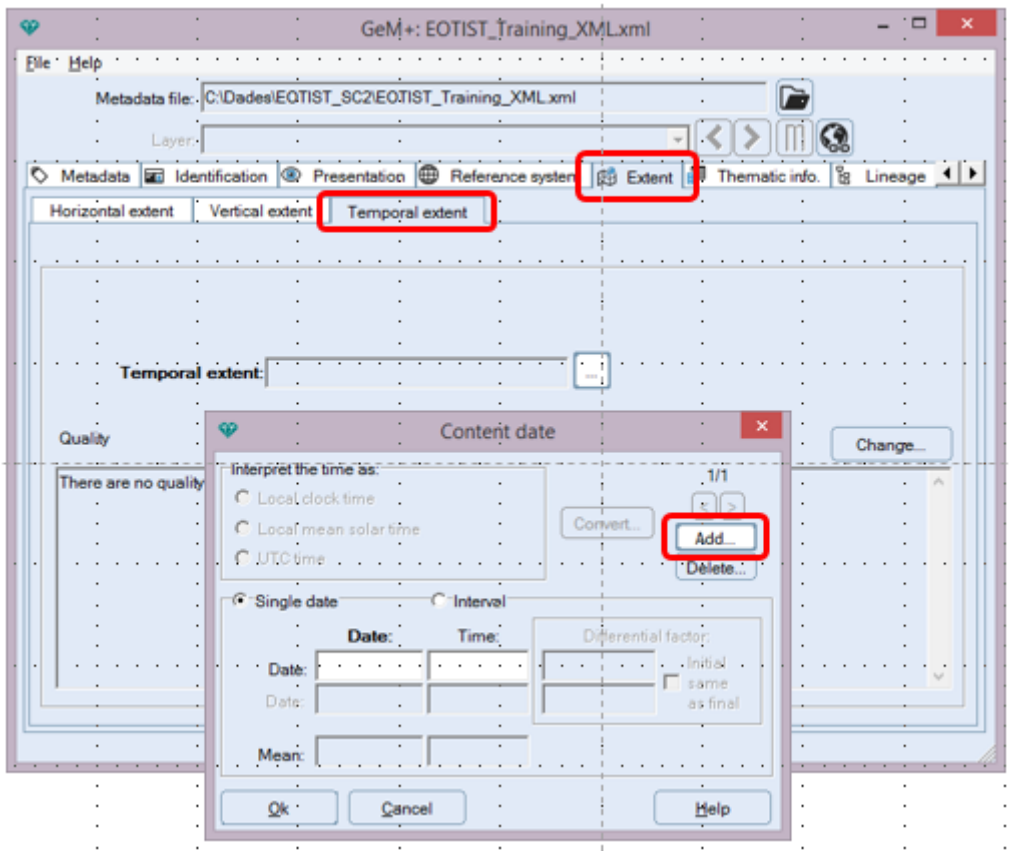

Figure 17: Temporal extent definition

Regarding temporal metadata fields, it is important to distinguish the Temporal extent corresponding to the content information than the **Dataset reference date.** This is the date of creation, publication or revision of the dataset, but the temporal content of the information is many times different. For example, we can generate today (2020) a 1956 land cover map; 2020 is the Dataset reference date and 1956 is the Temporal extent.

We define the Data reference date in *Presentation | General Presentation* tab:

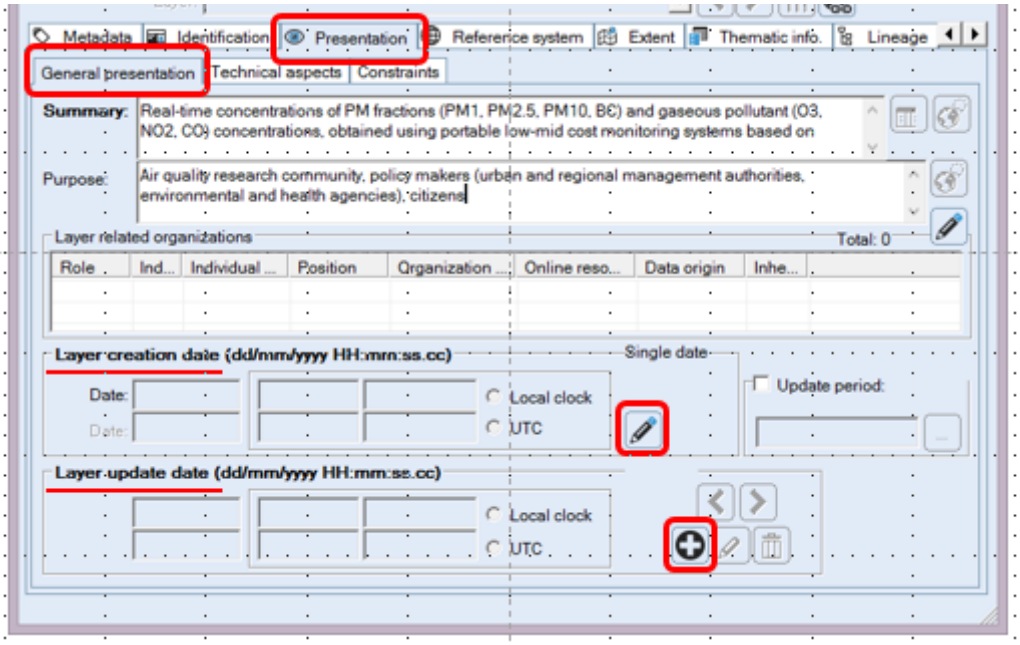

Figure 18: Dataset reference date definition

LINEAGE

Accessing to the *Lineage* tab, first of all we can add a general textual description of the lineage:

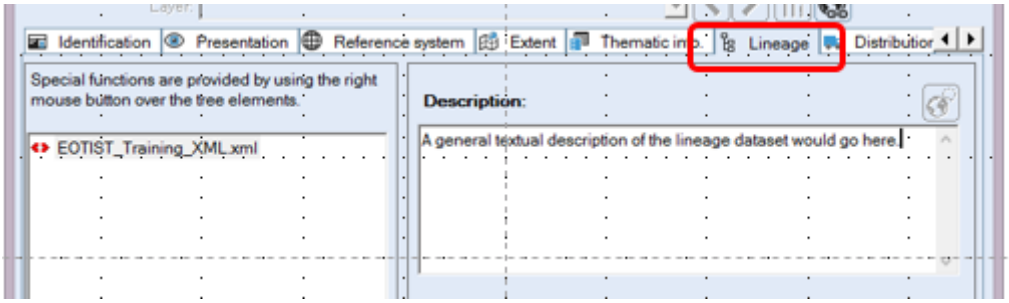

Figure 19: Lineage general description (or statement)

Moreover, we can add by right clicking a list of ordered processes (on the top, the last one) that explains the track processing chain of involved data until the generation of the final output dataset. We can inform about relevant details of each process step, too.

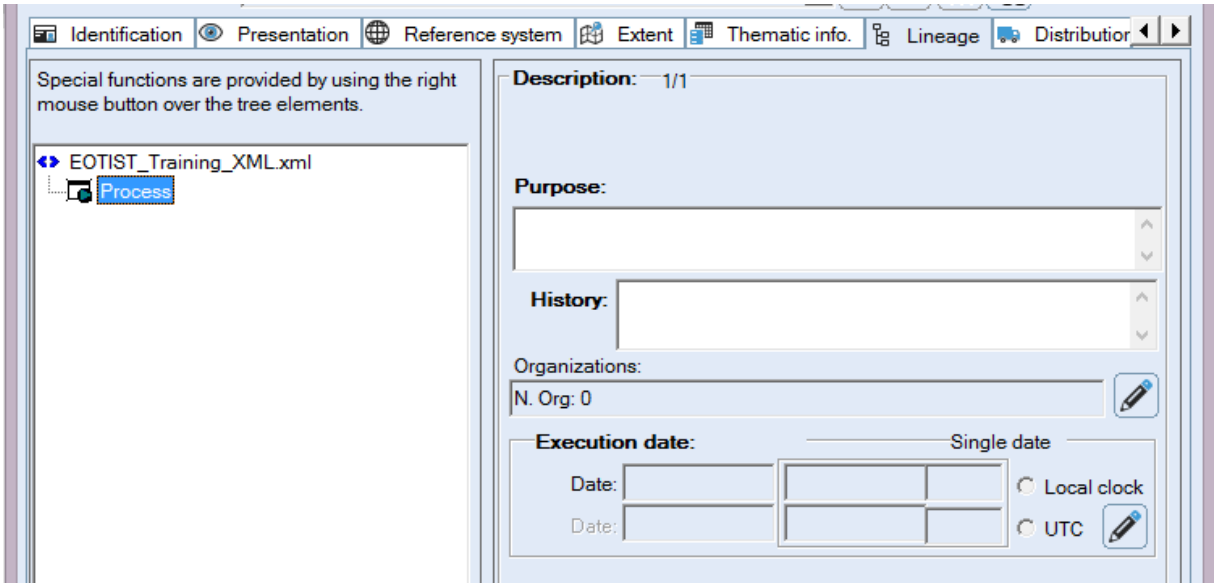

Figure 20: Lineage process step definition

Usually, the *Purpose* field contains an overall information about the objectives of the module or application associated to the process. The *History* field gives more particular details about the involved process, such as the parametrization of the process, the command line of a script or module execution, etc. The *History* and *Execution date* are relevant for the reproducibility and traceability of the process.

Finally, we can add involved sources (e.g. other datasets, etc). They can be defined for a certain process or for the whole dataset (depending on from which context menu you trigger the *Add Source* option).

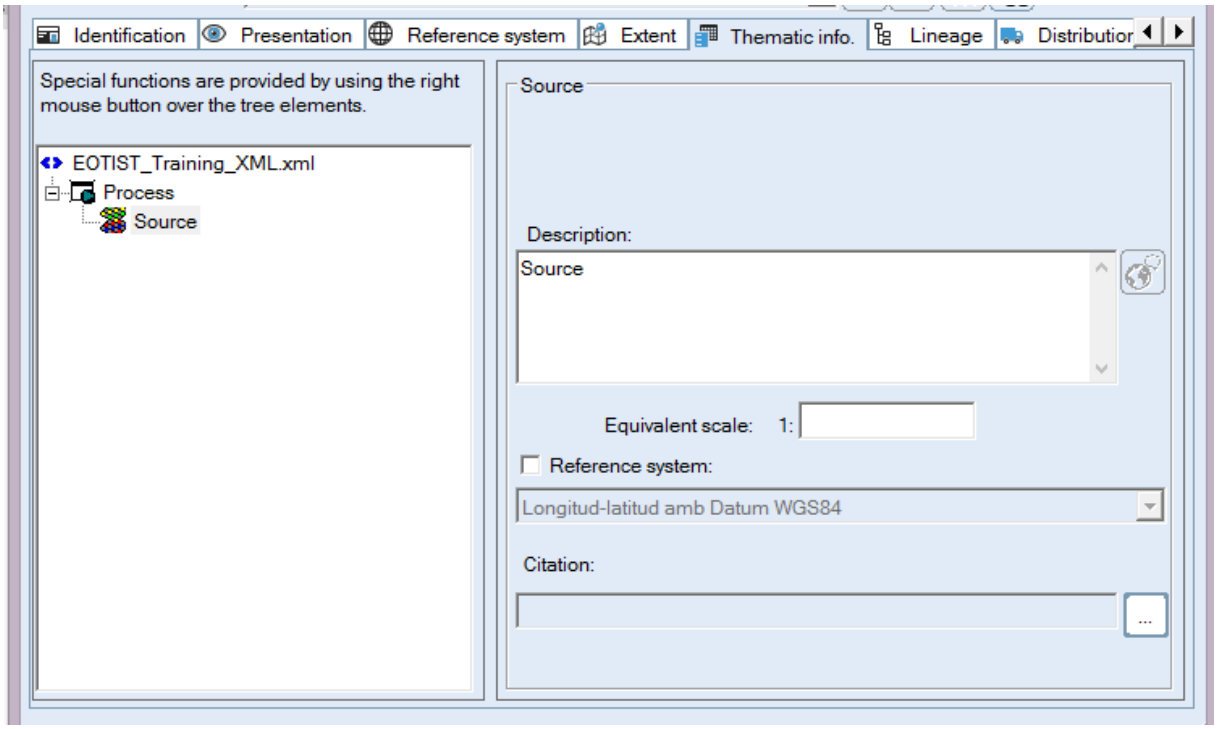

Figure 21: Lineage source definition

XML FILE READING

As a simple double-check (additional to GeM+ verifications) we can read the XML with any other software, even an internet browser:


```
<?xml version="1.0" encoding="UTF-8"?>
- <gmd:MD_Metadata xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://schemas.opengis.net/iso/19139/20060504/gmd/gmd.xsd
  http://www.isotc211.org/2005/gmx http://schemas.opengis.net/iso/19139/20060504/gmx/gmx.xsd" xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:gmx="http://www.isotc211.org/2005/gmx" xmlns:gts="http://www.isotc211.org/2005/gts" xmlns:gml="http://www.opengis.net/gml"
  xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:gmd="http://www.isotc211.org/2005/gmd">
    - <gmd:fileIdentifier>
        <gco:CharacterString>fc5a7ffb-7c0a-4cc0-bd8b-1e2ed8795e6d_IS_LMS_NOA_v0</gco:CharacterString>
      </gmd:fileIdentifier>
    - <gmd:language>
        <gmd:LanguageCode codeListValue="eng"
          codeList="https://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#LanguageCode">eng</gmd:LanguageCode>
      </gmd:language>
    - <gmd:characterSet>
        <gmd:MD_CharacterSetCode codeListValue="utf8"
          codeList="https://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#MD_CharacterSetCode">UTF-8</gmd:MD_CharacterSetCode>
      </gmd:characterSet>
    - <gmd:hierarchyLevel>
        <gmd:MD_ScopeCode codeListValue="dataset"
          codeList="https://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#MD_ScopeCode">Layer</gmd:MD_ScopeCode>
      </gmd:hierarchyLevel>
    - <gmd:hierarchyLevelName>
```


Figure 22: XML in the Internet Explorer browser

Attention: It is strongly recommended to avoid any editing with non-specialized software.

BASIC GEMM FUNCTIONALITIES

SAVE XML

At any step of the process of metadata generation, we can save the XML file, accessing to the File menu.

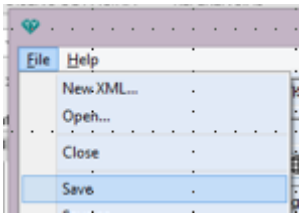

Figure 23: Save the XML file

In case of an existing incoherence with the standardization rules, we must solve if before saving.

OPEN XML

Accessing to the file menu and browsing to the corresponding folder, we can open an existing XML file in order to update its content.

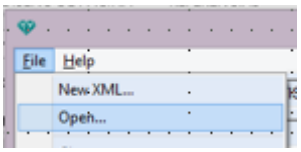

Figure 24: Open a XML file

PRETTY VISUALIZATION

In the *Metadata* tab, there are some options for checking and visualization the current content of the XML file following some predefined profiles. The Core ISO is the most recommended.

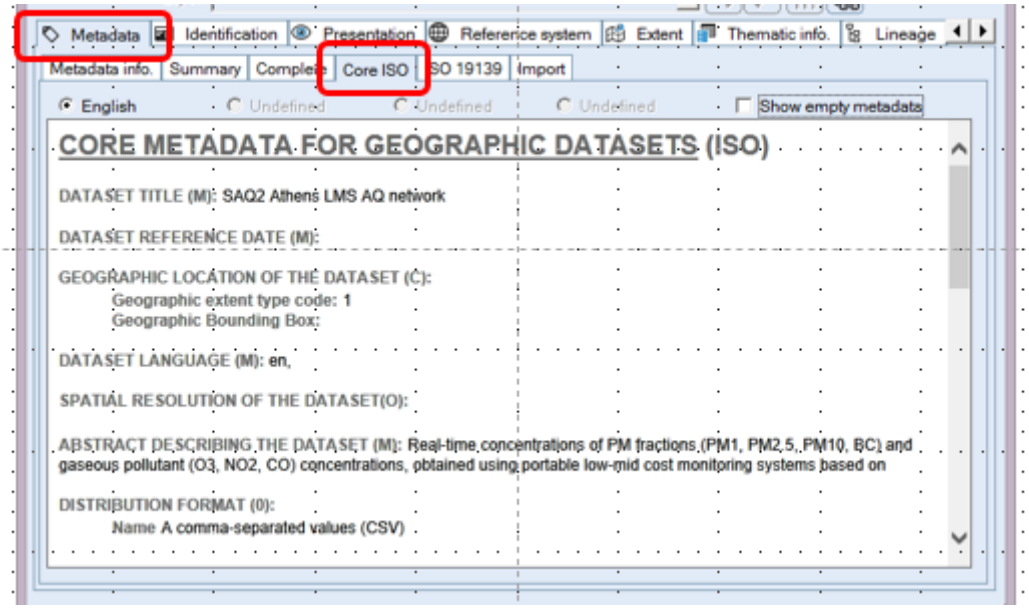

Figure 25: CoreISO metadata fields and content visualization in pre-formatted style

REFERENCES

Closa G, Masó J, Zabala A, Pesquer L, Pons X (2019) A provenance metadata model integrating ISO geospatial lineage and the OGC WPS: Conceptual model and implementation. *Transactions in GIS* 23(4): 683-705. DOI: 10.1111/tgis.12555.

ISO 19115:2003 Geographic information — Metadata. (https://www.iso.org/standard/26020.html)

ISO 19115-1:2014 Geographic information — Metadata — Part 1: Fundamentals (https://www.iso.org/standard/53798.html)

Pesquer L, Pons X, Masó J (2007) Implementación integrada en el SIG de los diversos tipos de transformaciones de Datum (Integrated GIS implementation of various types of Datum transformations). Actas de la 7ª Setmana Geomàtica. Sensores de alta resolución y sus aplicaciones [CD_ROM]. 2007. File: pap034.pdf. D.L. B-9432-2007.