Spatial data infrastructure of Navarre (Idena)

S. FONTANO, M. A. JIMÉNEZ, M CABELLO

(S. FONTANO, M. CABELLO: Trabajos Catastrales, S.A. Ctra. del Sadar s/n –Edif. El Sario. 31006 Pamplona (Spain)
sfontano@tracasa.es/mcabello@tracasa.es
M.A. JIMENEZ: Government of Navarra. Public Works Department, Avda. San Ignacio 3. 31002 Pamplona (Spain)
ma.jimenezdecisneros.fonfria@navarra.es)

Key words: Metadata, Corporate information systems Spatial Data Infrastructure, Navarre,
SITNA (Territorial Information System of Navarre), IDENA (Spatial Data Infrastructure of Navarre).

SUMMARY

Navarre has made a great effort in the last several years to have a real Territorial Information System, (better known by the acronym SITNA) and it has configured itself as the component within the Corporate information system of Navarre that places within reach of all citizens, data of a very diverse kind which refers spatially to the territory of the community.

Despite the undoubted success of the implementation of this initiative, the passing of time made it clear that there was an important gap in the loading of data, because it was carried out without taking into account the question of metadata.

At that time, there was a lack of standards, experiences and coordination among institutions, in short, an unclear horizon concerning this matter that advised against the design of a metadata system with an uncertain future.

The evolution of this matter in the last two years in Europe has seen a radical change after the publication of the proposal of European Directive INSPIRE[1] (Infrastructure for Spatial InfoRmation in Europe) that established an infrastructure for spatial information for the area of the European Union, the election of the ISO-19115 standard and the following multiplication of SDI initiatives in all fields.

This context frames the IDENA initiative (Infrastructure for Spatial Data of Navarre) that tries to give SITNA those characteristics that it lacked: Catalogue of Meta information, Interoperability with other SDIs and Distribution.

The present article describes the general approach of the IDENA project, focused on IDENA architecture and ist components as well as the resolution of the metadata work flow within the IDENA project and the interoperability with other Spatial Data Infrastructures

1 INTRODUCTION

As it is well known, INSPIRE is an initiative of the European Commission whose goal is to promote the creation of an Infrastructure for European Spatial Information, establishing what geographic information is pertinent, with conditions of harmonization and quality to the service of the training, application, follow-up and evaluation of the community policies with a territorial dimension or impact.

The general situation of the Spatial Information in Europe is that of great fragmentation of dataset and sources, vacuums in the availability, lack of harmonization among datasets of different geographic scales and duplication of information collections. These problems make identification, access and use of the available data difficult. To solve these problems and fulfill the principles of INSPIRE, the documentation of all existing spatial data in an SDI becomes essential with the goal of making its characteristics known. This way, the principal objective of an SDI: DISCOVER, ACCESS, USE the geographic information, is achieved. The key to all this is the metadata.

The SITNA model is a Corporate Information System that tries to offer concrete and advanced consultancy services, analysis and management of the different backgrounds of identified users by designing appropriate tools for each of them.

Besides this, Navarre did not have a Spatial Data Infrastructure recognized within the INSPIRE framework, which lead to a readjustment of the System that would allow, maintaining the current operation and tools, the collection of the concepts contemplated by INSPIRE for the development of the SDI, as well as the initiatives that were already in progress, establishing the objectives and the steps to follow for the IDENA implementation (Spatial Data Infrastructure of Navarre)[2].

2 IDENA: SDI FOR NAVARRE

2.1 What we understand as SDI

An Infrastructure for Spatial Data tries to identify and place within reach of the whole world, the geographic information that it has, properly documented through standardized metadata. From this point of view, it is therefore understood, that SDIs constitute real catalogues of cartographic products available through the Internet, that allow the user to access and obtain geographic information and that look for the interoperability with other SDIs on all levels (local, regional, national, international...). Basically, one can say that an SDI is made up of a base of metadata, with a catalogue whose consultation interface is the same as other SDIs as well as consultation tools. All this allows the localization of what data and services are available with similar characteristics, on a specific subject, in a specific spatial location. In short, consult the metadata that interests us offering, at the same time, the possibility of visualizing it. All this greatly facilitates the correct use of the information.

---

[1] Commission of the European Communities, 2004. "Proposal for a Directive of the European parliament and of the council establishing an infrastructure for spatial information in the Community (INSPIRE)", Brussels. http://www.ec-gis.org/inspire/.

[2] The prototype of IDENA was presented in the JIDEE'04 conferences in Saragossa, the official presentation was held on March 9, 2005.

The IDENA initiative (Infrastructure for Spatial Data of Navarre) began during 2004, to complement and give the Territorial Information System of Navarre (SITNA) those characteristics that it lacked: Catalogue of Meta information, Interoperability with other SDIs and Distribution.

IDENA will be the Infrastructure for Spatial Data of Navarre, that will fulfill the objectives outlined by INSPIRE for the elaboration of regional SDIs and will be interoperational with other SDIs, like the Spanish or European. Therefore, it should also have the necessary technological elements and provide the geographic data contemplated by INSPIRE.

2.2 IDENA Objectives

The main objective of IDENA, consequently, is to provide access through Internet to the geographic information of the Corporate Information System of Navarre in an integrated and open way. Initially, we formulate a serie of generic goals.

For data:

1. Documentation of all system information according to the background metadata defined from standards.
2. Make the search for information possible based on specific search criteria (this criteria is based on the INSPIRE directive) through a metadata catalogue (Localization service).
3. Allow the free viewing of the results of our searches, for metadata as well as data (Viewing service).
4. Offer the free download of basic data (Download service).
5. Facilitate the acquisition of paid data through electronic commerce (Paid service).

For components:

1. Development of a map viewer.
2. Implement services for access to spatial data services.
3. Development of a feature server (WFS), a coverage server (WCS) and a map server (WMS).

All this, respecting the standards and protocols recommended by INSPIRE (ISO 19100 and Open Geospatial Consortium) for interoperation with other SDIs on different levels (national, international, etc.).

3 IDENA ARCHITECTURE

As mentioned earlier, a SDI has to be accessible for any user through the Internet. The architecture designed by IDENA is shown in the following figure:

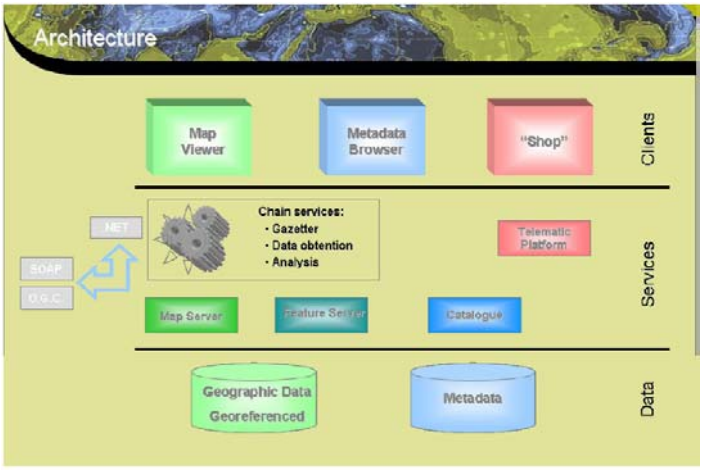


Figure 1. IDENA Architecture

The IDENA portal is made up of three main elements: a Map Viewer, a Metadata Search Engine and the Shop.

- Map viewer. This tool will allow viewing, consulting, identifying, etc. the spatial information of Navarre. At the same time, it will also allow the connection with other information servers (geographic data of other SDIs that follow the same standards and protocols), to view spatial data on other scales (other Spanish communities, European...).
It offers standard navigation tools, and also has others such as transparency, legend view, access to thematic information for a specific point on the map, save maps, etc.

In the final product, the Map Viewer will work with geographic data and georeferenced data, and will have a map server and a feature server. It also will offer other complementary services like the Gazetteer, data extraction and analysis.

- Metadata Search Engine. This tool allows metadata searches based on different criteria established by INSPIRE.
The result of this search presents a list of all the data including these criterias. Each result will mention 5 principal characteristics (title, summary, editor, format and additions).

Once the data of interest is found and selected by the user, it could access the following options:

- Visualization of the Map Viewer: “See map”.
- Consultation of the metadata (the main metadata will be seen and complete metadata can also be accessed): “See details”.
- Downloads, shop (for free data, with registered users or the sale through Internet): “Products”.
- Link to SITNA: “Go to SITNA” .
- Link to Website: “Go to website” .

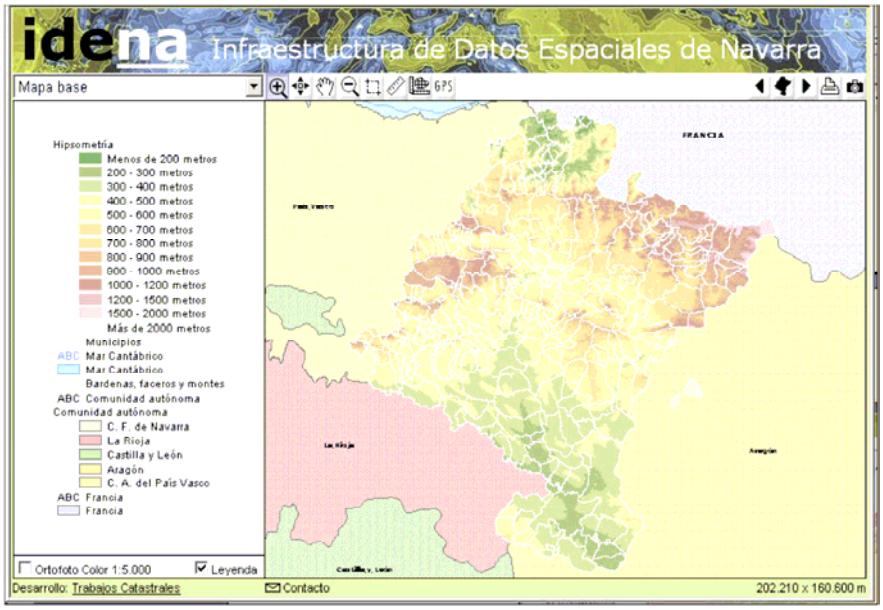


Figure 2. Map Viewer IDENA

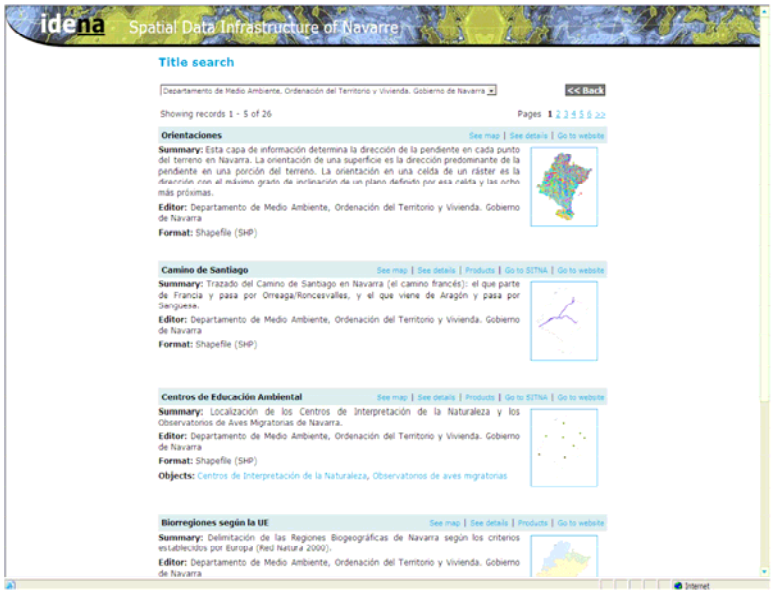


Figure 3. List of search results

- **Shop**: It will include the following tasks or elements: shopping cart, on-line downloads, physical product orders, payment, tracking of orders and authentication.
Through the telematic platform of the Government of Navarre, it will be possible to buy geographic information (with electronic signature, a payment method and Safe Electronic Communication). This utility is not operative in the IDENA portal jet.

4 METADATA IN IDENA

4.1 Definitions of Metadata

The most common definition of metadata is " data about data" . It describes the content, quality, condition and other characteristics on itself. It helps to locate and understand the spatial data available.

INSPIRE mentions that metadata must be maintained up-to-date by a person responsible for the data and must present a background compatible with the metadata standard ISO19115 (obligatory within INSPIRE). We will return to this question.

The main functions of metadata are the following:

- Organize the data to allow users to know it better and, as a consequence, use it, avoiding duplications and detecting errors. All this helps maintain the investment in the data.
- Facilitate the transfer of data by offering information on file formats, volume, localization, etc.
- Allow the on-line distribution of the data incorporating addresses to download files, either free or pre-paid, through descriptions on how to carry out these processes.

- Facilitate the search of data in a complex corporate database. Metadata standards allow a user to make a consultation, which is redirected to the different metadata catalogues registered in the main server. The result is a list of metadata from different servers.
- Avoid the erroneous use of data by including descriptions of how there were created, what goal and how they should be used.
- Insure the integrity of the data as well as its safety.

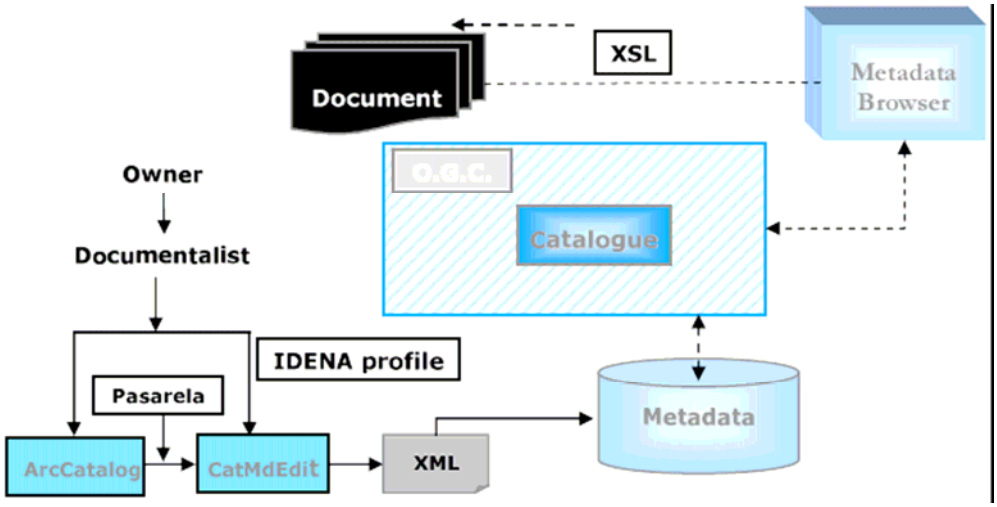
4.2 Standard of metadata used

The metadata can be shared among the different producers and users, so, it should be understandable for everyone. For this, metadata standards have been designed trying to create common points of view for the different actors involved in the spatial information making this documentation comprehensible for all those that " speak the same language" .

The International Organization for Standardization is a federation of national standardization organisms. They also receive the collaboration of international organizations, governmental or not. This organization is divided, internally, into technical committees in charge of elaborating standards for different themes. The committee that has elaborated this standard is the ISO/TC 211, designated to geographic information. Besides the metadata standard, there is an extensive list of other standards related to geographic information.

4.3 General Approach

Among the short-term objectives, we have already mentioned that offering the user the possibility of searching and identifying existing data, with precision, and reaching a basic level of interoperability that opens the IDENA data to any other SDI.



This figure shows the flow that the approach unites from the previous sections, and that we will now describe:

- In first place, is the metadata creation function for all existing data. This task has been carried out by the metadata administrator, according to the IDENA background (this will be commented, in detail, in later sections), combining the use of two tools: ArcCatalog and CatMDEdit (the first allows the extraction of some spatial attributes from the data such as format, number of objects, etc.; with the second, we can complete the metadata with more detail).
- Once the metadata registers has been created, we export from CatMDEdit to XML format (metadata exchange format), been stored and sent to become part of a metadata catalogue.
- The IDENA user can search according to defined criteria (Metadata Search Engine). These searches are done on the metadata catalogue, and the results are shown with a view, following the specific style of IDENA through the use of style formats (XSL). These style formats will be of two types: general and specific (first it shows the more important or highlighted metadata, and next all those established by the IDENA background).

4.4 Structure and exploitation model

The IDENA information warehousing has a layer-item structure. The layers (dataset) are groups of items fundamentally according to thematic criteria and information headlining (example: land registry will be the layer and the plots of land, buildings, etc. its items). From this point of view, the items are constituted as the essential element of the territorial information. However, when we talk about document cartographic series or metadata this division is not very clear as well as it is not when depth level to use is different according to layers.

The layers of available information are very wide types, from cartography in different scales, spatial planning, environmental resources, land registry, postal addresses, administrative divisions, agriculture, etc. and, of course, the raster images. There are some layers much well known than others, that means that we have to structure layers in different levels. Also it is necessary to structure the meta information in layers (or " parent" levels) and items (or " child" levels) with different levels of integration. The following figure shows some examples of existing dataset and the structure they have adopted. The metadata creation tool indicates the higher

hierarchical level each data belongs to and what are those that hang from each level. The general idea is offer through the catalogue client application all the integrated metadata (basic and detailed) for the data in any level of disintegration.

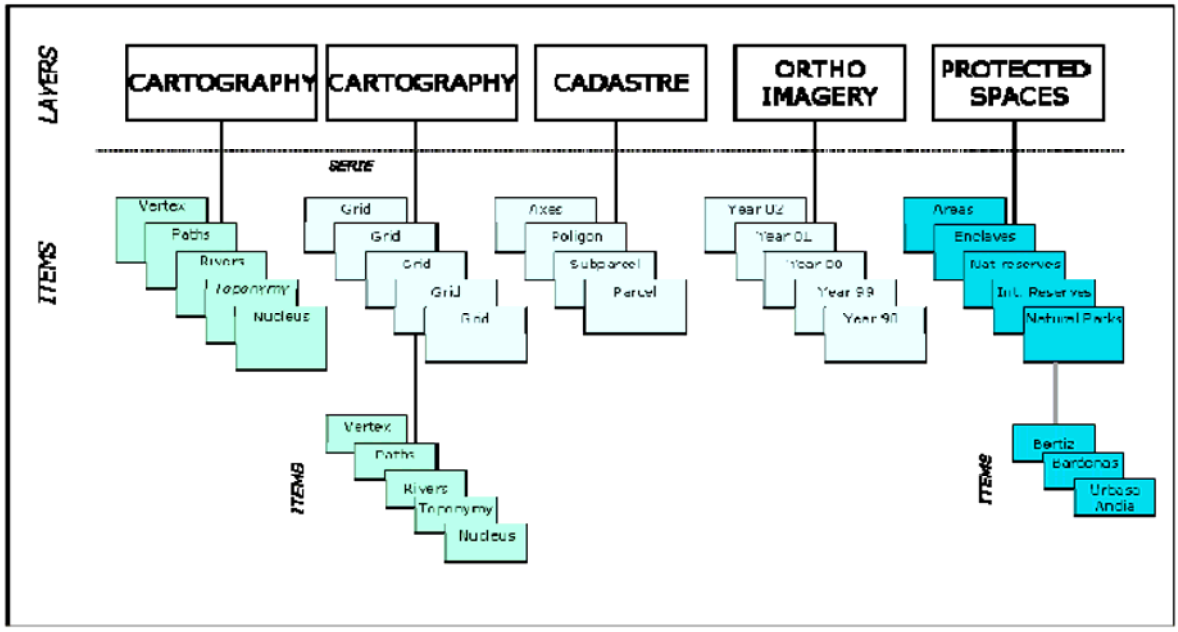


Figure 5: Structure of the warehousing of IDENA metadata (example for some data).

The search engine, according to the search criteria established by the user, offers the " layers" as results (see structure of the previous figure) that meet the criteria. Each layer offers information and access to all the included items.

4.5 Background description

The metadata described in the ISO19115 standard are very numerous and propose a very extensive documentation of the data that could lead to dejection and the risk of limiting creation advancement. The problem is, on one hand that it deals with a very general standard that should be useful in any discipline related to territory and in any country; on the other hand, that many of these elements will never be used.

For this reason, a more limited nucleus is proposed, in such a way that, although serving to collect basic information for more habitual searches, it avoids the need for excessive effort in the creation of metadata.

The SG NEM (Work Subgroup of the Spanish Metadata Nucleus) of the Superior Geographic Council, for the Infrastructure for Spatial Data of Spain (IDEE), has been intensively working on the definition of the NEM background.

NEM is the acronym for the Spanish Metadata Nucleus, a minimum group of metadata items, recommended for use in Spain when creating metadata for geographic data. That NEM is based on the nucleus of the ISO19115 metadata standard, in the Dublin Core Metadata, in the description elements of Quality and in other additional elements considered of interest in cataloguing.

The same way that a nucleus for the IDEE has been established, IDENA has established its own background adapted to its needs and based on the NEM.

IDENA Metadata Nucleus = Spanish Metadata Nucleus (ISO Core 19115 + Dublin Core + Quality + Additional NEM Elements) + Additional IDENA Elements

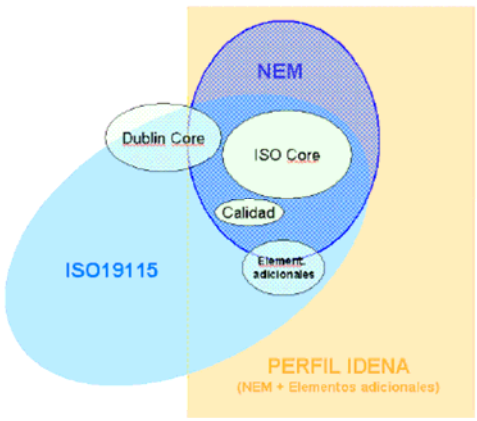


Figure 6: The background of IDENA metadata

The background of IDENA Metadata is:

- 22 elements of the ISO Core:
- 7 are obligatory, in other words, the metadata can not be considered valid if at least it does not collect information from these fields (marked as M in the table on the following page).
- 15 are optional (marked as O and C in the table on the following page).
- 4 additional elements from the Dublin Core without correspondence with any ISO Core element.
- 6 Quality elements. These elements are included to complete the element '/LQHDJH' of the ISO Core in order to determine the data quality.
- . 4 Additional elements of the NEM background.
- 6 Additional and specifically elements answering the needs from IDENA .

Finally, we have 36 elements from the NEM plus 6 elements specifically for IDENA.

Next, there is a more detailed revision of all the elements that configure the IDENA background.

4.6 Metadata Background

4.6.1 The ISO 19115[3] Core Elements

The international standard defines a group of very extensive metadata elements, but in the truth time, the group of elements used is quite small. However, it is essential to use a basic number of elements for the metadata. The next table shows the principal nucleus of the standard (the &RUH), the minimum necessary to document the data.

(M) are obligatory; (O) optional and (C) conditional.

Dataset title (M) (MD_Metadata > MD_DataIdentification.citation > CI_Citation.title)	**Spatial representation type** (OR) (MD_Metadata > MD_DataIdentification.spatia lRepresentationType)
Dataset reference date(M) (MD_Metadata > MD_DataIdentification.citation > CI_Citation.date)	**Reference system**(OR) (MD_Metadata > MD_ReferenceSystem)
Dataset responsible party(OR) (MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty)	**Lineage**(OR) (MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)
Geographic location of the dataset (by four coordinates or by geographic identifier) (C)	**On-line resource** (OR)
(MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or EX_GeographicDescription)	(MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Dataset language (M) (MD_Metadata > MD_DataIdentification.language)	**Metadata file identifier** (OR) (MD_Metadata.fileIdentifier)
Dataset character set(C) (MD_Metadata > MD_DataIdentification.characterSet)	**Metadata standard name** (OR) (MD_Metadata.metadataStandardName)
Dataset topic category (M) (MD_Metadata > MD_DataIdentification.topicCategory)	**Metadata standard version**(OR) (MD_Metadata.metadataStandardVersion)
Spatial resolution of the dataset (OR) (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) (OR) (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent or EX_VerticalExtent)	**Metadata date stamp** (M) (MD_Metadata.dateStamp)

Table 1: ISO Core[4]

4.6.2 Dublin Core[5] Elements

There are 4 elements included in the Dublin Core standard that do not have correspondence with the elements collected in the ISO Core (although they could find correspondence with elements of the ISO19115).

The Dublin Core is a standard for the creation of metadata for general resources, not only for geographic information.

[3] International Organization for Standardization, 2003. " International Standard ISO19115. Geographic information – Metadata" . www.iso.org

[4] Source: The OpenGISTM Abstract Specification. Topic 11: OpenGISTM Metadata (ISO/TC 211 DIS 19115). Version 5.

[5] Dublin Core Metadata Initiative, 2003. "Using Dublin Core – The Elements" . http://dublincore.org/documents/usageguide/elements.shtml

This standard contains 15 elements, 11 of which have correspondence with the elements of the ISO Core, but not with the other 4. These 4 elements are:

- Contributor: Contributor (would correspond with the element Credits of the ISO 19115): Recognition of those people that have contributed, one way or another, in the creation or modification of the data (through economic means, work, support, etc.).
- Relation: Relation (would correspond with the element ‚Integration information of the ISO 19115): Relationship to related resource.
- Rights Management (would correspond with the element ‚Information on legal Constrictions / Restrictions): Information to offer: restrictions on access, use, others restrictions and legal restrictions.
- Type: (would correspond with the element Hierarchy Level): Nature or type of content of the resource. The type includes terms that describe general categories, functions, types or integrated levels for the content.[6]

4.7 Quality elements

The ISO19115 standard collects information on the quality reports carried out on the data (determines what quality controls have been performed on the data).

A dataset can have passed one or various quality reports, which can be of the following types:

- Commission: Excess of data in the file.
- Omission: Lack of data in the file.
- Logical consistency: Degree of adherence to the logical rules of data structure, attributes and relationships.
- Position accuracy: Accuracy of the position of the objects.
- Thematic accuracy: Accuracy of attributes and relationships. Determines the fidelity of the values of the attributes, such as name, length, class code, population, etc., assigned to the objects with respect to the true value of the characteristic that they present.
- Temporary accuracy: Accuracy of temporary attributes and relationships.

Each one of these types of determined reports should specify: the name of the measurement, the description of said measurement and the result (value obtained).

Besides the quality reports, there are other metadata that also complement the quality information:

- Level: Hierarchy level of the data for the quality control.
- Declaration (state): General explanation of the knowledge of the data producer of the data group lineage.

4.8 Additional NEM elements

Besides the earlier mentioned elements, the NEM includes 3 others, considered important for the information provided with the data:

- Key words of subject and place. Word/s used to describe the subject or place corresponding to the dataset. Name of the formally registered thesaurus or a similar authority key words source.
- Form of presentation: Way that the resource is presented.
- Intention: Summary of the intentions that resource has been created.
- Specific use: Ways the resource is being used.

4.9 Additional IDENA elements

The previously commented elements correspond with the Spanish Metadata Nucleus (NEM); however, in the case of Navarre, they have established, for the moment, 6 additional elements, because they have been considered essential (distribution of information) or in some cases, interesting to complete the previously known information. These elements are:

- Alternative title: Short name, name in another language or acronym for which the cataloged information is known.
- Parent Identifier: identification number of the metadata file from which this metadata is a part (child).
- Frequency of maintenance and update.
- State: State of the resource, if it is completed, in progress, obsolete, etc.
- Supplementary information: Any other descriptive information on the dataset.
- Distribution. This section includes all the information that the standard has referring to the distribution of metadata, except for that already contained in the Core (On-line Resource and Distribution Format).

[6] This element is useful to determine if a resource is an integrated data or not, and if it is, what level of integration it presents (if it is a series or a page, a dataset or an item, an attribute...).

5 INTEROPERABILITY WITH OTHER IDES

Another of the INSPIRE requirements deals with the interoperability between different SDI. At this moment IDENA may interoperate with any SDI that use the OGC (Open Geospatial Consortium) standard, such as IDEE (SDI of Spain), INSPIRE (European SDI), GEODATA.GOV (North American SDI)...

The system has been designed to compare data across different regions, and visualize the datasets of Navarra close to any other region. Moreover, the advanced user could combine Navarra datasets with their own datasets by downloading those spatial data they need and work with all of them in more detailed analysis

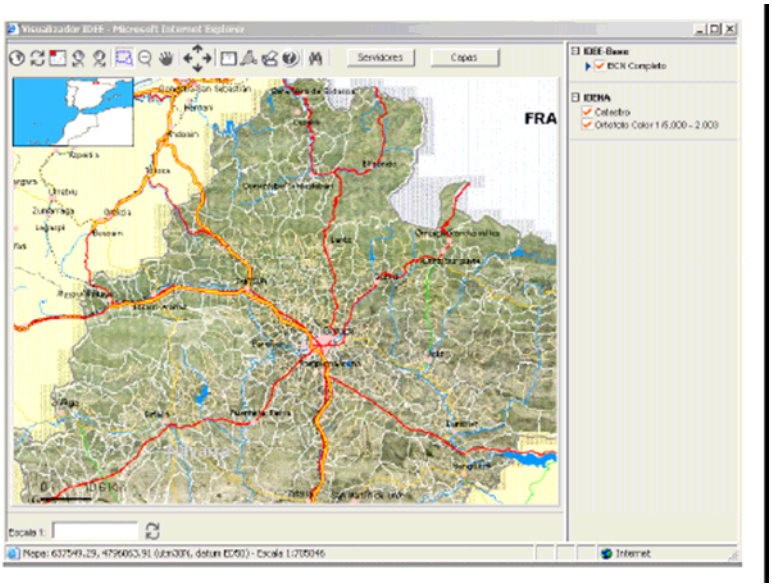


Figure 7. Visualization of IDENA information layers in the Spanish SDI Viewer

6 CONCLUSIONS

This article has focused, fundamentally, on the following aspects:

- IDENA architecture and its components (see figure 1).
- Resolution of the metadata work flow within the IDENA project, as well as on its different components (see figure 4).
- Establishment of the information structure (see figure 5) to create a better comprehension for users (simplify and structure search results).
- Definition of the IDENA metadata background.
    - Interoperability (see figure 7).
    -

Finally we could conclude:

Navarre and its Territorial Information System (SITNA) determinedly believe in the SDI world and, for this reason, have started a plan for the documentation of their entire information warehouse, which is currently in a fairly advanced phase.

IDENA is the answer Navarra need to comply with requirements from INSPIRE and Spanish SDI. It provides consistence to Territorial Information System of Navarra (SITNA), completing services profiles offered. It has been developed in a very short time: just six month from order to official presentation. That shows the great versatility of SITNA and the great technological answer capability from our company TRACASA.