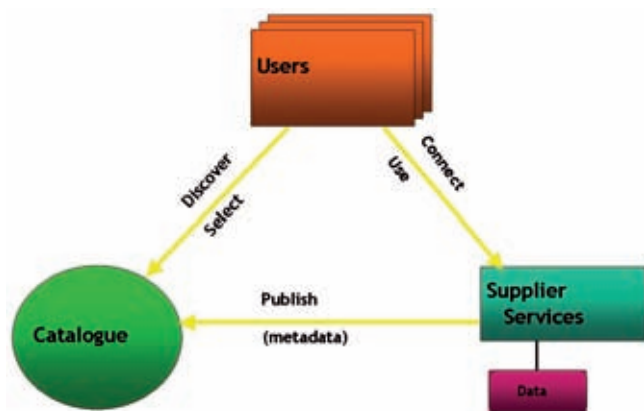


Part 2: Metadata

Standards in Practice

When building a geo-information infrastructure, specialists tell us that careful attention should be given to metadata. But what is metadata, how does it work, and what standards apply?

By Huibert-Jan Lekkerkerk



Use of metadata for searching, discovery and use of (geographic) datasets in a geo-information infrastructure.

There is much discussion about the exact definition of metadata. The most common remark is that metadata is information about information. This, however, does not solve the problem. Say we want to describe a building. To the land surveyor who is contracted to survey the building, the main body of data he collects are the coordinates. All other data such as the quality, specifications etc. are metadata. But the project developer will probably need additional data such as build quality, and will consider this the main body of data.

Metadata is usually collected with a specific purpose in mind. Common purposes are the discovery and use of datasets. The metadata is then used to answer questions such as the 'who, what, where, when and why' of a dataset. This article will only consider metadata used for the discovery and use of (geographic) datasets. It does not consider metadata for single objects; these will be discussed later in the series.

Metadata in Geographic Practice

The use of metadata is as old as the publication of geographic data. As long as charts have been published, they have had extensive legends.

With the introduction of digital data, the metadata concept was forgotten. Many organizations switched to automated systems in which it was

not possible to store metadata with the dataset. The metadata was then added to the legend in the publishing phase when the paper charts were printed. With the introduction of GIS, this changed again. Information was now acquired digitally and, after conversion, stored in the GIS as a digital product, usually with a limited set of metadata.

When, after a period of time, the dataset then needs to be used or published, it is difficult to obtain information about the purpose of the original dataset. A number of programs are available to create and manage metadata, but they are usually only used for the central archive and not for the local, specific datasets within an organization.

Metadata Standards

If missing metadata poses a problem within a single organization, then imagine what it is like within a geo-information infrastructure. Within an organization one can usually obtain the metadata, albeit with some difficulty, but when publishing to the outside world this is no longer feasible.

For this reason the EU INSPIRE directive has laid down a set of implementing rules for the use of metadata. In these implementing rules an obligatory set of core metadata is defined for specified government-managed datasets. Individual organizations can then choose to extend this minimum set to cover the needs within a specific organization or country.

The INSPIRE metadata core set and most national core sets are based on the ISO 19115 metadata standard. This standard consists of an extensive library of metadata elements of which a few are marked as obligatory. The INSPIRE

Metadata

What it is for: Describing the content and origin of datasets with respect to searching, discovering and using these datasets.

Relevant standards

- ISO 19115
- INSPIRE Implementing Rules for metadata
- Dublin Core

Technical implementation

- XML as exchange format
- ISO 19139 when using metadata in geo-services

Legal basis: INSPIRE Directive (for specific datasets managed by governments)

implementing rules have selected additional elements from the ISO 19115 library and made these obligatory as well, thus creating a new core INSPIRE set. Besides the ISO 19115 metadata standard for geographic information, there is another important metadata standard, Dublin Core (DC). This is a limited, general set of metadata that can be used for a myriad of datasets and information sources. Dublin Core is generally used within governments and is not exactly equal to the ISO 19115 elements. The two standards can, however, be mapped onto each other in such a way that both can be used in a combined environment.

How it Works:

The principle behind a metadata standard is very simple. Users use specific software to enter the necessary list of metadata for a specific dataset. Examples of metadata elements from the INSPIRE core set are:

- Title of the dataset
- Temporal elements, for example the acquisition and publication of the dataset
- Geographic boundary of the dataset
- Subject of the dataset
- Keywords
- Contact details for the responsible organization
- Abstract of the contents
- Details about the web service where the dataset can be obtained / used

If the metadata software is integrated with the GIS system, some elements such as geographic boundary are automatically generated from the dataset. Other elements will remain constant within an organization and only need to be filled out once.

Metadata Catalogue

After being created, the metadata can be published to a so-called catalogue. Users can search through the catalogue for datasets that, suit their purpose. The searching process is not unlike that

```

<gmd:language>
  <gmd:languageCode codeList="LanguageCode" codeListValue="eng">eng</gmd:languageCode>
</gmd:language>
<gmd:contact>
  <gmd:CI_ResponsibleParty>
    <gmd:individualName>
      <gco:CharacterString>Huibert-Jan Lekkerkerk</gco:CharacterString>
    </gmd:individualName>
    <gmd:organisationName>
      <gco:CharacterString>Geo Informatics</gco:CharacterString>
    </gmd:organisationName>
    <gmd:role>
      <gmd:CI_RoleCode codeList="#CI_RoleCode" codeListValue="author">author</gmd:CI_RoleCode>
    </gmd:role>
  </gmd:CI_ResponsibleParty>
</gmd:contact>
<gmd:dateStamp>
  <gco:Date>2007-01-12</gco:Date>
</gmd:dateStamp>
<gmd:referenceSystemInfo>
  <gmd:MD_ReferenceSystem>
    <gmd:referenceSystemIdentifier>
      <gmd:RS_Identifier>
        <gmd:code>
          <gco:CharacterString>ETRS89</gco:CharacterString>
        </gmd:code>
        <gmd:codeSpace>
          <gco:CharacterString>EPSG</gco:CharacterString>
        </gmd:codeSpace>
      </gmd:RS_Identifier>
    </gmd:referenceSystemIdentifier>
  </gmd:MD_ReferenceSystem>
</gmd:referenceSystemInfo>

```

Relationships between the various standards and sets of metadata for geographic information (source: Geonovum; adapted).

of an Internet search engine, the main difference being that one can define geographic search parameters as well. Once a dataset is discovered, the information on the web service is used to get the dataset. Most geographic portals can display the information in a viewer so that users think they have direct access to the data from the portal. The service can, however, also be used to view the dataset in one's own GIS if the publisher permits.

Technical Format

In order to publish metadata in a standardized way, it is not only important to define the con-

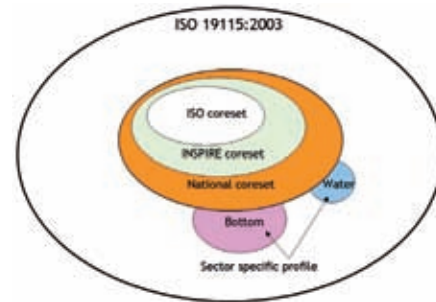
tent of the metadata file but also to define the technical format. Both Dublin Core and ISO 19115 use XML as the technical format. The advantage of using XML is that the metadata files can be checked with regard to integrity, structure and missing elements. XML is furthermore a

W3C (Internet) standard and can be read easily by a number of software packages.

The exact way in which ISO 19115 metadata elements need to be put into an XML format that is suitable for web services can be found in another standard: ISO 19139. Although XML can be generated with any generic text editor, it is advisable to use specific management software that uses information from the internal geographic database and converts it to XML.

Legal Basis

For a number of government-managed data categories INSPIRE states that metadata for



Detail of an XML metadata file that conforms to INSPIRE specifications.

new datasets has to be available by 2011. Beyond this legal basis, it is always necessary to have metadata available if datasets are published to the outside world. The mandatory (core) set of metadata elements is relatively small. As such implementation is not a matter of difficult and extensive software implementation but more one of backlogged datasets. In the past, many organizations have invested in numerous datasets that lack adequate metadata. Bringing this up to date will cost a lot of time (and therefore money).

Huibert-Jan Lekkerkerk

(hlekkerkerk@geoinformatics.com) is editor of *GeoInformatics* and a freelance writer and trainer in the field of positioning and hydrography.