Metadata

Metadata Concepts

What is "Metadata"?

Metadata is “data about data”.

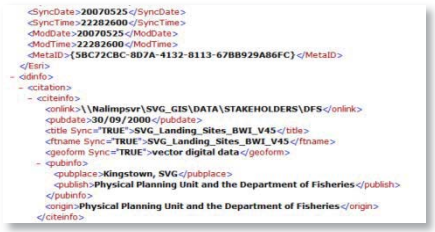
> “Metadata is a summary document providing content, quality, type, creation, and spatial information about a data set. It can be stored in any format such as a text file or database record. Because of its small size compared to the data it describes, metadata is more easily shareable. By creating metadata and sharing it with others, information about existing data becomes readily available to anyone seeking it.
> Metadata makes data discovery easier and reduces data duplication.”
> Metadata and GIS,
> ESRI, White Paper, October 2002

- Metadata is a formal data documentation (information about datasets) and it is critical to data discovery
- A metadata record is a set of information fields, usually presented as an XML document, which captures the basic characteristics of data as an information resource
- It represents the ***who***, ***what***, ***when***, ***where***, ***why*** and ***how*** of the resource
- **Who** collected and who distributes the data?
- **What** is the subject, projection of the data?
- **When** was the data collected?
- **Where** was the data collected?
- **Why** was the data collected (what is the purpose)?
- **How** was the data collected? **How** should it be used?
- **How** much does it cost?

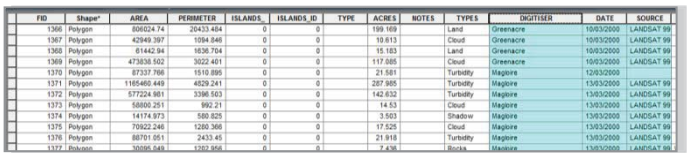
Metadata - Stored as…

Metadata can be stored as:

- an extra file (e.g an XML)

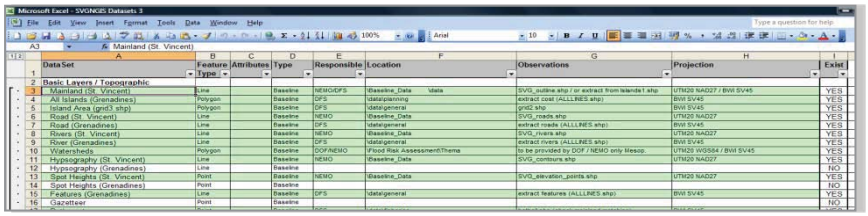


- within a file (as an attribute)



	FID	Shape*	AREA	PERIMETER	ISLANDS_	ISLANDS_ID	TYPE	ACRES	NOTES	TYPES	DIGITISER	DATE	SOURCE
	1366	Polygon	806024.74	20433.484	0	0		199.169		Land	Greenacre	10/03/2000	LANDSAT 99
	1367	Polygon	42949.397	1094.846	0	0		10.613		Cloud	Greenacre	10/03/2000	LANDSAT 99
	1368	Polygon	61442.94	1036.704	0	0		15.183		Land	Greenacre	10/03/2000	LANDSAT 99
	1369	Polygon	473838.502	3022.401	0	0		117.085		Cloud	Greenacre	10/03/2000	LANDSAT 99
	1370	Polygon	87337.766	1510.895	0	0		21.581		Turbidity	Magiore	12/03/2000	
	1371	Polygon	1165460.449	4829.241	0	0		287.985		Turbidity	Magiore	13/03/2000	LANDSAT 99
	1372	Polygon	577224.981	3396.503	0	0		142.632		Turbidity	Magiore	13/03/2000	LANDSAT 99
	1373	Polygon	58800.251	992.21	0	0		14.53		Cloud	Magiore	13/03/2000	LANDSAT 99
	1374	Polygon	14174.973	580.825	0	0		3.503		Shadow	Magiore	13/03/2000	LANDSAT 99
	1375	Polygon	70922.246	1280.368	0	0		17.525		Cloud	Magiore	13/03/2000	LANDSAT 99
	1376	Polygon	88701.051	2433.45	0	0		21.918		Turbidity	Magiore	13/03/2000	LANDSAT 99
	1377	Polygon	30095.649	1282.956	0	0		7.436		Rocks	Magiore	13/03/2000	LANDSAT 99

- as a single extra database documenting summaries

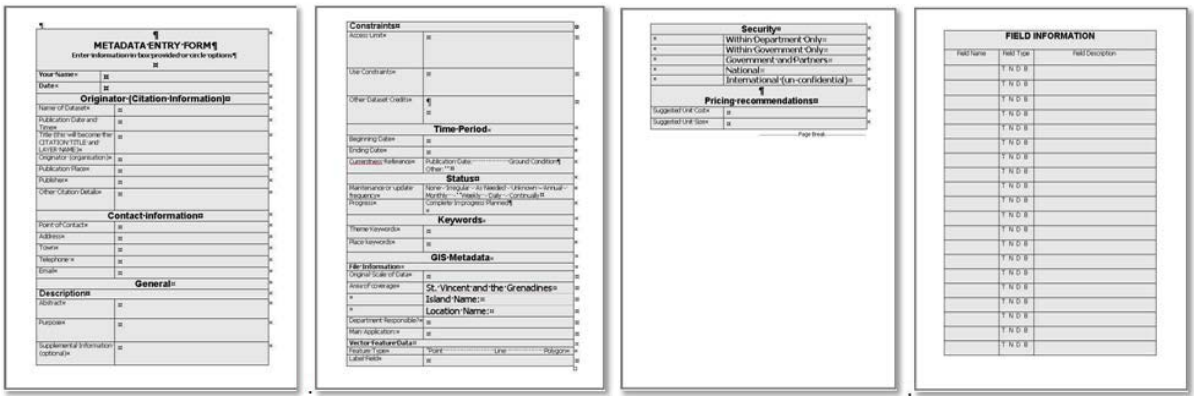


Metadata Elements - Category Overview:

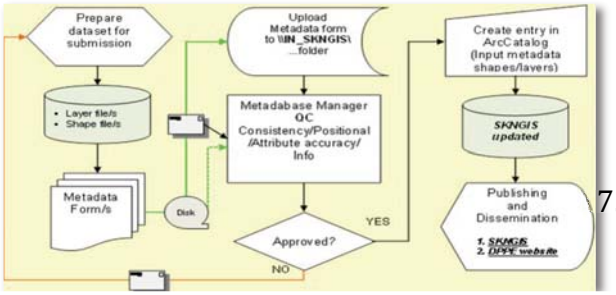
- **Identification** - tittle, originator, description, area covered, currency, rule for acquiring or using the data, restrictions, liability
- **Data Quality** – positional and attribute accuracy, source of info, method used for producing the data
- **Spatial Data Organization** – method used to represent spatial data (vector, raster, etc.)
- **Spatial Reference** - projection, type, scale, size
- **Entity and Attribute** - entity type, attributes, domain for attribute values
- **Distribution** - contact for distribution, available formats, how to obtain dataset, physical media, fees, etc.
- **Metadata Reference** - currency of the metadata information, responsible party

Metadata production

- Fill in metadata entry form to best of ability



- Provide new dataset(s) and include symbology for layer file



- Submit for QC of lyr / shp files.
- Include transformation to projection if necessary
- Provide updates to metadata as dataset changes

Concepts and use in search and documentation

Main uses of Metadata

- Organize and maintain an organization's internal investment in spatial data
- Provide information about an organization's data holdings to data catalogs, clearinghouses, and portals (it is the entry point of an SDI and for each SDI node)
- Provide information to process and interpret data received through a transfer from an external source

Benefits of using Metadata

- Facilitates data maintenance
- Facilitate data discovery
- Enables reuse of data
- Informs potential users of inappropriate uses of a dataset

Challenges of using Metadata

- Exchanging records -> Use standards!
- Creating metadata takes time and patience -> Use templates and tools that automatically extract or set metadata (like *GeoNode*). Imperfect metadata is better than none
- Maintaining metadata is essential, especially for Web services and downloads (beware of URL changes!) -> set up processes or tools for automatic checkups

Metadata standards

- Some standards defines the metadata content (fields, values, mandatory v. optional)
    - The most comprehensive is ISO 19115
- From this standard, custom profiles can be defined for specific countries, themes, etc.
- Other standards define appropriate exchange formats
    - The most accepted is ISO 19139, which defines an XML-based format

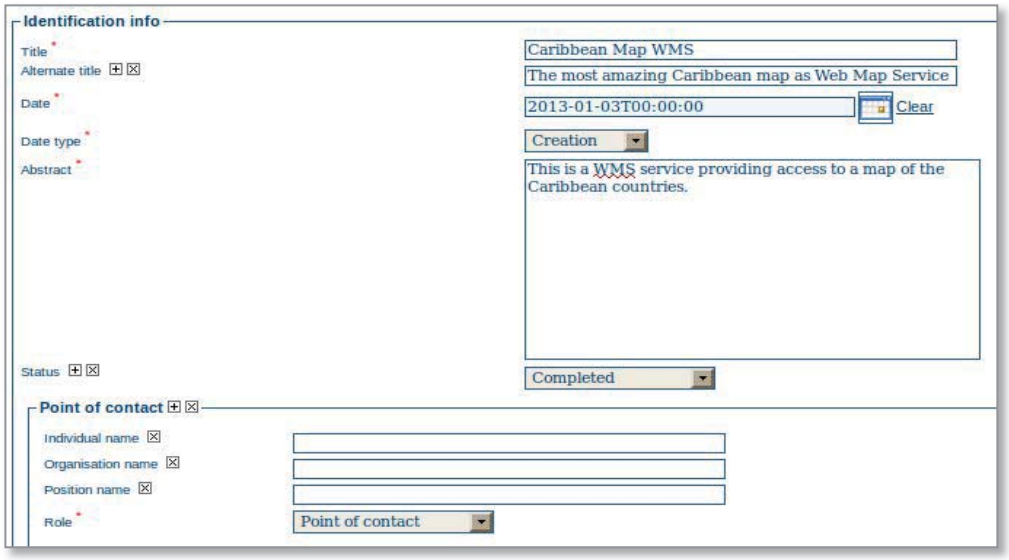
ISO 19115

The information included in the ISO 19115 standard supports the following use:

- **discovery** - data needed to identify and locate the sets of geographic data that exist for a geographic location;
- **access** - data needed to acquire an identified set of geographic data;
- **fitness for use** - data needed to determine if a set of geographic data meets the user's need and to support the user in applying the geographic information appropriately; and
- **transfer** - data needed to obtain a copy of a set of geographic data.

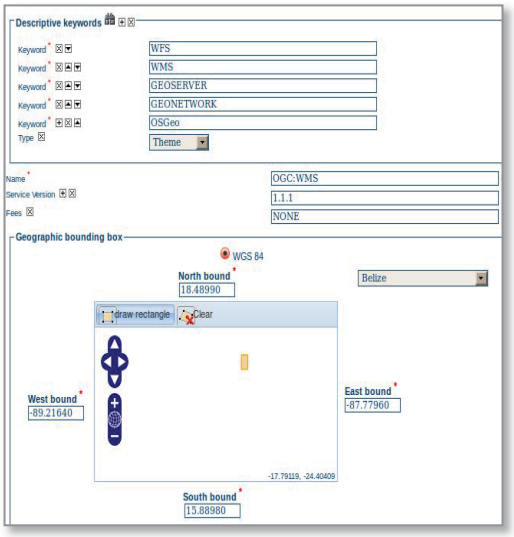
Metadata Content

Identification: Basic information about the data set. Examples include the title, the geographic area covered, currency, and rules for acquiring or using the data



Information in abstract: data model

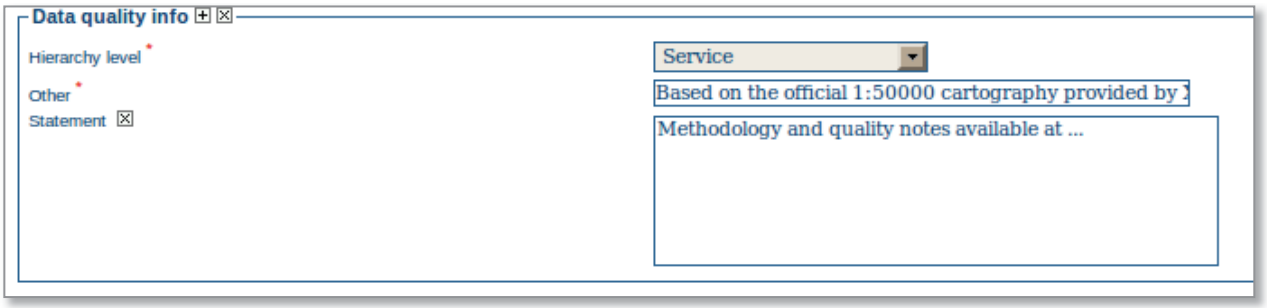
- Entity types and their attributes and the domains from which attribute values may be assigned
- The names and definitions of features, attributes, and attribute value



Data quality

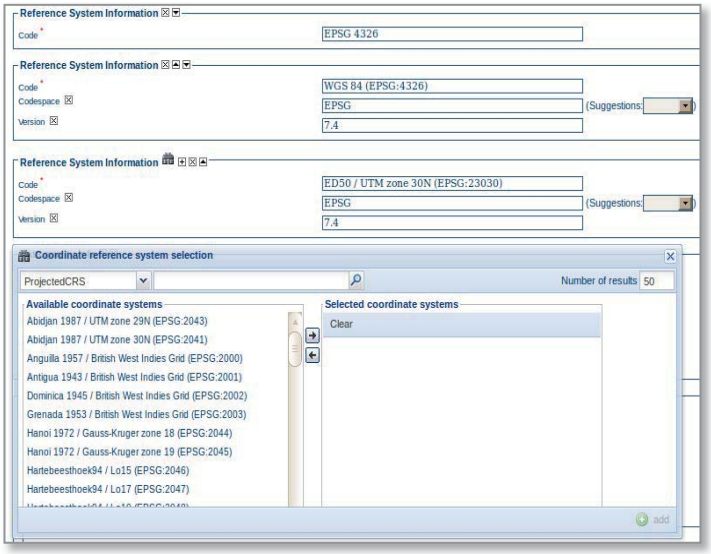
- Positional and attribute accuracy
- Completeness and consistency,
- Sources of information
- Methods used to produce the data
- Appropriate uses (scales)

An assessment of the quality of the data set. Examples include the positional and attribute accuracy, completeness, consistency, the sources of information, and methods used to produce the data



Spatial Reference

- Description of the reference frame,
- Means of encoding,
- Coordinates in the data set,



Examples include the name of and parameters for map projections or grid coordinate systems, horizontal and vertical datums, and the coordinate system resolution

Entity and Attribute

- Information in abstract: data model
  - Entity types and their attributes and the domains from which attribute values may be assigned
  - The names and definitions of features, attributes, and attribute value

Information about the content of the data set, including the entities types and their attributes and the domains from which attribute values may be assigned.

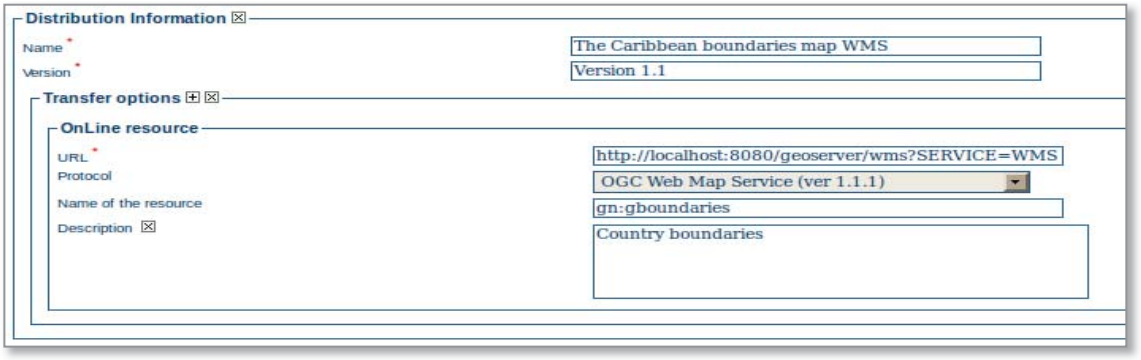
Examples include the names and definitions of features, attributes, and attribute values

Distribution information

- How to access the information?
- Multiple transfer options are possible
- Contact for physical media
- Fees

Information about obtaining the data set.

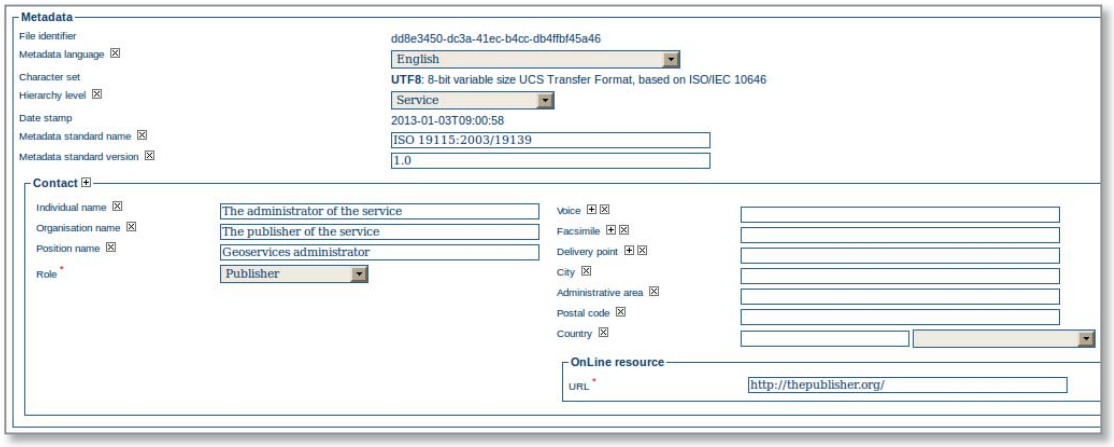
Examples include a contact for the distributor, available formats, information about how to obtain data sets online or on physical media (such as cartridge tape or CD-ROM), and fees for the data



Metadata Reference

- Language, date, update status
- Who is responsible for the metadata?

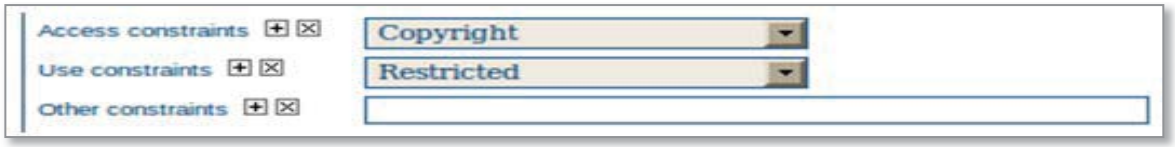
Information on the currency of the metadata information and the responsible party. Examples include currency and information about the organization that provided the metadata



Access and use constraints

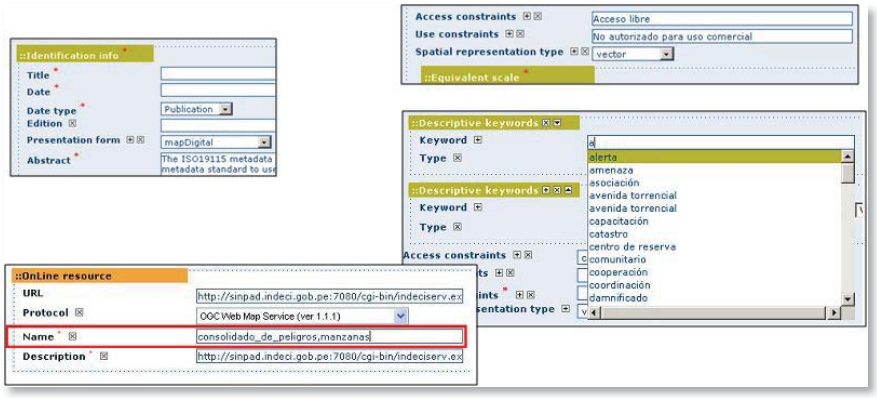
- Use constraint statements can be crafted to express scale, geographic, or temporal limitations to the data
- Liability statements should be written by legal staff to ensure that the legal requirements for use of the data are fully outlined

In general, it is far better to publish your dataset with limitations within your metadata than to later attempt to generate them in response to an inquiry or lawsuit

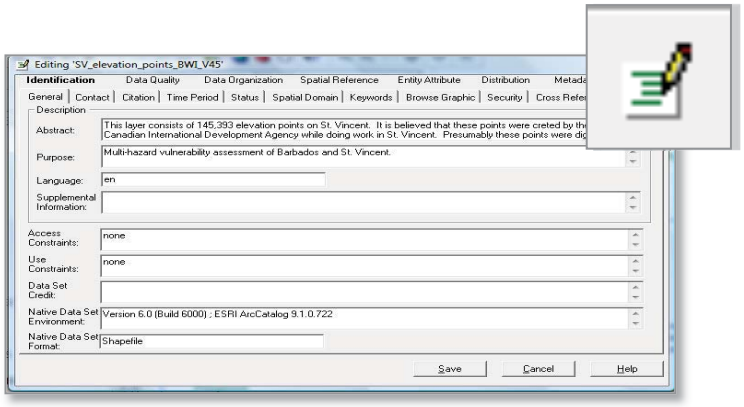


Metadata Tools

As seen in GeoNetwork



As seen in ESRI's ArcCatalog



Conclusions

- **Metadata is important**
  - To know what data exist
  - To know the quality of the data
  - Who is responsible for the metadata

- **You can use it to**
  - Search for suitable data

- **Help the management of the NGIS**
  - Using the metadata for searching and querying
  - Add new metadata using the forms
  - Help the Metadatabase Manager to improve the information about the NGIS

Metadata References

- Metadata Guidelines - Land Information Council of Jamaica - 2006
  http://www.licj.org.jm/licj/documents/metadata.html

- International Standard ISO/FDIS19115, **Geographic Information - Metadata**

- ESRI White Paper (October 2002), "**Metadata and GIS**" - 4. Vermeij, Bert, "implementing European Metadata Using ArcCatalog", ESRI ArcUser, July - September
  www.esri.com/library/whitepapers/pdfs/metadata-and-gis.pdf