What kind of data is needed to identify climate impacts?
How can data be managed and organized through data catalogues?

Metadata Standards, Data Catalogues and Data Requirements for the Identification of Climate Impacts

A Desk-Study

Prepared by:
Martin Dresen
geoSYS, Nansenstr. 17, 12047 Berlin, Germany
martin.dresen@geosysnet.de

Commissioned by
The GIZ-project
"Inventory of Methods for Adaptation to Climate Change"

September
2011

Table of Contents

List of Figures and Tables

1. Introduction

The International Climate Initiative of the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) has commissioned the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) to develop an inventory of methods and tools for identifying and analyzing climate change impacts and prioritizing adaptation needs and measures in cooperation with partners from Indonesia, Mexico, Tunisia and the Philippines.[1]. This study seeks to provide a knowledge base for the project's component II that aims to promote models for strengthening access to climate data and information with the objective to improve the quality of climate impact modeling and vulnerability assessments.

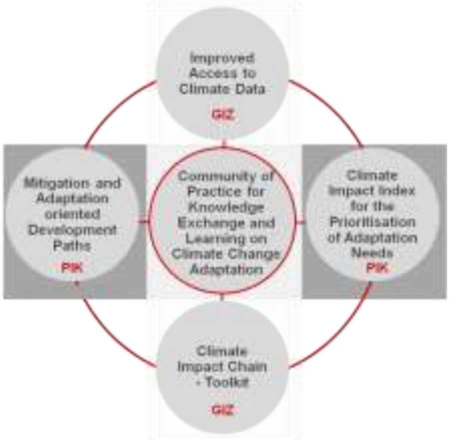


Figure 1: Project structure and working packages (GIZ 2011, Project description)

When dealing with the need for better access to climate information sufficient data integration strategies are needed to address the following key issues:

- **Where to find data?**

A lot of the data that is needed for climate impact analyses in developing or emerging countries is nonexistent, has not been structured or backed up – not to mention the often lacking storage and publishing tools or strict data policies in different countries. Therefore an inquiry about data existence and availability is very often necessary. Another obstacle often is the given structure of the data, especially for climate change issues, where data also have a temporal scale in addition to the spatial resolution. The data structure, the models or the storage might have changed over the years. Data might be corrupted, lost or stored as analog data. But even data published on the Internet are not inherently discoverable using standard web search capabilities because they are generally embedded within files or databases, the contents of which cannot easily be queried or catalogued by web technologies that are used by all major web search engines[2].

- **How to find comparable or integrable data?**

The syntactic and semantic heterogeneity of data from different sources makes data integration very difficult. Additionally, data are rarely annotated with sufficient attribute information or metadata, to make their interpretation unambiguous by investigators other than those who digitized or collected the data. The quality of data or data descriptions cannot be easily detected or evaluated. This also applies to different map scales and/or time series of data. A global dataset for instance cannot be used for local or regional studies, as it is usually based on generalized data, which is not precise enough for such studies.

---

[1] BMU 2011

[2] National Research Council 2010

- **How to organize the information about data and data sources?**

The manifold collectors and publishers of data often belong to different sectors, which do not necessarily have a focus on spatial questions. In this context deviating structures have emerged according to data management. Despite international standardization efforts there are no common practices or rules for the organization, administration, storage or query of data or information about data. This especially applies to interdisciplinary and transnational questions that arise in relation to climate change.

The study consists of two chapters that cover the main objectives: Chapter 2 answers the questions of what data is needed based on case studies from Tunisia and Indonesia. In order to analyze these data requirements the study uses the concept of climate impact chains. A climate impact chain is a general representation of how a given climate stimulus propagates through a system of interest via the direct and indirect impacts it entails[3]. Different climate impact studies from Indonesia and Tunisia will exemplify the data needs and the requirements and availability of data. Chapter 3 highlights how data can be organized by using data catalogs, metadata and standardizations. It is shown how access to data relevant for climate impact and vulnerability assessments can be improved. Finally examples from the EU and the US illustrate the possibilities of standards, catalog systems and the use of web-based information systems for data research and retrieval. At the end of the study we provide a list of hyperlinks, links to successful examples, projects and data catalogs.

2 What kind of data is needed?

In the following Figure a climate impact chain illustrates the relationship between climate stimuli and climate impacts, such as sea-level-rise and migration. Although the example seems to be relatively simple according to the impact chain, it is much more difficult when related to the data requirements. The overall impact chain implicates comprehensive data requirements to enable for example the analysis, evaluation or prediction of climate system vulnerability or the development of adaptation strategies.

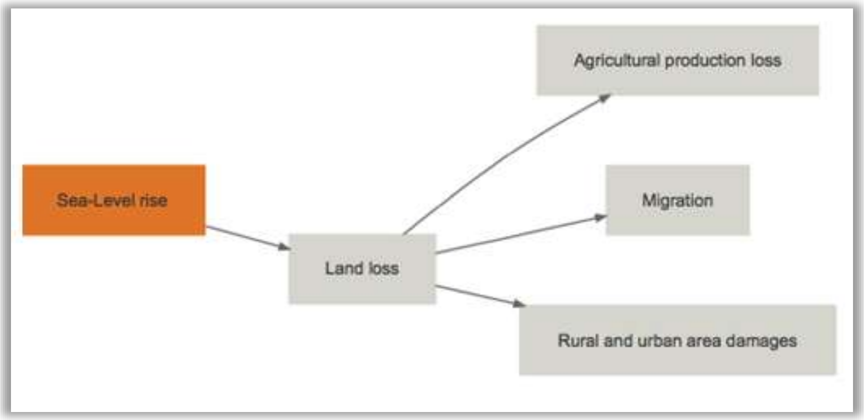


Figure 2: Exemplary Climate Impact Chain (http://cigrasp.pik-potsdam.de/about/impactchains)

[3] http://cigrasp.org

With respect to the example shown above the climate impact chain consists of three parts, namely the climate stimulus with sea-level-rise, the direct impact land loss and the different indirect impacts agricultural production loss, migration and rural and urban area damages. The starting point for any project related to climate change issues may include the preparation of an impact chain for analyzing possible causes and impacts and their relation to each other. The data question on the climate stimulus data is only briefly elaborated, as these are based on complex climate models, but existing model data will be used for the analyses. The visualization and evaluation of a 1 meter sea-level-rise for example requires a comprehensive digital elevation model (DEM) to derive inundation areas. Detailed land use / land cover (LULC) data and special statistical information about crop, yield and productivity of the agricultural areas to analyze the agricultural production loss is also needed. For the detection of migration, socioeconomic data has to be integrated, while for the evaluation of possible rural and urban damages data about infrastructure (roads, buildings etc.) is required.

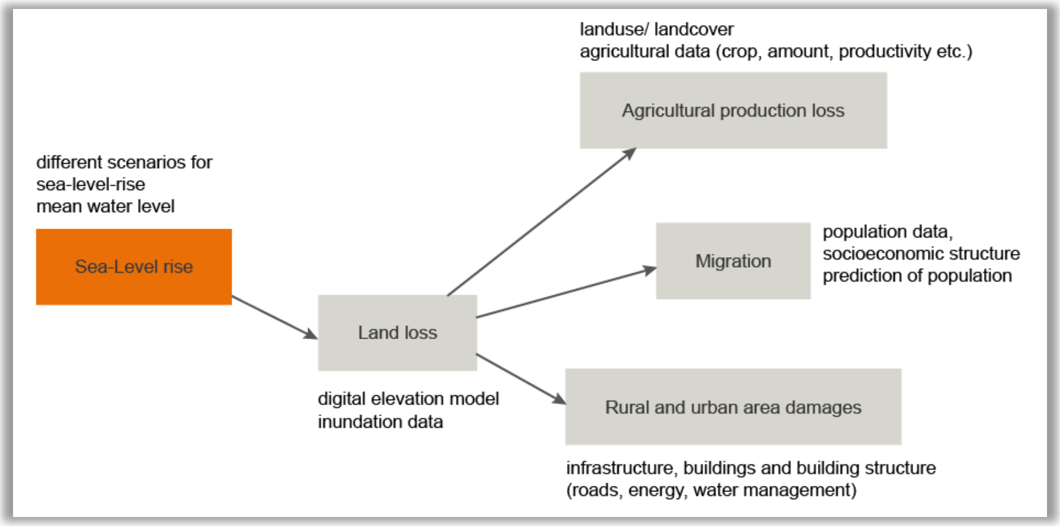


Figure 3: Data Requirements for Exemplary Climate Impact Chain

2.1 Evaluation of different case studies

The evaluation of data requirements for the analysis of climate impacts and estimation of possible adaptation measures are accomplished by interpreting the results from different studies from Indonesia and Tunisia. First the type of data that was required is highlighted before examining which data were available in sufficient quality. Additionally it will be checked, if problems occurred in relation to the use of data or the adopted spatial resolution and map scales[4].

[4] GTZ 2010d

2.1.1 Tunisia

The first case study from Tunisia examines the patterns of climate change vulnerability in the agriculture and water sectors in the southern region of the country by the example of olive production in the governorate of Médenine.

The Tunisian study was divided into two phases that comprised of an identification process followed by the organization of a workshop to discuss the results and to re-adjust the identified patterns. We focus on the first phase, as it discusses the data requirements for the study. For analyzing climate impacts, several datasets from biophysics, socio-economic and technological factors were considered which influence agricultural productions. Finally, olive production was used for the case study because of the sufficient availability and access to applicable statistical data.

The study's methodological approach included the preparation of a geodatabase with different socioeconomic and environmental datasets, which are considered to be important for the overall impact analysis. The datasets (climate, socioeconomic, natural resources, water, soils, ecological data, and infrastructure) were implemented on a regional basis (governorate level).

Table 1: Used data and scales in Tunisian project[5]

	Spatial scale	Temporal scale	Data sources
Climate			
Daily rainfall	Rain gauges of: Houmet Souk, Médenine, Sidi Toui, Bir Soltane, Béni Khédache	Available data from 1973 to 2008	World climate data : www.tutiempo.net/en/ National climate data : www.meteo.tn IRA's collected data
Daily temperature (Min, Max)	Weather stations of : Houmet Souk, Médenine,	Available data from 1973 to 2008	World climate data : www.tutiempo.net/en/ National climate data : www.meteo.tn IRA's collected data
CC projections for 2020 and 2050	-	-	We used the projections made for the national CC study in 2007 which adopted the HADCM3 model
Soil			
Texture	Governorate (1/200.000)	-	Soil map (Agriculture map) (MARH, 2002)
Depth	Governorate (1/200.000)	-	Soil map (Agriculture map) (MARH, 2002)
Topography	Governorate (1/200.000)	-	DEM (Agriculture map) (MARH, 2002)
Land use			
Current olive land use map	Governorate (1/200.000)	-	Agriculture map, (MARH, 2002) Updated with Google map and CRDA Médenine

However, the database "…has not yet been fully applied and used by the concerned agencies because of the difficulties in gathering suitable complete input data and ensuring their update continuously"[6]. Additionally, the use of current high resolution remote sensing data was found to be important for locating water stressed olive orchards and to initiate

[5] Information on used data kindly provided by Mohamed Ouessar from the Institut des Régions Arides (IRA), Tunisia
[6] GTZ 2010d, 22

measures. It can be summarized that data quality (spatial resolution and age) and data availability have been proven to be the limiting factors for the project. More weather stations would be required for a full coverage of the study area and the structure of the socioeconomic data available (land tenure, income, demography etc.) was not sufficient for modeling.

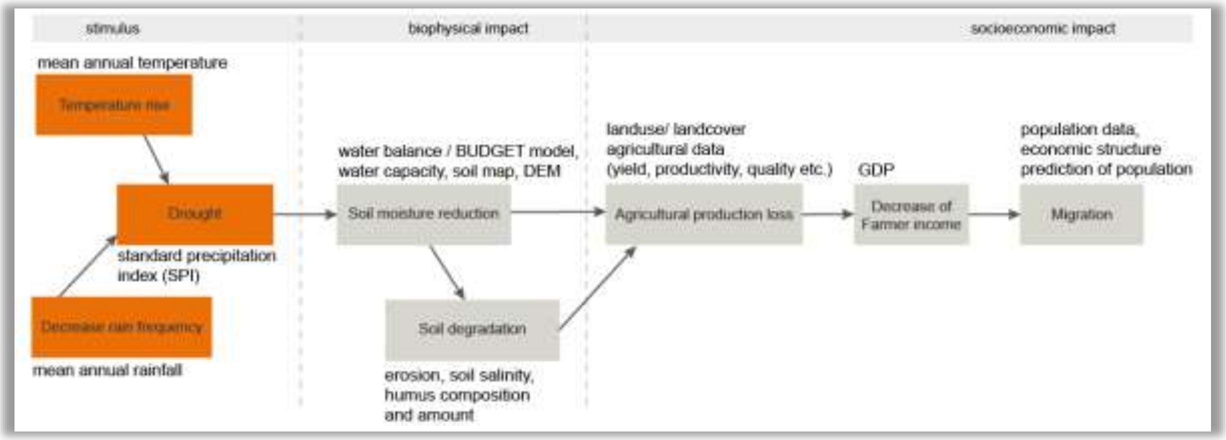


Figure 4: Data Requirements and Impact Chain of Tunisian Study

2.1.2 Indonesia

The following analyses of data requirements in Indonesia are based on two different studies. The first case study focuses on the evaluation of patterns of climate change vulnerability in the agricultural and water sector[7] while the second study elaborates on patterns of vulnerability of the coastal zone[8].

With respect to the first case study, the data requirements were in the beginning identified by means of a literature review and are largely based on the results of a comparable Indian study about climate change vulnerability in the agricultural sector. O'Brien et al. specify the following parameters and data requirements:

- Soil conditions - Depth of soil cover and severity of soil degradation; Types of degradation: erosion, chemical deterioration (salinization, loss of nutrients), physical deterioration (water logging), stability of terrain.
- Ground water availability - recharge from the canals, surface water bodies and change in land cover.
- Human capital - Represented by adult literacy rates.
- Social capital - gender equity, female child mortality rates and female literacy rates.
- Irrigation availability - net irrigated area as percentage of net sown area.
- Infrastructure quality - District-based availability of facilities for transport, energy, irrigation, banking, communication, education, and health.

Examining these parameters it seems unrealistic that all required data will exist. Using all of these parameters will just be possible at a local or subregional scale and anyhow will include a lot of surveying.

---

[7] GTZ 2010e
[8] GTZ 2010f

The data requirements and use of data for the Indonesian study are divided in two parts, climate data on the one hand and environmental and socioeconomic data on the other, because of its different data structure and acquisition.

The used climate data consists of monthly rainfall data, monthly flooding and drought occurrences data, as well as modelled rainfall data from global climate models. For defining capacity and vulnerability indices of villages, socio-economic survey data was used, while for some of biophysical data satellite interpretation with GIS techniques were required to derive the data (land use etc.). A lot of different analyses were undertaken to derive the capacity and vulnerability indices, so that we can only provide an exemplary overview about the different steps taken here. The determination of water deficits are one of the main impacts that were examined.

An overall climate impact chain and the resulting data requirements will be given in Figure 6 in combination with the results from the other Indonesian study about the coastal zone. The objectives of this study are to identify physical and socio-economic indicators of typical coastal vulnerability by collecting available and accessible data and to identify patterns of vulnerability of the coastal zone based on sea-level rise scenarios for Indonesia.

The step-by-step analysis of the study includes two steps which are important for the data requirement. 1. Data was gathered from many sources and related institutions. If the required data based on the ideal indicator cannot be found, modification of the indicators is needed based on data availability or by using data which can be used as proxies. 2. It is presumed that the ideal formulated indicators in the previous step can be adapted according to data availability, such as from either data obtained from related departments, data from questioners, or from examination of location conditions to obtain actual vulnerability indicators.

Table 2: Used data and scales for coastal zone project in Indonesia

	Spatial scale	Temporal scale	Data sources
Administrative area	Unknown	Unknown	Regency administrative boundary
Slope and elevation data	-200m resolution	-approx. 2004	Down-sampled SRTM data
Population density	Regency, scale unknown	-Unknown	Average regency value
Prosperity	Unknown	-Unknown	Human Development Index (HDI)
Land cover	Unknown, 30 arc seconds (approx.. 1km) Grid	-Unknown	MODIS data
Infrastructure	Unknown, 5km Grid)	-Unknown	total length of road, number of seaport, airport, and buildings

An overview about the climate stimuli and different impacts is given in the following Figure.

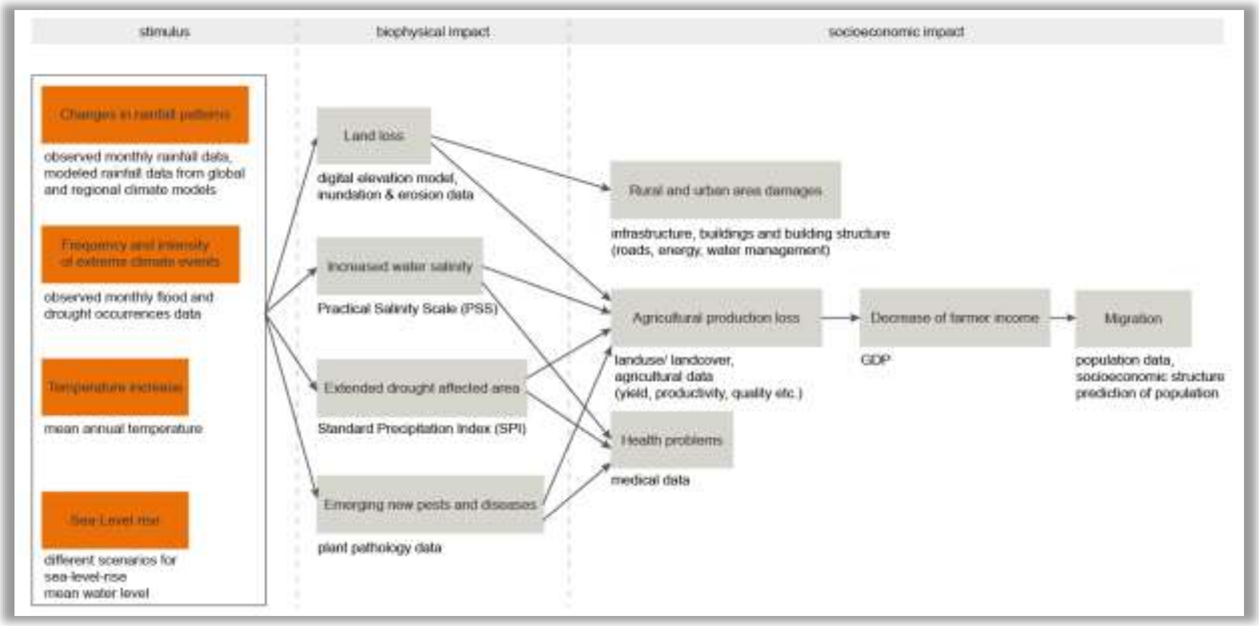


Figure 5: Data Requirements for Climate Impact Chain Indonesia Studies

2.2 Data availability and quality: conclusions

The last chapters have demonstrated the process on how to find out which kind of data will be needed for specific climate impact analyses, vulnerability assessment and possible adaptation measures. The standard approach for analyzing the data requirements is shown in the following Figure.

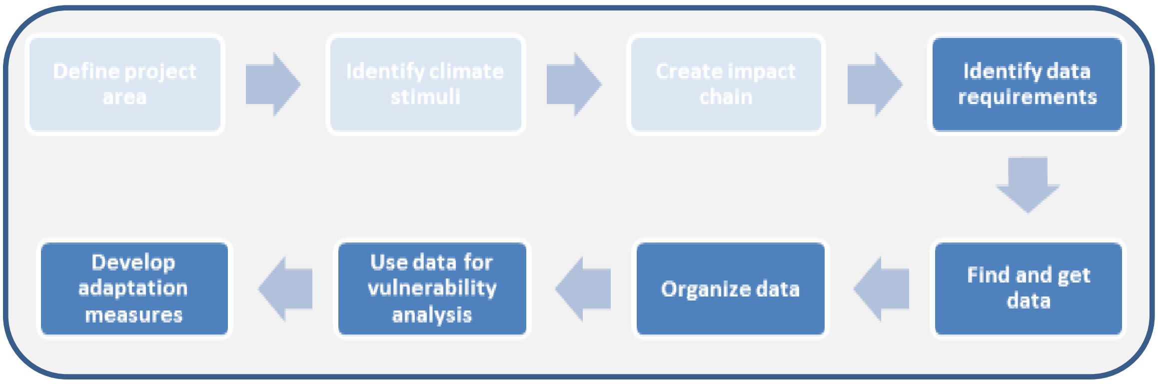


Figure 6: Standard approach for analyzing data requirements within CC adaptation projects

<u>Lessons learnt</u>

- Frequently the required data is not available and needs to be captured / digitized / collected. There is often a data gap in relation to the requirements for analyzing climate change.
- The spatial resolution and the age of the data have a considerable impact on the quality and precision of the results. They are crucial factors for assessing required adaptation measures.

- The spatial resolution of the data must be the same for all datasets. This also applies to the spatial scale, horizontally and vertically.
- Statistical or socioeconomic data are often based on abstract administrative units that do not fit to environmental structures. Therefore it might be difficult to analyze, predict or adapt to the human-environment systems.
- The process of data analyses for climate change issues is more complex than for most other spatial questions because of the manifold influencing variables and uncertainties.

The following Figure summarizes the data requirements, which have been evaluated from the impact studies. It is important to note that the table does not claim to be complete, it is rather meant to be a starting point for an analysis of data requirements.

Agricultural production loss	•land use/ land cover, digital elevation model, soil map, different scenarios for sea-level rise, mean water level, inundation data, standard precipitation index (SPI), mean annual rainfall, mean annual temperature, agricultural data
Increased forest fire frequency	•land use/ land cover, standard precipitation index (SPI), mean annual rainfall, mean annual temperature, infrastructure data
Land loss	•land use/ land cover, digital elevation model, soil map, different scenarios for sea-level rise, inundation data, standard precipitation index (SPI), mean annual rainfall, mean annual temperature, agricultural data, erosion data, humus composition
Migration / Relocation	•land use/ land cover, different scenarios for sea-level rise, mean water level, inundation data, standard precipitation index (SPI), population data, socioeconomic structure, prediction of population, agricultural data, infrastructure data, GDP, erosion data
Rural and urban area damages	•land use/ land cover, digital elevation model, different scenarios for sea-level rise, inundation data, population data, socioeconomic structure, prediction of population, agricultural data, infrastructure data, erosion data
Soil moisture reduction	•land use / land cover, digital elevation model, soil map, mean water level, standard precipitation index (SPI), mean annual rainfall, mean annual temperature, agricultural data, erosion data, humus composition
Urban water supply decrease	•land use/ land cover, digital elevation model, soil map, mean water level, standard precipitation index (SPI), mean annual rainfall, mean annual temperature, population data, socioeconomic structure, prediction of population, infrastructure data
Wetland loss	•land use/ land cover, digital elevation model, soil map, mean water level, standard precipitation index (SPI), mean annual rainfall, mean annual temperature, erosion data

Figure 7: Data requirements for different climate change impacts

As indicated previously, another crucial factor for any type of project is the importance of using appropriate scales. In the following figure lists the different planning levels in relation to the suitable map scales and spatial resolutions.

The relevance of using suitable spatial resolutions for any kind of planning level needs to be emphasized.

Planning level	Map Scale	Spatial resolution
Local	1:5-10.000	1-3 meters
Regional	1:25.000	5-10 meters
National	1.50-250.000 (depending on the area size)	25-500 meters
Global	1: 10.000.000	1-5 kilometers

Figure 8: Project scales and data requirements

It has been discussed what kind of data is needed. Although it has not yet been shown where to obtain such data, some ideas will be given when discussing how to organize and store data and information about data. This will also give an idea about the structure of the data and its related information.

3 How to organize the data?

The organization of data facilitates its collection and retrieval. Questions to be asked in this context are:

- Is there a possibility to find out in which catalog to look for the required data?
- Can I find out the age of the data?
- Who has published the data? Is there a contact person for questions?
- Which methods have been used to produce the data and which scales and resolution are they based on?
- May I save and bundle catalog data for a future use?

The questions underline the need to estimate, if the data is sufficient for questions regarding the analysis of climate impacts. Subsequently examples are given on how catalogs can be implemented and considered, which of the solutions might be suitable for the study’s needs. Advantages and disadvantages of the different possibilities will also be highlighted.

3.1 Data Catalogs – An Information Storehouse

There are three types of data catalogs. Whereas analog card catalogs are out-of-date, digital catalogs are widely-used. The digital versions contain the file-based cataloging, that could use simple table structures and the web catalog, which requires an internet connection, but therefore no software other than an internet browser.

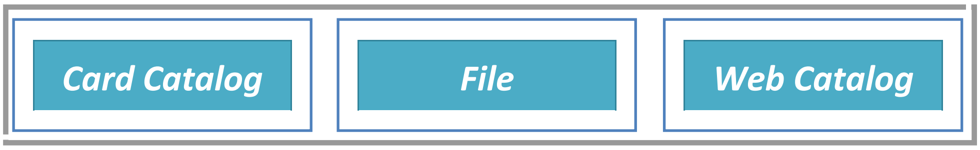


Figure 9: Data Catalog Systems

A data catalog is an application that provides access to data, services and applications taking into account subject-specific, temporal and spatial criteria. Normally a catalog is not administering the data itself but metadata, i.e. information about the data. If the catalog has a network interface it is named a web catalog or catalog service.

> *The term **data catalog** is used to describe the set of service interfaces which support organization, discovery, and access of geospatial information. Catalog services help users or application software to find information that exists anywhere in a distributed computing environment. A Catalog can be thought of as a specialized database of information about geospatial resources available to a group or community of users.*
>
> *OGC 1999*

File-based cataloging may consist of simple table or text structures but may also be part of more complex database structures. The main difference to the web catalog is the offline status. The file-based catalog cannot be reached via the internet and can be compared in this regard to the library card catalog. But in contrast to the card catalog, there are a lot of other possibilities available to transfer the catalog information.

Imagine a table, such as a Microsoft Excel list, which contains all necessary information about the used and/or required data in our project. Necessary information in this context means all metadata which is or might be important for our project. Examples of such metadata are:

- Name of the dataset, Identifier or consecutive number
- Description of content
- Origin of the data (publisher, producer, contact details, copyright, restriction of use)
- Data contents (spatial extent of the data, data structure, data source, type, format, attributes, coordinate system, creation date, used scale)
- Quality of the data (resolution, accuracy, update cycles of data)

Other information may be recorded as well. The table already indicates that, according to the data volume, information may not be viewed and queried easily.

Similar structures like the one mentioned may also be created with text documents. Only databases support better data structuring and search functions. One example for such a database is the existing list and description of EPSG coordinate systems, which is available in different database formats[9]. In contrast to using the aforementioned tables or text catalogs, this database includes different forms to facilitate the entry, navigation and search for data. Thus, it is easy to navigate through the data and metadata of single coordinate systems or set up own records in the form. The structure of the coordinate system should however, be considered, which is always based on the same schema and for this reason the input mask is relatively simple. If this is compared to the data requirements for climate change issues, these are much more complex with different structures ranging from vector and raster data to statistical data. This demands more complex input masks which should somehow be standardized to support the querying of all entries.

Figure 10: EPSG database view, (MS Access version), http://epsg.org

Web catalogs show considerable differences to the before mentioned solutions; they do not require any additional table or database software but an internet browser and an internet connection. Web catalogs may be viewed as a key technology for localizing, administering and maintaining distributed spatial resources (geo-data, applications and services). The term catalog service is also very often used to underline the possibilities of integrating catalog data into different systems, applications or software solutions (GIS etc.) by using standardized interfaces. Numerous examples of web catalogs or catalog services respectively exist that can be evaluated and used for data research. An example of such a web catalog will be given below, which is provided by the Federal Geographic Data Committee (FGDC) and focusses on US geo-

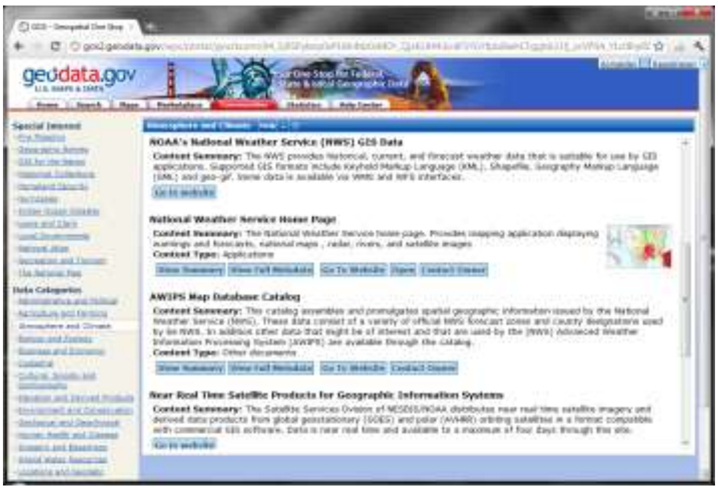


Figure 11: Example for a web catalog (http://geodata.gov)

9 http://www.epsg.org/

data.

What are the main features of web catalogs especially in relation to the mentioned solutions? Web catalogs usually are clearly arranged and represented. They contain assorted categories which give access to other datasets or metadata. They often support direct downloads of data and metadata and always have an integrated viewer to visualize the data. Basically the flexibility and adaptability of the web layout much better reflects the needs of structuring and visualizing information than any of the other solutions.

As simple as it is to enter own data or information into a web catalog, integration of other web catalogs may also be easily done. Thus, if an entry already exists, it can be copied and integrated into other catalogs without entering it again.

The import of existing catalog data is supported via standardized interfaces and is called metadata harvesting. Going back to the main feature of web catalogs which are the possibilities to access services via the internet, to use or edit metadata entries, hybrid catalogues also exist, which can be used online and offline as long as there is a regular reconciliation of the data. An example of such a catalog is given in the following Figure.

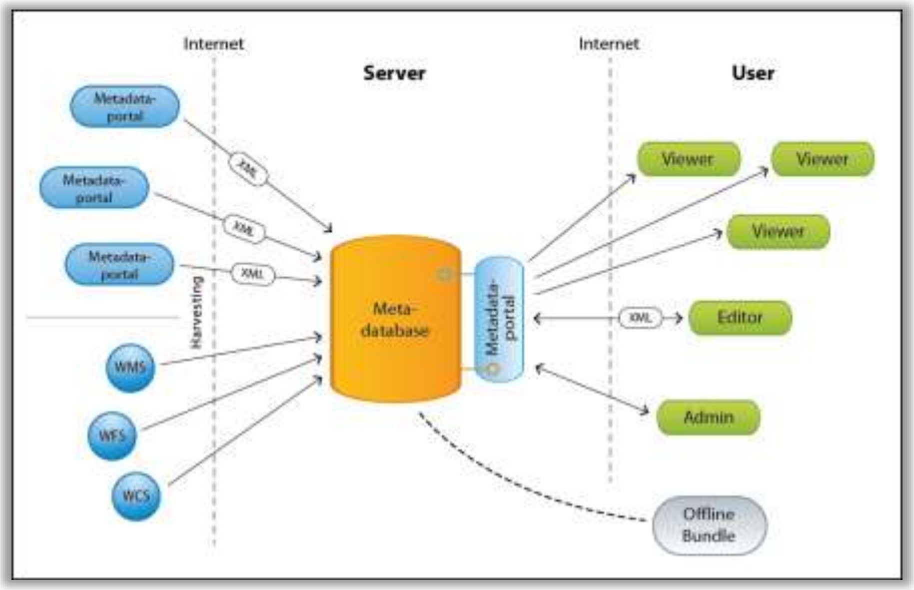


Figure 12: Web Catalog with online/offline components

The special feature and attractiveness of these kinds of catalogs is that there is no internet connection required to use the catalog and to query or update metadata. Therefore it is a local installation of the catalog. Metadata can be synchronized whenever there is an internet connection or if there is no internet connection available metadata files or updates can be sent by post to update the particular catalog. Such an online/offline scenario seems to be very useful for regions, where there is no or only unreliable internet connection. This is very common in developing countries and especially in rural areas.

3.2 Catalog software and its application

At this point a short review about different catalog software applications and catalog services is given. More information about different catalog software can be found in the link list at the end of this study. Three software solutions have been selected, which are widely used for different reasons.

deegree catalogue manager

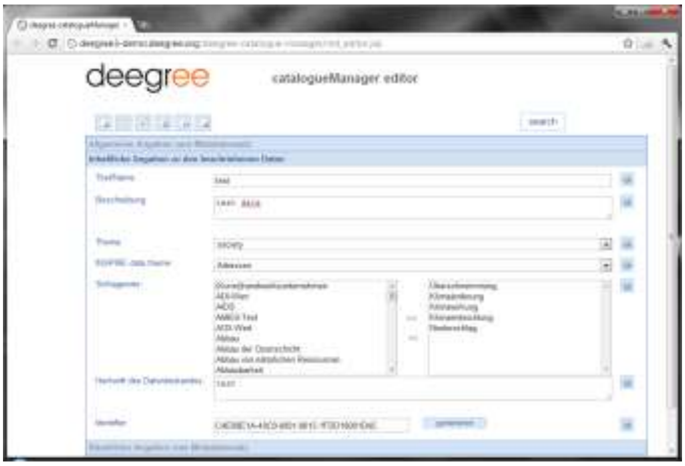


Figure 13: deegree Catalogue Manager (http://deegree.org)

deegree is an open source geospatial software package with implementations of different standardized web services and includes a geoportal, a desktop application, security mechanisms, and various tools for geospatial data processing and management. The integrated catalogue manager allows for all kinds of importing and managing of metadata. It has been selected as the standard implementation for the European spatial data infrastructure. As deegree is composed of open source tools it is widely used worldwide, with a focus on Europe and Asia.

ESRI ArcCatalog

ESRI is the world's largest commercial GIS software producer. ESRI software is used by more than 350,000 organizations worldwide including most U.S. federal agencies and national mapping agencies. ArcGIS is a system that allows for easy editing data, maps, globes, and models on the desktop and to use the results on a desktop, in a browser, or in the field via mobile devices, depending on the specific needs of the organization. Metadata capabilities are included in ArcCatalog.

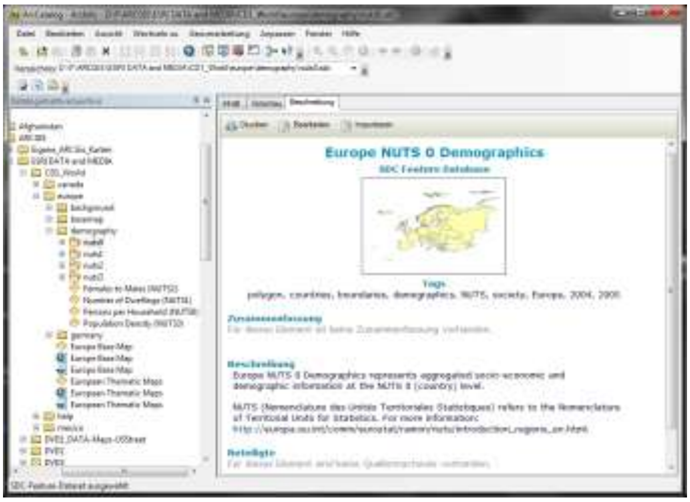


Figure 14: ESRI ArcCatalog (http://esri.com)

Geonetwork

GeoNetwork is a catalog application to manage spatially referenced resources. It provides powerful metadata editing and search functions as well as an embedded interactive web map viewer. It is currently used by numerous Spatial Data Infrastructure initiatives across the world. GeoNetwork has been developed to connect spatial information of communities and their data using a modern architecture, which is at the same time powerful and low cost, based on the principles of Free and Open Source Software (FOSS) and International and Open Standards for services and protocols.

Figure 15: GeoNetwork Metadata Administration (http://geonetwork-opensource.org)

3.3 Implementation requirements

It has been demonstrated how data catalogs work and what kind of features they have or support. We have also seen some examples of catalog software.

Now the requirements for such catalogs will be examined as were done for the data as well. The requirements of what a catalog should be capable of are listed below:

- The catalog must support different data types (vector, raster, tables) and different data entry forms for the different data structures.
- It should be capable to store and import metadata from different sectors or fields (climate, socioeconomic, topographic, etc.).

- Different kinds of catalog access' should be supported (online, offline, maybe also file-based)
- The main metadata information (name, identifier, description, data source information, data contents, quality of the data) should be supported and the catalog should be flexible for new or adapted data structures
- The technical level and requirement should be scalable according to the projects' complexity and financial volume. The maintenance effort for the catalog should be as low as possible.

In the following the advantages and disadvantages of each catalog type will be evaluated in order to obtain an idea about the application fields and usability with respect to climate impact analyses.

Table 3: Requirements, advantages and disadvantages of catalog types

	File-based	Web-based	
	User	Client	Server
Description	Text (txt, csv, doc, odt etc.) Table (xls, xlsx, ods, mdb etc.)	XML, browser-based	XML, server technology, web technology
Hardware	No special requirement	No special requirement	High-performance hardware recommended
Software	Software for using file format	No special requirement, internet browser	Webserver, mapserver, catalog software, geoportal
Internet	Not required	Required *	Required
Accuracy/ quality	varying (mostly not standardized, user-dependent	Depends on server configuration	If defined (mandatory fields, selection lists)
Costs	Low – medium	Low	High
Actuality	Low - medium - high	High	High
Service/ support	Important	Not necessary, easy	Very important
Versions/ maintenance	Depends on users, policies, mostly no versions available	Not necessary	Can be implemented easily
Data availability/ content	Low – medium	Depends on server	Very good
Data access / performance	Medium, depends on data volume	Medium – fast / Very good	Fast access / high performance
Multiuser Capabilities	Restricted, 1 editor	yes	yes
Advantages/ Disadvantages	Easy if known software, structure and entries can be changed easily / no import possibilities, map viewer, mostly no standards used, software required	Easy, complexity can be adapted or changed, different languages, profiles, requirements / internet access required	Harvesting of data possible, integration with other components (map viewer, standards import etc.) / complex server technology

* Offline solution possible

3.4 All about standards and metadata

The increasing availability and use of spatial data for diverse questions leads to the need for standards and standardized access from a variety of applications. Different data formats have to be imported; frequently different software applications have to be used to solve complex spatial problems.

Spatial data represent an extremely rich domain that requires special attention, perhaps more so than any other type of data. The enormous variety of ways of encoding geospatial data and the large number of classification schemes, vocabularies, terms, and data definitions used by producers and consumers of data make it particularly challenging to process requests or queries for geoinformation. Within any organization trying to integrate geodata, it is important that there is an agreement on the proper definition and use of metadata. In principle, proper metadata can provide the foundation for interoperability, by defining the meaning of each of the terms that underlie the data management process.

Geospatial metadata commonly document geographic digital data such as GIS files or databases. Earth imagery may also be used to document spatial resources including data catalogs or web mapping applications. Metadata records include core library catalog elements such as Title, Abstract, and Publication Date. An overview about the core components is given below.

<u>Core components of metadata records</u>

- **Metadata Record Information** - information about the metadata record including the language in which the record is written, a unique file identifier, the metadata standard, a point of contact for the metadata, and the record date of the metadata.
- **Identification Information** – citation-level information about the data including the title, abstract, purpose for creation, status, keywords, (theme and place), and extent (temporal, vertical and horizontal).
- **Constraints Information** – information about legal and security limitations to data access and use.
- **Data Quality Information** – information about the processes and sources used to develop the data and/or accuracy assessments performed.
- **Maintenance Information** – information about scope and frequency of data updates.
- **Spatial Representation** – information about the mechanism used to represent spatial data (grid, point, vector).
- **Reference System Information** – information about the reference systems used to represent geographic position and time.
- **Content Information** – information about the data set entities and attributes.

A metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource. It represents the who, what, when, *where, why and how of the resource.*
(http://www.fgdc.gov/metadata)

- **Symbology Information** – information about symbols used to represent spatial features.
- **Distribution Information** – information about the data distributors and methods for obtaining the data.
- **Metadata Extension Information** – information about customized, user-based, changes to the elements, domains or conditionality of the standard.
- **Application Schema Information** – information about the schema or data models used to structure the data

Metadata are structured facts that describe information, or information services. Metadata facilitates many things beyond enabling information discovery and access; it also informs about the appropriate use of products and services. Metadata is applicable to any topic/anything, not just to geographic information. For example, the business card is metadata about oneself.

There are three types of geospatial-related standards:

- Content standards including land use and surveyor codes, data dictionaries, geographical place names, and bathymetry
- Access standards including ISO 19100 series (Geographic information), ISO 23950 (Information Retrieval – Z39.50), most OpenGIS Specifications
- Exchange standards including Geography Markup Language (GML), Scalable Vector Graphics (SVG), Uniform Resource Identifiers (URIs, also known as URLs)

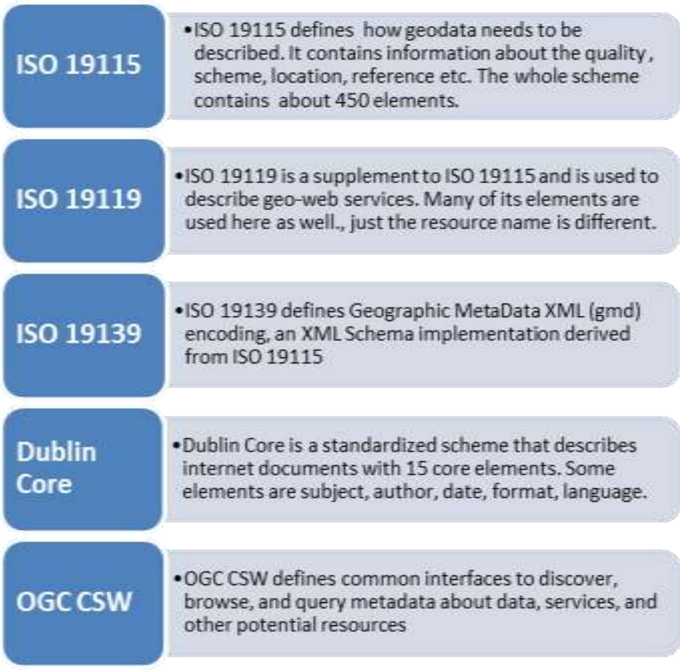


Figure 16: Important Metadata Standards

The most important organization for standards development applicable to business, government and society is the International Organisation for Standardisation (ISO). ISO has published an International Standard for Metadata ISO 19115 which is widely used. An overview about the main metadata standards is given in the side box. The format for the metadata itself will be discussed in the following chapter.

3.4.1 The standards format: XML

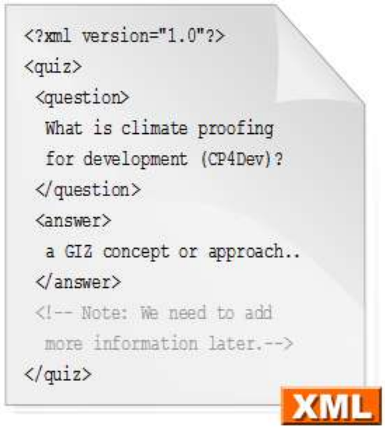
The Extensible Markup Language (XML) is a markup language, which was developed by the World Wide Web Consortium (W3C) to store and transfer data in a hierarchical structure.

XML is text-based and therefore independent from operating system and application software. It can be used everywhere and for every kind of application.

Additionally it is the base of other technologies, standards and conventions. Its interoperability and versatile usability make it also very attractive for metadata applications. Almost every geo-data portal or catalog implementation uses and supports XML as the main transfer (import and export) format and to store the metadata information.

XML is nonetheless not only used as a format for metadata but also for the data itself. Different standard geo-data formats use the XML structure to store the data. Two examples are GML (Geography Markup Language), which is used in GIS widely to transfer data and KML (Keyhole Markup Language). The latter is used as a standard format in many web mapping applications and Google Earth.

XML is an open standard for describing data and is used for defining data elements on a Web page and documents.



The structure of the XML documents is very easy and can be used by everybody who knows the element names, which are defined in the different standards. The XML document structure will only be examined briefly, as there are countless Web pages about XML available.

XML documents form a hierarchical tree structure that contains different elements which are closed by tags. A tag originally is an HTML structure with a name that is closed by angle brackets as shown below. A start tag, here “<title>” opens the element and the end tag “</title> closes the tag. The space in between is filled with the content.

```
<title>My XML Title Name</title>
```

XML documents are composed of such tags. The structure may also contain encapsulated tags, as shown in the side example.

3.4.2 The standards organizations: ISO and OGC

We have seen the standard format XML which was defined by the W3C. But so far we have not yet shown how the XML format is defined and filled with data. This is also important to ensure compatibility and transfer. There are two international organizations which lead in developing standards for geoinformation: the OGC (Open Geospatial Consortium) and one of ISO technical committees – ISO/TC 211. The roles of these organizations in the process of standardization are rather different. OGC deals with technology issues and practical implementation, and ISO with the more formal and procedural approval of standards as the official documents. In practice, OGC develops the technical specifications and the ISO/TC 211 approves them as international standards.

ISO was founded in 1947 and brings together national organizations. Each country is represented, in principle, by only one. Currently, ISO members represent 160 countries.

The International Organization for Standardization widely known as ISO, is an international standard-setting body composed of representatives from various national standards organizations. Founded on February 23, 1947, the organization promulgates worldwide proprietary industrial and commercial standards.
http://www.iso.org

There are many technical committees and working groups developing the standards within ISO. One of these technical committees is ISO/TC 211, which aims to develop and implement standards for geographic information. ISO/TC 211 so far has developed more than 50 standards for spatial information. These standards specify methods, tools and services in relation to spatial data (access, harvesting, processing, presentation and transmission) in electronic form between users of different systems and applications.

The OGC, Open Geospatial Consortium, is an international consortium of more than 400 parties from around the world. Commercial organizations, non-profit organizations, government agencies and universities are involved in the OGC, that was established in 1994 to develop agreed and accepted specifications for interoperability issues in the field of geoinformation. The OGC Vision is a world in which everyone benefits from geographic information and services made available across any network, application or platform. Its mission is to deliver spatial interface and encoding specifications that are openly and publicly available for everybody.

The Open Geospatial Consortium (OGC®) is a non-profit, international, voluntary consensus standards organization. We are leading the development of standards for geographic content and services, sensor webs, and location services.
http://www.opengeospatial.org

Because of the growing interest in geoinformation, web mapping and its increasing relations with the Internet and mobile technologies, close cooperation of GI centers with organizations involved in solutions for the Internet (W3C) and e-business (e.g. OASIS) becomes increasingly important. The ebRIM standard for metadata catalogs developed by OGC in collaboration with the OASIS is an example of such co-operation. Although it is beyond the scope of this study to introduce all existing standardization organizations, some important examples of such organizations, which are involved in topics related to spatial data and metadata, should be mentioned:

- ISO/TC 211 – ISO 19100 series (Geographic information)
- OGC – geo web standards, GML, KML, sensor standards, netCDF, ebRIM Application Profile
- DCMI (Dublin Core Metadata Initiative) - interoperable metadata standards for the Internet, RDF encoding
- W3C – XML, web architecture, web services, HTML, CSS, Web Schema etc.

OASIS (Organisation for the Advancement of Structured Information Standards) – OpenDocument, SAML, ebRIM

3.5 The use of metadata in climate research and other subject areas

This chapter focuses on other subject areas than spatial data use different kinds of metadata. The main differences and characteristics will be highlighted and analyzed. The main focus will be on climate data requirements, on environmental data and briefly on regionalized statistical data. In general, ISO standards and norms are widely used in climate research as well as in other subject areas. The ISO standard ISO 19156, which introduces the concepts of observation and spatial sampling features, is used for complex observation of environmental parameters. It has been implemented into the Climate Science Modeling Language, a standards-based data model and GML application schema for atmospheric and oceanographic data, which was developed in the UK and may be also used world-wide in the future. A data format which is used for time-based data and is very common in the US is NetCDF, an OGC standard.

According to the use of metadata in climate research there is one intrinsic problem. Many countries have their own Climate Data Management System and therefore individual metadata structures. Although a new WMO Information System is planned to improve the interoperability of meteorological information, this process will take some years. Many of the involved standards have to be improved and there is a long list of climate-related requirements, which will change future climate research and possibilities. As there is still a lot of work in progress, it cannot be foreseen which of the standards will be implemented and used for global change research issues. The often discussed standards are:

- SWE - Sensor Web Enablement and SOS - Sensor Observation Service
- O&M – Observations & Measurements (ISO 19156)
- SensorML – Sensor Model Language
- SES –Sensor Event Service (SAS v.2)
- SPS - Sensor Planning Service
- TML - Transducer Model Language
- WNS – Web Notification Service

Another important topic in climate research is the semantic heterogeneity of different relationships, synonyms etc. which is one of the main reasons for a missing interoperability of different climate data and services. The WMO approach to introduce a new Information System will also contain better semantic capabilities.

3.6 Metadata use cases

Two examples are given below that illustrate the importance of metadata and data catalogs. In this respect, numerous other examples exist which may however, not all be elaborated on in this study.

3.6.1 INSPIRE in the European Union

The INSPIRE directive came into force on 15 May 2007 and will be implemented in various stages, with full implementation intended by 2019. The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organizations and to better facilitate public access to spatial information across Europe. A European Spatial Data Infrastructure will assist in policy-making across boundaries. Therefore the spatial information considered in the directive is extensive and includes a great variety of topics and technical themes.

INSPIRE is based on a number of common principles:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed information for thorough investigations, generalized information for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- It should be easy to find out what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used[10].

What makes the INSPIRE project special is the huge effort that is needed to integrate, harmonize and standardize a bulk of geo-data information. So far, it is a successful example for implementing a spatial data infrastructure at a multinational level.

A similar concept could not be transferred to the countries which were mentioned in the study because of different initial situations (policy, organization etc.). But INSPIRE may serve as a best practice example for implementing a spatial data infrastructure under somehow adverse conditions. Furthermore, it can be used as a research field, with manifold publications, documents, open source software components, test beds, metadata test software etc. It is recommended to use the INSPIRE resources in this way.

*The term **spatial data infrastructure** (SDI) describes a framework of technologies, policies, and institutional structures that together facilitate the creation, exchange, and use of geospatial data and related information across an information-sharing community within an organization or more broadly the sharing of geo-spatial information at a local, regional, national or global level.*

[10] http://inspire.jrc.ec.europa.eu/index.cfm

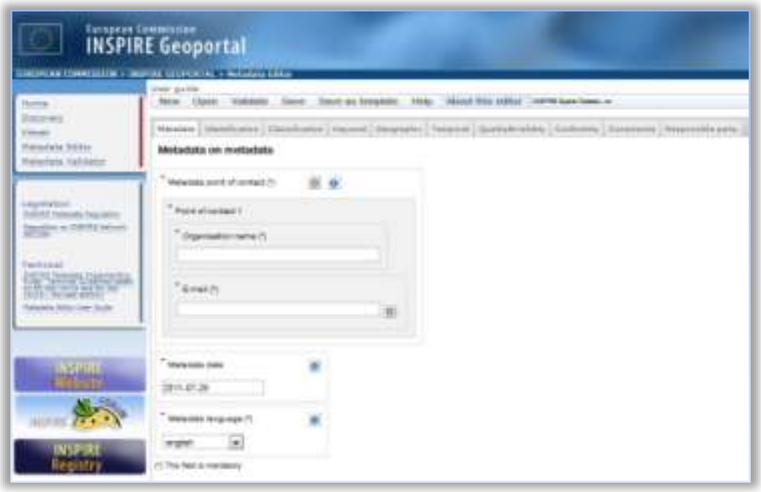


Figure 17: The INSPIRE Metadata Editor (http://www.inspire-geoportal.eu/index.cfm/pageid/342)

3.6.2 FGDC in the US

The U.S. Federal Geographic Data Committee was founded in 1990 to coordinate the data acquisition, administration and use of spatial information by federal authorities. This was done to build up an efficient Geoportal.

The FGDC as an example demonstrates the significance of standards. For a long time FGDC adopted data and catalog concepts that were developed by ESRI, a world-wide operating US-based geoinformation company. Although these concepts and models were adopted by many different countries, they were not accepted world-wide. However, analyzing global structures and developments like climate change requires global and somehow standardized geo-data. Though ESRI and FGDC developed interfaces to allow data connections and metadata imports, increasing costs and existing world-wide standards have changed FGDC's data and standards policy. In 2009 ISO 19115 has been accepted as the standard for geo-data in the US and Canada and the former FGDC standard will be replaced gradually. The North American Profile (NAP) of ISO 19115 will be accomplished by 2013.

What conclusions can be drawn from this? In a global world with global problems and cooperation in fields like climate change and risk analyses there is little scope left for national oriented spatial data infrastructures. Another reason may be FGDC's policy to turn away from company-driven development. Standards must be open and should not be restricted by a copyright law. As mentioned for INSPIRE as well, the FGDC Web page is a good starting point for SDI research and information about geo-standards and spatial metadata.

Figure 18: The FGDC Geoportal (http://geodata.gov)

4 Conclusion

The complexity of topics related to climate research has been shown. Firstly, the type of data required and data organization was highlighted as a starting point of the analysis of different climate impact studies. It was pointed out that the analysis of data requirements is of great relevance before the project starts. Otherwise the use of insufficient or unsuitable data may lead to wrong conclusions. In this respect the importance of considering the dependencies between planning levels, map scales and the corresponding spatial resolution needs to be accounted for. The introduction of different catalog types and systems facilitates the evaluation of different possibilities and the approach to organize and structure data by using metadata and standardized formats. Metadata and catalogs may be used for archiving and structuring data inventories; they allow specific queries for data, metadata and services, and last but not least may be used and integrated in regional, national or international spatial data infrastructures.

The different steps on how to use metadata and catalogs are highlighted below:

1. Publication – make data and their associated metadata available by using web services and/or combined online/offline tools
2. Catalog – aggregate metadata from published web services
3. Discovery – search a catalog to identify data series or datasets of interest
4. Access – acquire the selected data from a web service
5. Processing – apply complex analysis tasks or other processing routines to datasets, such as evaluating climate change impacts or vulnerabilities to estimate proper adaptation measures

This study has identified different bottlenecks in relation to the available data and its spatial resolutions. Furthermore the usability of catalog services for storing climate impact information has been proven. Current examples have demonstrated the applicability of data and metadata catalogs and the following links provide a deeper insight into the complex and comprehensive subject of this study.

5 Bibliography

BMU (Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit) (Ed.) 2011: International Climate Initiative. Inventory of methods for climate adaptation. Available online at: http://www.bmu-klimaschutzinitiative.de/en/projects?p=1&d=703 (Retrieved 2011-08-01).

European Commission (Ed.) 2008: Verordnung (EG) Nr. 1205/2008 der Kommission vom 3. Dezember 2008 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich Metadaten. Amtsblatt der Europäischen Union, L 326/12. Available online at: http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:DE:PDF (Retrieved 2010-03-02).

European Commission (Ed.) 2009: INSPIRE Roadmap. Available online at: http://inspire.jrc.ec.europa.eu/index.cfm/pageid/44 (Retrieved 2010-02-26).

European Parliament and Council (Ed.) 2007: Directive 2007/2/EC of the European Parliament and Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, L 108/1. Available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF (Retrieved 2007-11-21).

European Commission Joint Research Centre (Ed.) 2009: INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. Available online at: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf (Retrieved 2009-09-07).

GTZ (Deutsche Gesellschaft für technische Zusammenarbeit GmbH) (Ed.) 2010a: Climate Proofing for Development. Adapting to Climate Change, Reducing Risk. Eschborn.

GTZ (Deutsche Gesellschaft für technische Zusammenarbeit GmbH) (Ed.) 2010b: DRAFT TRAINING MANUAL ON CLIMATE PROOFING FOR DEVELOPMENT. Philippines.

GTZ (Gesellschaft für technische Zusammenarbeit) (Ed.) 2010c: Manual. Climate Proofing Tool Vietnam. Eschborn.

GTZ (Gesellschaft für technische Zusammenarbeit) (Ed.) 2010d: Case of olive production sector in the governorate of Médenine. Final report. Eschborn

GTZ (Gesellschaft für technische Zusammenarbeit) (Ed.) 2010e: Patterns of vulnerability in the agriculture and water sector in Indonesia. Final Report. Eschborn.

GTZ (Gesellschaft für technische Zusammenarbeit) (Ed.) 2010f: Study on Patterns of Vulnerability on Coastal Zone of Indonesia. Final Report. Eschborn.

OECD (Ed.) 2009: Integrating Climate Change Adaptation into Development Co-operation. POLICY GUIDANCE. Available online at: http://www.oecd.org/dataoecd/45/45/44887764.pdf (Retrieved 2011-04-06).

OGC (Ed.) 2006: Interoperability & Open Architectures: An Analysis of Existing Standardisation Processes & Procedures. OGC 05-049r1.

Trinh Cong Van (2009): Identification of Sea Level Rise impacts on the Mekong Delta and Orientation of adaptation activities. Available online at: http://www.vncold.vn/Modules/CMS/Upload/13/Science/TCVanEwv09_160509/identification%20sea%20level%20rise%20impacts%20on%20MK%20delta_5april.pdf (Retrieved 2011-08-08).

Worldbank (Ed.) (2007): The Impact of Sea Level Rise on Developing Countries: A Comparative Analysis. Washington.

6 List of acronyms

AP	Application Profile
ASDI	Australian Spatial Data Infrastructure
CF	Climate and Forecast Metadata Conventions
CP4Dev	Climate Proofing for Development
CSW	Catalogue Service Web
DCMI	Dublin Core Metadata Initiative
DEM	Digital Elevation Model
DTD	Document Type Definition
ebRIM	Electronic Business Registry Information Model
ECVs	Essential Climate Variables
EPSG	European Petroleum Survey Group
FGDC	Federal Geographic Data Committee
GDP	Gross Domestic Product
GEOSS	Global Earth Observing System of Systems
GIS	Geographic Information System
GIZ	Gesellschaft für internationale Zusammenarbeit
GMES	Global Monitoring for Environment and Security
INSPIRE	Infrastructure for Spatial Information in Europe
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization
LULC	Land Use / Land Cover
netCDF	Network Common Data Form
OAI	Open Archives Initiative
OAI-PMH	Open Archive Initiative – Protocol for Metadata Harvesting
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geospatial Consortium
PIK	Potsdam Institute for Climate Impact Research
RDF	Resource Description Framework
SDI	Spatial Data Infrastructure
SLR	Sea-level-rise
SRTM	Shuttle Radar Topography Mission
UML	Unified Modeling Language
W3C	World Wide Web Consortium
WCS	Web Coverage Service
WFS	Web Feature Service
WIS	WMO Information System
WMO	World Meteorological Organization
WMS	Web Map Service
WSDL	Web Services Description Language
XML	Extensible Markup Language
XSD	XML Schema Definition
XSL	Extensible Stylesheet Language
XSLT	Extensible Stylesheet Language Transformations
Z39.50	Client–server protocol for searching and retrieving database information
ZING	Z39.50 International Next Generation

7 Links

Standardization organizations

DCMI	http://dublincore.org/
ISO	http://www.iso.org
OASIS	http://www.oasis-open.org/
OGC	http://opengeospatial.org

W3C	http://www.w3.org/

Catalog software

CATMDEdit:	http://catmdedit.sourceforge.net
deegree Catalogue Manager:	http://deegree.org
ESRI ArcCatalog:	http://www.esri.com/
ESRI Geoportal Extension:	http://geoss.esri.com/geoportal
GeoNetwork:	http://sourceforge.net/projects/geonetwork
IME:	http://www.crepad.rcanaria.es/metadata/
INSPIRE Metadata Editor:	http://www.inspire-geoportal.eu/
M3CAT:	http://www.intelec.ca/html/en/technologies/m3cat.html
MetaD:	http://www.geoportal-idec.cat
MetaGenie:	http://www.gigateway.org.uk/metadata/
Preludio:	http://www.disy.net/preludio
TerraCatalog:	http://www.sdi-suite.de/index_en.shtm

Metadata and Geoportals

FGDC	http://www.fgdc.gov/	
INSPIRE	http://inspire.jrc.ec.europa.eu/	
geoSYS	http://gdi.geonetwork.web-mapping.com/	http://giz.csw.web-mapping.com