HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Risk Habitat Megacity
A Helmholtz Research Initiative 2007-2013

Sharing experience on project data management with *GeoNetwork opensource*

This document shares the experience on data management with GeoNetwork opensource in the Risk Habitat Megacity Research Initiative. It is directed at people who manage large volumes of data and facilitate access to geospatial data for Decision Makers, GIS Experts and Spatial Analysts enhancing the spatial information exchange and sharing between organizations and their audience.

Reflection on the experience

The almost two years of experience with data management in *Risk Habitat Megacity* provide some lessons that may be of use to other initiatives intending to set up their data system.

As far as the technology is concerned, the implementation of a standards-based Geographic Data and Information system based on *GeoNetwork opensource* is generally straightforward. A platform independent installer permits quick installation and configuration of a customized spatial data management system on a personal computer as well as a more complex server installation. This includes a relational database as the basis for a metadata catalogue and system functionality and a map server to provide geospatial information. The application can also be integrated into an existing IT-infrastructure connecting to different databases and web mapping applications. The customization of the user interface, e.g. adapting your institutions corporate design, or adapting the structure of the open source software requires experience and expert knowledge in XML and Java programming. Some tutorials at the *GeoNetwork opensource* homepage, user and developer mailing lists provide help and support.

The non-technical side of 'designing the system' and getting all involved participants to use the system takes much more effort and must not be underestimated. As the benefits are very often not fully appreciated and exchange of data via other more simple means is faster, users are often reluctant systematically use the data management system. Therefore a strict and binding data policy at the onset of a project is useful. Major aspects that need to be addressed at an early stage are willingness to share data, forms of communication, benefits and services, author rights and data publication.

Given the potential for long-term availability of project information, an agreement by all involved collaborators on options for further use of data after project termination is advisable at the beginning.

Potential to expand the functionality of *MetaCity*

Expanding the accessibility of spatial information

Geospatial information is a powerful means for analysing and visualizing problems as well as suggesting and communicating possible measures. The implementation of a map server guarantees the use of available large volumes of geospatial data and newly information generated by the research initiative for GIS and modelling applications or the creation of customized maps.

These data can be managed by a database management system (DBMS) or stored file based and are provided via Geo Web Services. Collaborating organizations can implement own map server applications like ESRI ArcIMS, GeoServer or UMN MapServer integrated in their specific IT infrastructure to deliver geospatial information for decentralized use.

The query, visualisation and analysis of geodata can be managed by various WebMapping applications that operate as a client. Those clients are external Desktop GIS or WebGIS applications or the integrated WebMapClient *Intermap*.

In this way results of the various research studies based on geospatial data are processed and can be published in the form of digital and interactive maps using *GeoNetwork opensource* web mapping tools instead of hardcopies or simple figures. This offers an opportunity for future exchange of (geo-) information with other organizations and initiatives. Original data with restricted access remain in possession of the respective distributor to ensure actuality of data and its protection from unauthorized use.

Enhancing spatial information exchange

For the research area – the Metropolitan Region of Santiago de Chile – the available data comprise raster data like satellite imagery and land-use classification data, as well as vector data like administrative boundaries at regional and local levels, street networks, hydrography and population statistics.

Enhancing the exchange of data and information

Furthermore, the software implements a number of catalogue interfaces to discover, browse and query metadata and services from distributed metadata repositories.

The harvesting functionality is a mechanism of data collection in accordance with rights to data access and data ownership protection. Through this functionality it is possible to collect public information from different GeoNetwork nodes or other metadata repositories and to copy and store these information locally in the database.

This feature enables the exchange of information with other research centres and building up a network including partnerships with governments, non-governmental organizations and local, national and regional organizations.

To activate this potential requires that collaborating organizations are running software applications which implement those catalogue interfaces and facilitates the use of standards.

Table of contents

Data management in *Risk Habitat Megacity*

The *Risk Habitat Megacity* research initiative is an interdisciplinary project that investigates sustainable futures of 'urban' Latin America. It reflects the joint work of a network of natural and social scientists and engineers from more than ten universities and research centres from Latin American countries and the Helmholtz Association in Germany in a close working relationship with the Economic Commission for Latin America and the Caribbean of the United Nations and local stakeholders. The initiative focuses on the Metropolitan Region of Santiago de Chile and includes other metropolitan areas in Latin America as reference cases.

Requirements for the data management

The availability of project data for joint use and the exchange of information play a major role in the close collaboration of research partners and for sharing information with others. [see Fig. 1]
The interdisciplinary and international nature of the initiative presents various challenges that data management has to face: accessibility from literally anywhere in the world, use of a range of languages, the necessity to incorporate a wide variety of data and data sources and the need for information exchange with other partners from research and practice.

The principle goals of data management in the initiative are:

>> To guarantee best possible (universal) access to project data for all collaborators including rights management;

>> To distribute documented data and provide software to retrieve available data sets;

>> To enable exchange of and communication about existing data and the current state of work between all participants;

>> To permit access for other interested partners to certain information and data.

In particular the efficient use of geospatial data as a common information resource implies that documented geospatial data are centrally available and can be certainly combined. Thus, decentralized access to geospatial data and information must be possible for all collaborators of the initiative.

To meets these requirements the development of a data infrastructure as a collection of technologies, policies and institutional arrangements that facilitate the availability of and access to data. [see Fig. 2]

Fig. 1:   Enhancing spatial information exchange and sharing

Fig. 9:   Creating interactive maps

Features
- Add, arrange, remove layers
- View metadata
- View and select styles
- Show legend
- Add individual markers
- Export map as PDF
- Send, up-/download a context

Use of geospatial Information

GeoNetwork opensource connects different parts of a modern spatial data infrastructure by integrating closely related specialized systems, e.g. the Metadata Catalogue, a Web Map Server and a Web Map Client in accordance with international standards. The possibility to search the catalogue for available web services and to add geospatial data directly to the embedded Web Map Viewer illustrates the integrative potential of *GeoNetwork opensource*. [see Fig. 8]

Fig. 8:   View and analyse spatial datasets

(1) Search the catalogue for data

(2) Read metadata

(3) Add directly to the map viewer...

(4) Or open in a external map client

Creating interactive maps

InterMap allows users to overlay map layers from distributed servers around the world to create a customized thematic composite maps on their computer. The client supports OGC-WMS compliant web services as well as ESRI ArcIMS map services. Users can add project data or geospatial data from predefined available Web Map Servers from outside the project e.g.:

>> the program visualization server of the GLOBE (Global Learning and Observations to Benefit the Environment) program which contains environmental data from NASA or

>> the OnEarth Web Map Server located at NASA's Jet Propulsion Laboratory, which contains worldwide satellite imagery

Also possible is the search of the project database for available Geo Web Services (WebMapServices/ WebFeatureServices) and adding these data directly to the viewer.

Users can likewise change layer styles or add markers with personal explanatory notes and print or export these maps as a PDF. [see Fig. 9]
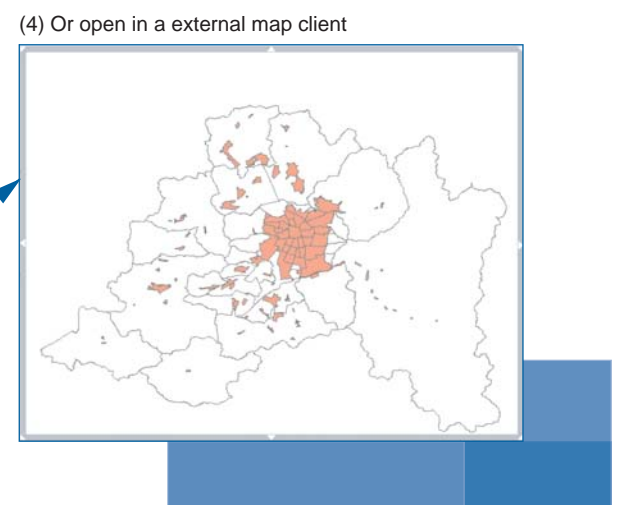
Implementation of a Data Infrastructure

A spatial data infrastructure (SDI) donates a framework to manage spatial data and metadata, users and tools that are interactively connected in order to use spatial data and information in an efficient way. For its implementation the focus is on data, services, standards and network.
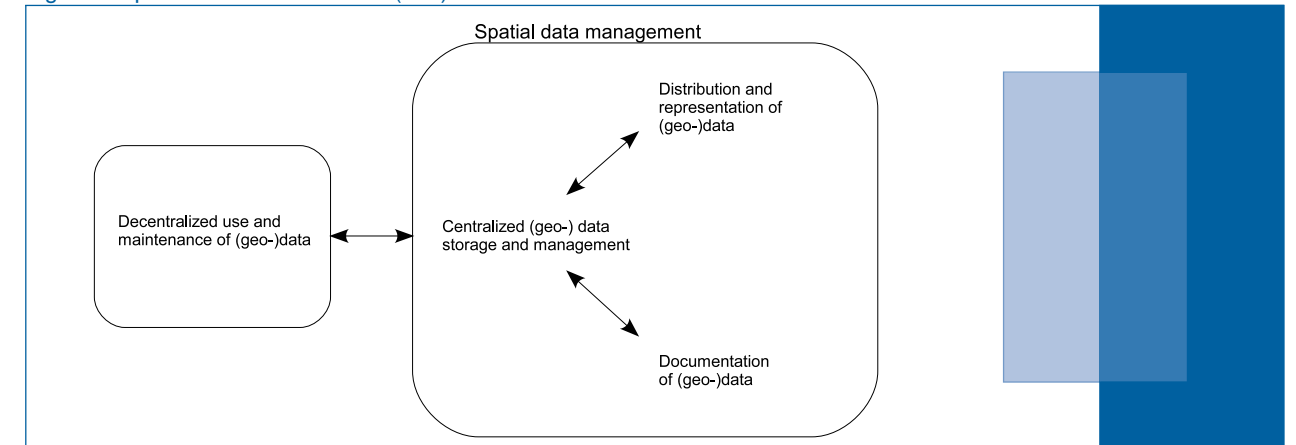
(1) The basis is the implementation of a metadata information system (MIS) to collect, manage and provide standard based metadata for all available project data. Metadata describe digital datasets to assess their suitability for a specific purpose. Each dataset is accompanied with additional information that ensures finding relevant data that fit particular needs.

Fig. 2:   Spatial data infrastructure (SDI)

Spatial data management

Distribution and representation of (geo-)data

Decentralized use and maintenance of (geo-)data

Centralized (geo-) data storage and management

Documentation of (geo-)data

ISO 19115 Geographic Information – Metadata

a standard of the International Organization for Standardization (ISO), defines how to describe geographical information and associated services, including contents, creation or publication date, spatial-temporal purchases, data format and quality, access and rights to use as well as contact details to communicate with related persons.

(2) The use of standards applied in the structure of metadata and software in use ensures interoperability and enables access to and the exchange of information between different systems.

(3) To enable the use of geospatial information as a common resource it is useful to store these data in a central geo-database. This ensures centralized data storage and management and decentralized use and maintenance of data.

(4) A web interface facilitates access to data and applications and enables the interactive networking and user related access to the data infrastructure. Thus, data and information become available online to all participants. This includes search and retrieval of data, upload of data into the database and edit of existing or creation of new metadata records.

GeoNetwork opensource is a standards-based Geographic Data and Information Management System, developed by the Food and Agriculture Organization of the United Nations (FAO-UN), the United Nations World Food Programme (WFP) and the United Nations Environment Programme (UNEP).
GeoNetwork opensource contains a number of attributes and features that are particularly useful to the data management needs of *Risk Habitat Megacity* and that led to its adoption.

Why *GeoNetwork opensource*?

The research initiative uses *GeoNetwork open-source* to manage a wide range of information including spatially referenced resources, to access project related datasets and to visualize spatial data through the internet. It permits both a centralized data management and a decentralized access to available data and information for all users from their local network.

Web based portal

The software is easy-to-use via web browser. The web interface supports the advanced search for various data in the project database or across multiple catalogues. It permits users to browse metadata sheets for detailed information, to directly access and download available datasets, to preview and analyse geospatial data in an interactive map viewer, to publish data by uploading them into the database and to edit existing or to create new metadata records using the online editing tools. [see Fig. 3]

Fig. 3:   Integrative structure of *GeoNetwork opensource*
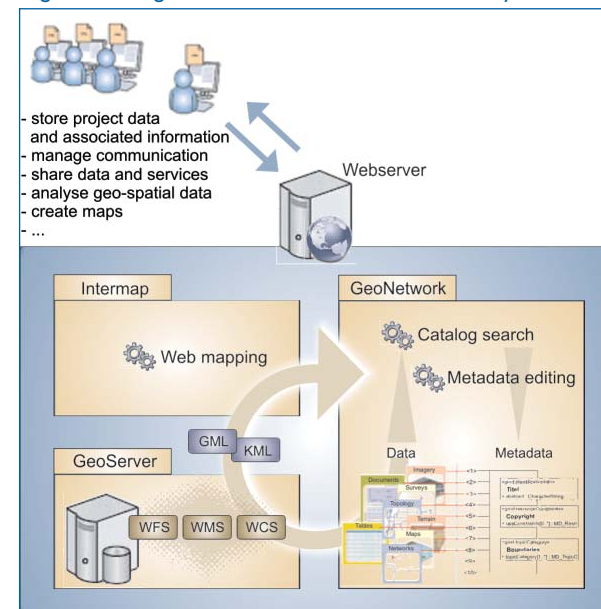
User and rights management

With an option to manage user and group accounts the software permits to assign different profiles and access rights to users. This enables to manage privileges for different user groups to access existing data and information resources.

Integrated Web Map Viewer

The embedded WebMapClient *InterMap* enables the visualization and analysis of geospatial data. Using *InterMap* different thematic layers from various sources can be combined to create interactive maps. This offers a simple possibility to generate information from geospatial data to support decision-making processes without additional software.

Use of standards

The management of data, the creation of metadata and the visualization of geospatial data uses international standards defined by the International Standards Organisation (ISO) and the Open Geospatial Consortium (OGC). The ISO standards for geospatial metadata define how to describe geographical information and associated services. The OGC implementations of standards for geospatial content and services and the OGC specifications for query metadata repositories enable data sharing and requesting geo-registered map images from one or more distributed geo-databases and retrieval of information from Web catalogues.

Free and Open Source Software

GeoNetwork opensource is free and open-source software, reducing costs for all users of the data management tools. It enables the modification of the application to fit specific requirements. *GeoNetwork opensource* is part of the Open Source Geospatial Foundation (OSGeo) software stack. Voluntary support, tutorials and user guides are provided through mailing lists, websites and online forums.

How protect data and ensure ownership rights?

The common management of project data owned or published by various collaborators at the different institutions raises the issue of managing and protecting ownership and use rights.

MetaCity provides three options for managing and protecting ownership and use rights.

>> Assigning privileges to different groups of users. For example, a user with a status as *Registered user* has by default only privilege to view metadata and to download protected data, but cannot upload data or create/edit metadata.

>> Permitting or restricting the access to a dataset by the owner (e.g. only showing the metadata but not downloading the actual dataset) or providing access only to particular user groups. For example, full access to datasets is given to project collaborators, while guests can't download any data unless permitted by the data owner.

>> Providing additional information on specific protection of privacy, copyrights or intellectual property and any special limitations on obtaining or using the resource in the metadata description.

The group and user management facilitates the data administration by authorized persons. In this way only the creator of a specific metadata record can upload data and edit or delete metadata. Only a *User administrator* can transfer metadata ownership to another user.

How to upload new datasets and create metadata records?

The metadata editor presents a powerful tool to add new records online. Using a standard based template system all collaborators may create metadata, add thumbnails, upload data, link to services and set access privileges to the dataset. [see Fig. 6+7] The templates are arranged in a way that the default view does not display the entire complexity of the ISO 19115 standard but can be expanded by new elements to fit specialized needs.

Adding and uploading a new document requires the description of identification information, e.g. title, abstract, descriptive keywords and its temporal and geographical extent. This information can be further specified by entering the current status of the document and its creation or publication date. All these information help users to find the right data.

For uploading geospatial datasets (e.g. satellite images, shapefiles or digital maps), the metadata templates offer the possibility to include specific information on the spatial representation typ, reference system, level of detail, data quality or available transformation parameters.

In addition, data upload involves the specification of contact information, access/use constraints or other restrictions on handling the resource as well as distribution information and options for obtaining the data.

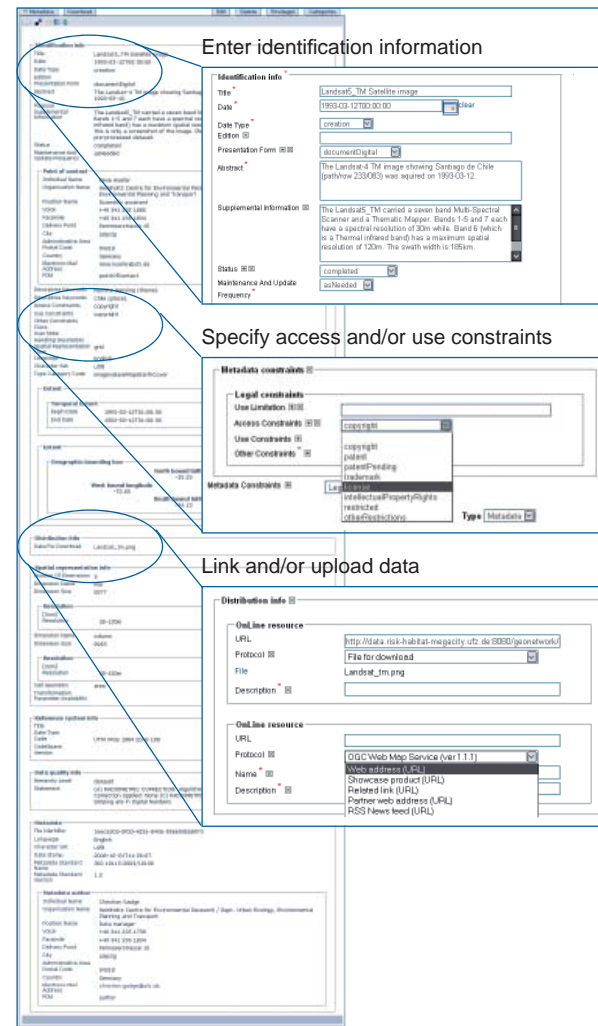Fig. 6:   Using the metadata editor

Enter identification information

Specify access and/or use constraints

Link and/or upload data

Fig. 7:   Assign privileges

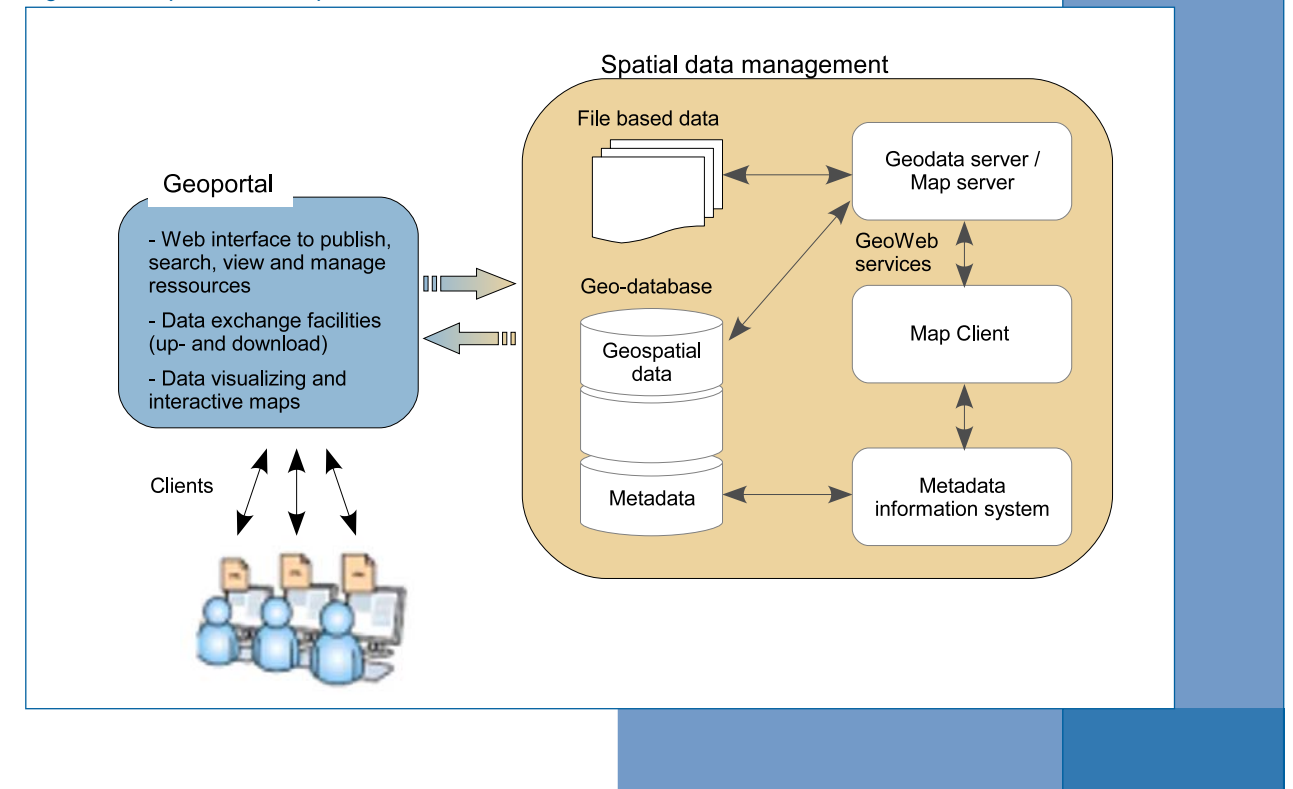# *MetaCity* - the *Risk Habitat Megacity* data portal

The implementation of a suitable solution to develop a (spatial) data infrastructure for the initiative has been done using several open source software components. [see Fig. 4]

Risk Habitat Megacity operates a modified application of *GeoNetwork opensource* that suits the particular demands, e. g. customization of the user interface or used metadata templates. The project database *MetaCity*, which holds all data of the initiative, is based on PostgreSQL, a powerful, open source object-relational database system. The PostGIS extension adds support for geographic objects in PostgreSQL, allowing it to be used as a spatial database for geographic information systems (GIS).

GeoNetwork opensource already includes a Web Map Server which enables the integration and visualization of geospatial data using the embedded Web Map Viewer *InterMap*. An Apache web server enables clients to communicate with and access to the application.

The application runs on a central server at the Helmholtz Centre for Environmental Research – UFZ. The portal provides an integrated user interface. Fast and easy access is possible via web browser at http://data.risk-habitat-megacity.ufz.de:8080/geonetwork/.

Fig. 4:   Components of a spatial data infrastructure

How to access *MetaCity*?

Collaborators of the *Risk Habitat Megacity* research initiative have a personal account to access and to work with the project database as an authorized user. Users belong to different status groups and own different privileges to manage these accounts:

>> User administrators with full rights for creating new users or changing users profiles within the own group;

>> Editors with full rights on creating new or editing/deleting old data within the own group, and

>> Registered users with right to download protected data.

All project collaborators have unrestricted access to available datasets. Their accounts include privileges to create new metadata records and upload data for joint use. In addition some user administrators were determined to manage the creation of new user accounts.

Other interested persons who do not belong to the initiative can be registered as guest. A guest account enables the search and download of data with explicitly defined accessibility rights to this group.

What kind of data does the database contain?

MetaCity contains a wide heterogeneous range of datasets in various stages of processing. These data were created and uploaded by the different collaborators for joint use. Available data are:

>> Reports, working papers, presentations and diverse tables (docs, pdf, ppt, xls, mdb, etc.);

>> Vector data (e.g. ESRI Shape files, OGC WMS/WFS): basic map data and thematic layers;

>> Raster data (e.g. GeoTIFF, IMAGE): satellite imagery covering the Metropolitan Region of Santiago de Chile and first results from the work with remote sensing data like land-use classification, digital elevation models or other modelling results;

>> Project-related information about external resources, websites or applications (e.g. homepages and data bases of Ministries or Data Observatories in Santiago de Chile).

How to retrieve the data?

The web interface facilitates to find data or a specific document. A powerful search function allows the search for free text within the entire record, keywords or a geographically referenced location of the dataset. Search criteria can be specified further to allow:

>> The search for elements related to the data content, e.g. keywords, title or abstract fields and specifics of the level of accuracy of the search;

>> The definition of the spatial extent, which allows users either to select particular areas of interest or to select a predefined region from the drop-down list. This is useful to find e.g. satellite images or thematic layers which cover a certain part of the research area or any dataset referenced to a specific geographical location;

>> The restricted search in terms of temporal extent, indicating a specific range of time referred to the data creation or publication date. It allows to apply further restrictions on the basis of additional parameters such as data source, data categories and data format e.g. to facilitate the identification of actual papers, latest state of work on a specific issue or data published by a particular user group.

The search results are shown as a list of the metadata records that match the request. Every record is associated with several options to retrieve the dataset, e.g. Show Metadata, Draw Map or Download Data. [see Fig. 5]

Fig. 5:   Data retrival options

Search the online database

Download your selected data sets in different formats or **display** and **analyse** them on-the-fly...

Show Metadata
Read additional information on the selected data set to assess their suitability for a specific purpose.

Draw Map
Display the selected data set as a map, dynamically drawn by the *Intermap* map client.

Download Data
Download the selected data set from the *Risk Habitat Megacity* server in different formats.