The Hydrodynamics of Astrophysical jets: Scaled Experiments and Numerical Simulations

GEOSPATIAL REFERENCE DATA ACCESS FOR EMERGENCY RESPONSE: DATA INVENTORY AND QUALITY ASSESSMENT

A. Ajmar [a, *],S. Balbo [b], A. Facello [b], F. Perez [b], S. Steffenino [b], R. Vigna [b]

[a] ITHACA (Information Technology for Humanitarian Assistance, Cooperation and Action), Via Pier Carlo Boggio 61, 10138 Torino

[b] Politecnico di Torino, DITAG, C.so Duca degli Abruzzi 24, 10129 Torino

KEY WORDS: Spatial Data Infrastructure, GMES, global geographic data inventory, OGC compliant on-line catalogue, metadata, spatial data quality assessment.

ABSTRACT:

Updated reliable and easily accessible reference base datasets are a key factor for the success of the emergency operations; the term Spatial Data Infrastructure (SDI) is often used to denote the relevant base collection of technologies, policies and institutional arrangements, that facilitate the availability of and access to spatial data. In the framework of the GMES initiative, ITHACA association is responsible for the implementation of an initial Global Monitoring for Environment and Security (GMES) service for geospatial reference data access covering areas outside Europe.

That activity is being accomplished by performing a complete and detailed inventory of available datasets and their preliminary classification according to general information acquired for each single data source, including maintenance, updates and distribution rules; all information are stored in an Open Geospatial Consortium (OGC) compliant on-line catalogue. Furthermore, identified datasets respecting the base requirements are submitted to a data quality evaluation process, intended to verify several criteria such as the completeness, the logical consistency, the positional, temporal and thematic accuracy, using a quantitative approach. Once the internal quality is verified, the level of concordance that exists between a product and user needs in a given context, defined as external quality, is estimated through a system of weights applied to each internal quality criteria. On the basis of the results of the data quality evaluation process, a specific data model is designed and implemented, together with Extraction, Transformation and Loading (ETL) procedures.

1. INTRODUCTION

1.1 The GMES program

GMES (Global Monitoring for Environment and Security) is the European Programme for the establishment of a European capacity for Earth Observation. It consists in a complex set of systems which collects data from multiple sources (earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors), processes these data and provides users with reliable and up-to-date information. The objective of GMES is to provide, on a sustained and operational basis, reliable and timely services related to environmental and security issues in support of public policy needs.

Policymakers and public authorities, the major users of GMES, will use the information to prepare environmental legislation and policies with a particular focus on Climate Change, monitor their implementation and assess their effects. GMES also supports the critical decisions that need to be made quickly during emergencies, such as when natural or man-made catastrophes and humanitarian crises occur.

Users will be provided with information through services dedicated to a systematic monitoring and forecasting of the state of the Earth's subsystems. Six thematic areas are developed: marine, land, atmosphere, emergency, security and climate change. A land monitoring service, a marine monitoring service and an atmosphere monitoring service contribute directly to the monitoring of climate change and to the assessment of mitigation and adaptation policies. Two additional GMES services address respectively emergency response (e.g. floods, fires, technological accidents, humanitarian aid) and security-related aspects (e.g. maritime surveillance, border control). GMES services are all designed to meet common data and information requirements and have global dimension. Into the GMES services (in particular Emergency Response, Land, and Security services) and to end-user applications reference data are among the most vital elements in order to provide a) the basic geographic framework on top of which additional spatial information can be produced and disclosed (e.g. land use/land cover maps, asset maps and damage assessment maps in response to crisis) and b) the set of relationships between the geographical components that will allow building the assessments, analyses and monitoring from combinations of datasets. Updated, reliable and easily accessible reference base datasets are a key factor for the success of emergency operations.

A Spatial Data Infrastructure (SDI) is often used to denote this reference base in terms of relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data.

ITHACA experience in this framework has lead the Association to be deputed for the implementation of an initial Global Monitoring for Environment and Security (GMES) service for geospatial reference data access (RDA) covering areas outside Europe.

This project has been differenced into 5 major activities or tasks:

- Analysis of non-European reference data availability and proposal of a strategy to access those data, subdivided into data inventory phase, data modelling phase and on-line metadata catalogue creation;
- Analysis of non-European reference data quality and consistency, defining and processing the quality and consistency indicators and analyzing them versus the requirements established in the previous task;

- Analysis of the UN-SDI initiative and contribute to it integrating GMES reference data;
- Identify reference data information gaps for areas outside EU;
- Demonstrator and validation.

This paper focuses especially on the first two phases on which a more advanced stage of implementation has been at present reached.

2. ANALYSIS OF NON-EUROPEAN REFERENCE DATA AVAILABILITY

2.1 Geographic data needs identification

In the first activity of the data inventory process, the spatial reference data required to support the possible organizations/projects interested in a RDA service outside Europe, with a specific focus on UN and GMES related initiatives, have been identified.

In order to correctly identify the base geographic data, the comprehensive Inventory Report prepared by UNGIWG TG 1 in conjunction with the Poverty Mapping Project Group (PMPG) of the FAO (Dooley, 2005) has been considered as a starting point. This report presents an inventory of global spatial data sources for the purpose of helping to identify a standard list of reference global databases for use across all UN agencies, suitable in particular for Emergency Preparedness and Response, Food Security and Poverty Mapping issues. This standard list of global databases has been then refined in order to correctly respond to the specific needs of the project and final users.

In particular, as GMES services must deal with different types of emergencies, considered reference data have to be broad and inclusive, with the aim to define a standard set of topics and sub-topics that can act as minimum backdrop on top of which adding the required thematic layers. Therefore, suitable spatial data have been separated into two groups for the geographic data needs identification:

- *Core Data*: the minimum geographic data baseline that is necessary for optimal use of all the different considered applications and activities. In the context of this phase, geographic data identified as fundamental and common to the different considered activities have been included in the Core Data. In particular, in the frame of GMES services and projects, the definition of Core Data can help to improve interoperability, reducing expenses resulting from the inevitable duplications;
- *Thematic Data*: considering in details each specific activity concerning GMES related projects and applications that will take advantage from the planned geospatial RDA service, additional specific geographic data can be identified.

In Table 1 are summarized the different data categories or topics and sub-topics containing the identified Core Data, whereas the chosen Thematic ones are presented in

Table 2. Currently, topics and sub-topics proposed for the Thematic Data constitute only a first proposal; they must be enriched and detailed according to the identified data sources.

TOPIC	SUB-TOPIC
BOUNDARIES: COSTAL, ADMINISTRATIVE AND AREAS OF SPECIAL INTEREST	Coastline and Maritime Boundaries; Country Political Level 0 Boundaries and Area of Dispute Boundaries; Sub-national Boundary Data (from 1^{st} Level on); Areas of conflict; Parks, conservancies, and Protected areas
SURFACE HYDROLOGY: SURFACE WATER BODIES, WATER POINTS, DRAINAGE AND WATERSHEDS	Drainage, Rivers and Flow Routing Databases; Surface Water Bodies (SWB) Databases; Watersheds and River Basin Databases; Water Points and Limnological Databases
HUMAN POPULATION: POPULATION CENTRES AND DISTRIBUTIONS	Population Centres Database ; Urban and Rural Population Density and Distribution Databases
TRANSPORTATION: ROADS, RAILWAYS, AIRPORTS, HARBOURS AND NAUTICAL ROUTES	Roads Databases; Railway line, Station, and Marshalling Yard Databases; Airport Databases; Harbour Databases; Nautical/Routes
BATHYMETRY AND TERRESTRIAL ELEVATION	Bathymetry Databases Terrestrial Elevation

Table 1. The topical areas of identified Core Data.

Two major requirements have to be met in the data inventory process, that can be defined as:

- research of globally consistent databases in regards to the geographic extent and scale (small to medium scale);
- research of geographic data with limited or absent constraints on their access and use in the projects. In particular, data freely accessible and available free-of-charge or at a low reproduction cost have a priority rank. At the moment, the review focuses on datasets freely available for academic, research and other non-commercial uses.

TOPIC	SUB-TOPIC
GEOPHYSICAL: GEOLOGY, GEOMORPHOLOGY, HYDROGEOLOGY, AND SOILS	Geology and Mineral Databases; Geo-morphology and Physiographic Databases; Hydro-geological/Aquifer Databases; Soils and Soil Properties Databases
SATELLITE IMAGERY, MOSAICS, LAND COVER AND VEGETATION DATA	Satellite Imagery and Mosaics; Satellite derivative Land Classification and Vegetation Databases
CLIMATIC DATA	Global Networks Databases; Satellite derivative Databases
AGRICULTURAL AND ECONOMIC ASSESSMENT	Agro-Ecological Assessment; Land Productivity Potential; National and Sub-national Agricultural Productivity data; UN global AEZ (agro-ecological zones); Food Insecurity, Poverty and Vulnerability databases
HUMAN HEALTH	Human Health Infrastructures and statistical Databases; Socio-Economic and Nutrition Indicators Databases; Poverty-Monitoring Indicators Databases
GLOBAL HAZARDS AND NATURAL DISASTERS	Earthquake and Tsunamis Databases Volcanic Databases; Floods and Landslides Databases; Fires and Droughts Databases

Table 2. The topical areas of identified Thematic Data.

2.2 Overview of geographic data inventory process

The main activity of the data inventory aims at identifying the potentially exploitable sources for all the geographic data necessary to fill the defined spatial needs. In particular, in this phase, the available up-to-date datasets, stand-alone or contained in databases, that supply the desired information for each sub-topic defined in Table 1 and 2 have been identified and described.

Data source identification is fundamentally based on the review of Web sites, geospatial data servers, and on-line searchable metadata databases. In particular, in addition to data sources identified by FAO in the aforementioned Inventory Report, metadata catalogues of major organizations that deal with geographic data (i.e. UNEP GRID[*], FAO[**]) have been considered.

Obtained outcomes are summarized in several tables indicating, for each sub-topic, all the available data sources identified during the inventory and compliant with the aforementioned restrictions. In these tables, the main features of the identified datasets have been described using the following fields (for their detailed description see Table 3): *Database ID* (database

[*] http://www.grid.unep.ch/data/geodataportal.php
[**] http://www.fao.org/geonetwork/srv/en/main.home

internal identifier, this field represents the source of the single data), *Database title* (name of the database that contains the cited dataset), *Database producer*, *Dataset name* (name of the dataset that supplies the geographic data for the specific sub-topic), *Dataset type*, *Dataset scale*, *Database last edition*, *Online resource access*, *Access constraints*.

Datasets are then proposed for the subsequent task of data quality and consistency analysis and for the task concerning gaps identification. Datasets are provided together with the indication of their suitability (identified categories are: datasets that have to be discarded, due to they are out-of-date or because of they are obtained just collecting other existing datasets without any relevant processing, datasets that have to be considered at a later stage, only if necessary and, finally, datasets that have to be considered).

The features described using this structure constitute the basis in support of the activities related to next tasks. Nevertheless, all the information collected during the data source review activity have been systematically held and organized into four different groups (
Table 3):

- *General Information*: the database source of the dataset is identified;
- *Technical Information*: a description of the content of the database is provided, together with details about the format and type of the data, the scale and the reference system;
- *Maintenance Information*: information concerning database creation and updating;
- *Distribution Information*: a description of the main database providers, with information about their access and use instruction and constraints.

As previously mentioned, data inventory process is aimed at researching data with the expressed requirements, gathering information necessary to the following tasks and correctly achieving the final step of metadata compilation, as will be described in the paragraph 2.4.

	FIELD	FIELD CONTENT
GENERAL INFORMATION	Database ID	Database internal identifier
	Database title	Name by which the cited resource is known
	Database alternate title	Short name or other language by which the cited resource is known
	Database producer	Identification of the institution, organization or company that is responsible for the production and maintenance of the resource
	Database sources	Information about the source data used in creating the resource or, at least, about the provider of source data
	Online resource	Online sources from which any other descriptive information about the database can be obtained
	Notes	Additional information
TECHNICAL INFORMATION	Database content	Description of the content of the database and list of the different datasets contained
	Datasets type	Information about grid or the vector spatial objects in the datasets
	Datasets format	Name of the transfer format(s) for the datasets
	Datasets scale	Level of detail of data, expressed as a scale factor or a ground sample distance
	Database geographic extent	Geographic coverage of the database (global, continental, national, or sub-national)
	Database spatial reference system	Information about the spatial reference system (name of reference system, projection, ellipsoid and datum used)
	Notes	Additional information
MAINTENANCE INFORMATION	Database reference date	Reference date (creation or publication or revision) for the cited database
	Database last edition	Last edition of the cited database
	Last edition date	Date of the last edition of the cited database
	Maintenance	Information about the frequency of updating and revision (i.e. changes, additions) of database
	Notes	Additional information
DISTRIBUTION INFORMATION	Access instructions	Information about the distributor of and options for obtaining the resource
	Access constraints	Any restrictions or limitations on accessing and obtaining the resource
	Use constraints	Any restrictions or limitations or warnings on using the resource; moreover limitations affecting the fitness for use of the resource
	Notes	Additional information

Table 3. Fields used for the databases description

2.3 Data model design

Data model design has been approached using a bottom-up approach, being the result of a reengineering effort. It started from the results of the review of global reference data holdings, with the objective to reorganize and harmonize the available information. It has been developed taking into consideration the outcomes of the user needs assessment and the requirements of the designed application scenarios. The implementation stage will follow an iterative process made of draft, submission, testing and adaptations. Data modelling process followed several distinct and consecutive phases, as described in Figure 4:

- Categorization of the reality: based on the outputs of the process of categorization of reference and thematic elements, a subdivision in topics and sub-topics is proposed, with the objective to organize available data and knowledge;
- Conceptual schema: it consists in the definition of entity classes, based on the previously defined sub-topics, and relationship assertions. The resulting schema is platform independent;
- Logical schema: consist in the description of tables and columns, object oriented classes, XML tags, etc. The logical schema, also definable as entity-relationship model, is the translation of the conceptual schema into a machine-readable format. That model is still in major part independent from the storage technology;
- Physical schema: describes the physical means by which data are stored. It represents the physical implementation of the logical schema in the production environment.

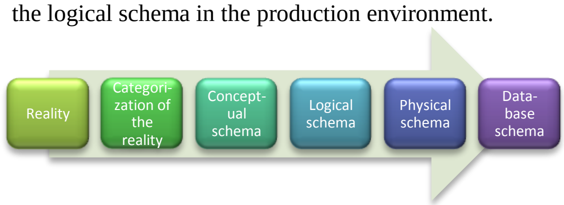

Figure 4. Data modelling phases

2.4 On line metadata catalogue

In order to provide search and discovery functionalities on the identified reference datasets, an Open Geospatial Consortium (OGC) compliant web based catalogue has been implemented.

The catalogue application that has been used is the Free and Open Source project GeoNetwork which provides instant search tool to the users on local and distributed geospatial catalogues. An effective search is provided by the use of metadata to specifically describe geographic data; to allow this functionality, metadata for all records found during the review process have been created. In particular, the ISO standard 19115:2003, that provides information on what elements should be included when describing a geographic resource, and ISO 19139:2007, that provides details on how this element should be coded in an XML document, have been taken into consideration.

The International Standard Organization defines a set of metadata elements to be used for describing the data; typically only a subset of the full number of elements is used. However, it is essential that a basic minimum number of metadata elements required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data) is maintained for a dataset, typically for catalogue purposes. These elements, named Core Metadata Elements, have been included in produced metadata. Moreover, further metadata elements, considered Optional by the ISO standard, have been also considered as crucial and proposed as mandatory during the metadata editing phase.

Following tables summarize the elements used for the dataset description, pointing out which are Core or Optional for the ISO standard 19115, and highlighting also which metadata elements, between the optional ones, have been compiled for all the datasets in the framework of GMES project (indicated with “xxx”). It has to be noted that some of proposed metadata elements are automatically updated by ESRI ArcCatalog, used to optimize the metadata creation process: these are indicated in tables with an asterisk “*”.

In particular, collected information have been grouped into 5 sections:

- metadata information (Table 5);
- distribution information (Table 5);
- resource identification information (Table 6);
- spatial representation – vector (Table 6);
- reference system information (Table 6).

			ISO 19115 reference	
Metadata section	**Metadata entity**	**Metadata element**	**Core metadata**	**Optional metadata**
Metadata Information	Metadata language		x	
	*Metadata character set			x
	*Last update		x	
	Metadata contact	Individual's name	x	
		Organization's name	x	
		Contact's position	x	
		Contact's role	x	
	*Scope of the data described by the metadata			
	*Scope name			
	*Name of the metadata standard used			x
	*Version of the metadata standard			x
Distribution Information	Distributor	Contact information		
		Individual’s name	x	
		Organization's name	x	
		Contact's position	x	
		Contact's role	x	
		Available format		
		Format name	x	
		Format version	x	
		Available format		
		Format name		x
		Format version		x
		Transfer options		
		Online source		
		- Online location (URL)	x	
		- *Connection protocol		
		- Function performed		xxx
		- Description		xxx
		Online source		
		- Online location (URL)	x	
		- *Connection protocol		
		- Function performed		xxx
		- Description		xxx

Table 5. Considered elements for metadata creation in the section Metadata Information and Distribution Information.

			ISO 19115 reference	
Metadata section	**Metadata entity**	**Metadata element**	**Core metadata**	**Optional metadata**
Resource Identification Information	Citation	*Title	x	
	Reference date	Date	x	
		*Type of date (creation)		
	Edition			x
	Edition date			x
	Party responsible for the resource	Individual's name	x	
		Organization's name	x	
		Contact's position	x	
		Contact's role	x	
		Contact information		
		Phone - Voice		x
		Address		
		- City		x
		- Administrative area		x
		- Postal code		x
		- Country		x
		- e-mail address		x
	Themes or categories of the resource		x	
	Theme keywords	Keywords		xxx
	Place keywords	Keywords		xxx
	Abstract		x	
	Dataset language		x	
	Resource maintenance	Update frequency	x	
	Resource constraints	Constraints - Limitation of use		x
		Legal constraints		x
		Access constraints		x
		Use contraints		x
		Other contraints		x
	*Spatial representation type			x
	*Processing environment			
	Spatial resolution	Dataset's scale - Scale denominator		xxx
		Ground sample distance		xxx
		Precision of Spatial data		xxx
		Units of measure, scale		
		- Units		xxx
		- *Conversion to metric		
	Resource's bounding rectangle	*Extent type:		x
		*Extent contains the resource		x
		*West longitude:		x
		*East longitude:		x
		*North latitude:		x
		*South latitude:		x
	Other extent information	Geographic extent:		x
		Bounding rectangle:		x
		*Extent type:		x
		*Extent contains the resource		x
		*West longitude:		x
		*East longitude:		x
		*North latitude:		x
		*South latitude:		x
	Point of contact	Individual's name	x	
		Organization's name	x	
		Contact's position	x	
		Contact's role	x	
		Contact information		
		Phone - Voice		x
		Address		
		- City		x
		- Administrative area		x
		- Postal code		x
		- Country		x
		- e-mail address		x
Spatial Representation - Vector	*Level of topology for this dataset			x
	Geometric objects	*Name:		x
		*Object type		x
		*Object count		x
Reference System Information	Reference system identifier	*Value		x

Table 6. Considered elements for metadata creation in the section Resource Identification Information, Spatial Representation, Reference System Information

3. DATA QUALITY ANALYSIS PROTOCOL

3.1 Overview of data quality process

The data quality assessment process has been performed into two phases: the first one concerns the definition of quality itself, while the second one aims at turning this concept into a quantitative approach.

Several authors distinguished two definitions of spatial data quality: internal quality and external quality. Internal quality corresponds to the level of similarity that exists between the data produced and the "perfect" data that should have been produced, that are also called "nominal ground". The evaluation of internal quality does not use the nominal ground that has no real physical existence since it is an "ideal" dataset, but uses a dataset of greater accuracy than the data produced, which is called "control data" or "reference data" (Devillers & Jeansoulin, 2006).

Instead the concept of external quality corresponds to the level of concordance that exists between a product and user needs, or expectations, in a given context. It is also often defined as "fitness for use" or "fitness for purpose" (Devillers & Jeansoulin, 2006). The evaluation of external quality can imply criteria that describe internal quality.

In concordance with these two definitions the proposed evaluation process can be divided into two phases, the first of which recalls the internal quality concept is differenced in the following phases:

- 11 Quality Indicators have been defined on the basis of the ISO specifications (ISO 19113, ISO 19115), which have been modified and adapted in relation to the aim of the GMES project;
- some confidence intervals have been set for each indicator, in order to define up to 5 levels of quality;
- for all candidate datasets selected for each sub-topic, comparisons, measurements and statistics are produced, in order to obtain a value for each indicator;
- for each indicator, a score between 1 and 5 is assigned to the current dataset, on the basis of the value obtained and the confidence intervals previously defined.

The result of this phase is a table, or Quality Matrix, where datasets and scores related to each indicator are summarized.

The second phase implies the evaluation of the external quality, which may be defined only in accordance with end users needs:

- a weight included in the range 0-1 has to be assigned to each indicator for the single sub-topic: this value represents the importance of any single quality indicator for the current sub-topic in respect to the final service. Some possible end users have been contacted and asked to define the weights, which then can be collected in a vector;
- finally, a vector of the total scores for each considered dataset is calculated as the result of an ordinary matrix product between the quality matrix and the vector of weights. Datasets are then ranked in order to reveal which is the most suitable for satisfying the end users' needs.

3.2 Quality Indicators

Hereinafter a short description of the selected Quality Indicators and their intervals of confidence is presented:

3.2.1 Geographic Extent: spatial coverage of the dataset. Usually, datasets with large nominal scale have a small extent and vice versa: when the end user needs a wide field of view, the extension can be insufficient to cover the area of interest and an operation of spatial merging of different datasets is necessary. This introduces an issue of integration and linking of different datasets in a cross border situation. On the purpose of creating a global service, the larger is the extent of a dataset the higher is its score (Table 7).

3.2.2 Licensing and constraints: different data holders deliver their spatial data with different license agreements. Two different elements are taken into consideration: the cost of the data and the possible limitation of use (Table 7).

Score	Geographic extent	Licensing and constraints
1	Local	Commercial data with limitations
2	Sub national	Commercial data
3	National	-
4	Continental	Free data with limitations
5	Global	Public domain data

Table 7. Requirements and respective scores for the indicators 1 and 2

3.2.3 Scale Denominator: this information is usually provided in the metadata or published by the data holder; the reference intervals are reported in Table 8. In case it is not provided, some geoprocessing operations are performed in order to give an estimation of the scale.

3.2.4 Update: different datasets may have different update rate: the more recently the data have been updated, the more reliable the data are assumed to be (Table 8). The indicator refers to the level of updating of the data used to produce a specific datasets (i.e. for a Digital Elevation Model, the date range of the satellite/aerial images used to create the model itself).

Score	Scale denominator	Update
1	> 1.000.000	From 6 to 10 years ago
2	>500.000 - <=1.000.000	From 3 to 5 years ago
3	>250.000 - <=500.000	In the past 2 years
4	>100.000 - <=250.000	Annually (planned)
5	<= 100.000	Continuous

Table 8. Requirements and respective scores for the indicators 3 and 4

3.2.5 Fitness for use in cartographic representation(I): the generic concept of Fitness For Use varies on the basis of the applications in which the data are exploited: the present Quality Indicator considers map representation both in printing and in displaying (i.e. Web applications). Some datasets need some geoprocessing operations in order to obtain the best performance in terms of representation. The number and the importance of the geoprocessing operations are used to evaluate the data quality: in general the higher and the more complex are the operations, the lower is the quality (Table 9).

Score	Fitness for Use I
1	Need of huge correction
2	Need of intermediate correction, both on natural boundaries and abstract edges
3	Need of few corrections on natural boundaries
4	Need of few corrections on abstract element edges
5	No need of correction

Table 9. Requirements and respective scores for the indicator 5

3.2.6 Fitness for use in territorial analysis(II): in the field of application of land and territorial analysis, data aggregations based on their spatial location are often needed. The accuracy in respect of the administrative boundaries, in terms of complexity of the potential geoprocessing operations needed to obtain them, and the possibility to contain in the attribute table a clear reference to the country of membership, e.g. extended country name or code version, are more important than the appearance in phase of visualization, including inconsistencies among different sub-topics (Table 10).

Score	Fitness for Use II
1	Focus on natural boundaries
2	High level of correction needed to extract administrative boundaries
3	Intermediate level of correction to extract administrative boundaries
4	Low level of correction to extract administrative boundaries
5	Focus on administrative boundaries

Table 10. Requirements and respective scores for the indicators 6

3.2.7 Integration: the implementation of a service for geospatial data needs of a perfect integration among datasets, considering both the same or different sub-topics (e.g. hydrology and elevation data). The integration issue becomes even more complex if the datasets have different nominal scales, extents, geometrical and thematic structures. Based on this assumption, the datasets are considered more suitable if they're supplied as integrated in a thematic or multipurpose database, since the relationship issues between datasets are solved by data holders (Table 11).

3.2.8 Data integrity: it implies the respect of topological relationship by the geometric shape of the features belonging to a dataset: some topological rules can be defined at sub-topic level on the basis of their specific role. Each feature is compared with the whole dataset and their spatial relationship is controlled in order to find violations of the rules by mean of a specific software. The value of the indicator consists in the percentage of the sum of the errors, for all the controls performed on a dataset, over the number of features observed in that dataset (Table 11).

Score	Integration	Data integrity
1	Stand alone dataset	> 5%
2	-	2% - 5%
3	Part of thematic database	0.5% - 2%
4	-	0.1% - 0.5%
5	Part of comprehensive database	< 0.1%

Table 11. Requirements and respective scores for the indicators 7 and 8

3.2.9 Positional accuracy: it has been evaluated through the visual comparison of the shape of a sample from the dataset in exam with features referring to the same object but coming from dataset with a higher accuracy (at least an order of magnitude). Since the accuracy is different, during visualization at scales bigger than its reference scale, the dataset in exam could appear simplified while the reference one normally appears more detailed: the bigger is the simplification of the features or the possible errors in them, the lower is the score assigned to the dataset (Table 12).

3.2.10 Thematic accuracy: random controls are performed on attribute values: a number of feature is investigated in order to verify the correctness of what reported in the fields of the attribute table of the respective dataset. The percentage of the number of errors found over the number of features observed constitutes the value assumed by the indicator (Table 12).

3.2.11 Completeness: the absence of gaps in spatial data is one of the major requests arose from the user needs assessment. Datasets are randomly investigated in order to reveal missing features: the percentage of missing features over the number of features observed constitutes the indicator (Table 12).

Score	Positional Accuracy	Thematic Accuracy	Completeness
1	Coarse	> 5%	> 5%
2	Low	2% - 5%	2% - 5%
3	Intermediate	0.5% - 2%	0.5% - 2%
4	High	0.1% - 0.5%	0.1% - 0.5%
5	Fine	< 0.1%	< 0.1%

Table 12. Requirements and respective scores for the indicators 9, 10, 11

4. CONCLUSIONS

Present paper describes a methodology for building a reference geographic database for areas outside Europe starting from a list of core and thematic topics. An inventory of available data sources allows to obtain a complete overview of the state-of-the-art of the geographic information. A data quality evaluation process, based on indicators, is the key for identifying, among the possible sources, those that are authoritative, reliable, up-to-date and appropriate for the purposes of the final users.
On the basis of inventory and data quality outcomes, a data model can be defined and a geodatabase implemented with information deriving from selected sources. A metadata catalogue application supports the search and discovery of the geodatabase content.
The developed methodology is complete and exhaustive, and allows to met all requirements in terms of building a system for reference data access for emergency response; due to the huge amount of data to be inventoried, it constitute a significant resources consuming activity. In order to increase efficiency, several tools for optimizing the process are under consideration. Moreover, the effectiveness of the solution will be tested through the implementation of web applications devoted to specific emergency management activities, such as flood alert systems and disaster response.

5. REFERENCES

Devillers R. & R. Jeansoulin, 2006. *Fundamentals of Spatial Data Quality*. ISTE Ltd, London

Dooley J.F.Jr., 2005. An inventory and comparison of globally consistent geospatial databases and libraries. Environment and Natural Resources Series No.19 - FAO, Rome 2005 - online resource (accessed Apr. 2011) http://www.fao.org/docrep/008/a0118e/a0118e00.htm#Contents

GMES in brief. http://www.gmes.info/pages-principales/overview/gmes-in-brief/ (accessed 11 Apr. 2011)

ISO/TC 211, 2004. 19113 Geographic Information – Quality Principles, International Organization for Standardization (ISO).

ISO/TC 211, 2004. 19115 Geographic Information – Metadata, International Organization for Standardization (ISO).