

POLITECNICO DI TORINO

DOCTORATE SCHOOL

Ph.D. in Environment and Territory – XIV Cycle

Final Dissertation

**A reference data access service in support
of emergency management**

**Data quality assessment protocol, publication and
exploitation of the results**



Simone Balbo

Tutor

Prof. Piero Boccardo

Doctorate course coordinator

Prof. Attilia Peano

April 2012

Table of Content

| | |
|---|----|
| INTRODUCTION..... | 1 |
| I. GLOBAL MONITORING FOR ENVIRONMENT AND SECURITY | 5 |
| I.1 GMES in brief..... | 5 |
| 1.1.1 Access to the services..... | 6 |
| 1.1.2 Cost of the services..... | 7 |
| 1.1.3 Services availability..... | 7 |
| I.2 European Commission notes on GMES..... | 7 |
| I.3 Services specification..... | 8 |
| 1.3.1 Land Monitoring..... | 8 |
| 1.3.2 Marine Environment Monitoring..... | 8 |
| 1.3.3 Atmosphere Monitoring..... | 8 |
| 1.3.4 Security..... | 9 |
| 1.3.5 Climate Change..... | 9 |
| 1.3.6 Emergency Management..... | 9 |
| I.4 Projects Launch | 11 |
| I.5 7 th Framework Programme..... | 11 |
| I.6 SAFER..... | 12 |
| 1.6.1 Cambodia floods example..... | 12 |
| I.7 Final notes..... | 13 |
| 2. INITIAL GMES SERVICE FOR GEOSPATIAL REFERENCE DATA ACCESS..... | 15 |
| 2.1 Tender Specifications and Monitoring..... | 16 |
| 2.1.1 Objective..... | 16 |
| 2.1.2 General considerations..... | 18 |
| 2.2 Technical Proposal and Stakeholders | 19 |
| 2.2.1 Contractor..... | 19 |
| 2.2.2 Fundamental activities | 20 |
| 2.3 Identification of users and needs assessment..... | 23 |

| | | |
|--------|---|----|
| 3. | PROJECT ACTIVITIES STATUS | 25 |
| 3.1 | Data inventory and related activities | 25 |
| 3.1.1 | <i>Data inventory</i> | 25 |
| 3.1.2 | <i>Data modelling</i> | 27 |
| 3.1.3 | <i>Metadata and OGC compliant on-line catalogue</i> | 28 |
| 3.1.4 | <i>Demonstrator applications</i> | 28 |
| 3.2 | Analysis of data quality and consistency | 29 |
| 3.3 | Analysis of the UNSDI initiative | 29 |
| 3.3.1 | <i>Current status of the UNSDI</i> | 31 |
| 3.4 | Identification of reference data information gaps | 31 |
| 3.4.1 | <i>Gaps in the indicator Scale Denominator</i> | 32 |
| 3.4.2 | <i>Gaps in the indicator Update</i> | 32 |
| 3.4.3 | <i>Proposal of technical provision</i> | 32 |
| 3.5 | Demonstrators and Validation | 34 |
| 3.5.1 | <i>Fires Monitoring System</i> | 34 |
| 3.5.2 | <i>ERDS</i> | 36 |
| 3.5.3 | <i>GeoNode for CoSA</i> | 37 |
| 4. | DATA QUALITY ASSESSMENT PROTOCOL | 39 |
| 4.1 | Data Quality definition | 40 |
| 4.2 | Data Quality Assessment procedure | 41 |
| 4.3 | Data Quality Indicators definition | 42 |
| 4.3.1 | <i>Geographic Extent</i> | 42 |
| 4.3.2 | <i>Licensing and Constraints</i> | 42 |
| 4.3.3 | <i>Scale Denominator</i> | 43 |
| 4.3.4 | <i>Update</i> | 44 |
| 4.3.5 | <i>Temporal Extent</i> | 44 |
| 4.3.6 | <i>Fitness for Use</i> | 44 |
| 4.3.7 | <i>Thematic Richness</i> | 45 |
| 4.3.8 | <i>Integration</i> | 45 |
| 4.3.9 | <i>Data Integrity</i> | 46 |
| 4.3.10 | <i>Positional Accuracy</i> | 47 |
| 4.3.11 | <i>Thematic Accuracy</i> | 47 |
| 4.3.12 | <i>Completeness</i> | 48 |
| 4.4 | Overall Evaluation | 48 |
| 4.4.1 | <i>Quality Matrix</i> | 48 |
| 4.4.2 | <i>Vector of Weights</i> | 49 |
| 4.4.3 | <i>Final Scores and Ranking</i> | 49 |
| 4.5 | Notes on the protocol application | 50 |
| 4.6 | Conclusions on the protocol | 50 |
| 5. | DATA QUALITY PROTOCOL APPLICATION | 53 |
| 5.1 | Country Political Level 0 Boundaries | 53 |
| 5.1.1 | <i>Datasets introduction</i> | 54 |
| 5.1.2 | <i>Indicators evaluation</i> | 55 |
| 5.1.3 | <i>Overall evaluation</i> | 61 |

| | | |
|-------|--|-----|
| 5.2 | Urban and Rural Population Density and Distribution..... | 63 |
| 5.2.1 | <i>Datasets introduction</i> | 64 |
| 5.2.2 | <i>Indicators evaluation</i> | 64 |
| 5.2.3 | <i>Overall evaluation</i> | 71 |
| 5.3 | Conclusion and comments | 72 |
| 6. | GEONODE: A WEB APPLICATION FOR DATA SHARING..... | 75 |
| 6.1 | GeoNode introduction..... | 76 |
| 6.2 | GeoNode functionalities | 76 |
| 6.2.1 | <i>Data sharing</i> | 77 |
| 6.2.2 | <i>Data uploading</i> | 77 |
| 6.2.3 | <i>Data downloading</i> | 78 |
| 6.2.4 | <i>Map creating</i> | 79 |
| 6.2.5 | <i>Metadata management</i> | 81 |
| 6.2.6 | <i>Data sensitivity</i> | 81 |
| 6.3 | GeoNode and the OpenDRI initiative | 81 |
| 6.4 | GeoNode components..... | 82 |
| 6.4.1 | <i>PostgreSQL and PostGIS</i> | 83 |
| 6.4.2 | <i>GeoServer</i> | 83 |
| 6.4.3 | <i>Open Geospatial Consortium and OGC standards</i> | 84 |
| 6.4.4 | <i>GeoNetwork</i> | 87 |
| 6.4.5 | <i>GeoExplorer</i> | 89 |
| 6.4.6 | <i>OpenLayers</i> | 90 |
| 6.4.7 | <i>Django</i> | 91 |
| 7. | GEONODE DEVELOPMENT FOR QUALITY ASSESSMENT RESULTS | 95 |
| 7.1 | Development goals | 95 |
| 7.2 | Medium choice | 96 |
| 7.3 | Development approach | 96 |
| 7.4 | Data model..... | 98 |
| 7.5 | Publication of the quality assessment results..... | 100 |
| 7.6 | Selection of the best quality dataset..... | 101 |
| 7.7 | Open issues and further development..... | 104 |
| | CONCLUSION | 105 |
| | BIBLIOGRAPHY | 109 |
| | WEBOGRAPHY | 111 |
| | ANNEX A – GEONODE QUALITY APPLICATION CODE | |
| | ANNEX B - MALAWI DISASTER AND NATURAL RESOURCES SPATIAL DATA SHARING PLATFORM: TECHNICAL PROPOSAL | |

Introduction

In the field of natural disasters recovery and reduction and of emergency management geo-referenced information is strongly needed. In my personal experience obtained in the three years period spent at ITHACA, during the shorter at GFDRR Labs and through the work done indirectly with UN-WFP, after a natural disaster occurs, the experts in geomatics are often asked to provide answers to questions such as: where did it occur? How many people have been involved? How many infrastructures have been damaged and to what extent? How much is the economical loss?

Geomatics can give answer to all these questions or give significant help in addressing operations in order to get the answers. The goal can be reached both with the use of base reference data, the ones usually contained in the classic cartography, and by exploiting value added information coming from satellite and aerial data processing, classic surveys and GPS/GNSS acquisition on the field.

The activities object of this thesis have been performed in the framework of a European project that aim to provide services for Earth Observation: it is called Global Monitoring for Environment and Security. The document is organized in order to follow the project logical line rather than covering the respective timeline.

In chapter 1 the European Commission initiative called Global Monitoring for Environment and Security is introduced. At the international level the environmental variables are taken into deep consideration for investigating long term processes such as the climate change and monitored for natural disasters early detection and mapping. The European Union in recent years invested much in research and in development of a set of global monitoring services of the environmental variables for citizens' security.

Since the services have a strong spatial and mapping component and since the technology used is centred on the internet and the information digitalization is a need, the geomatics acquires a key role in this framework. Also in this context geospatial base reference data are fundamental because they give information on the Earth status and allow to detect the changes in natural components such as soil, air and water.

In chapter 2 an EU call for tenders regarding the implementation of an initial GMES service for geospatial reference data access is introduced as part of the preparatory actions of the GMES

initiative. In particular environmental emergency management and land monitoring are the main purposes of the data access; moreover the project subdivision into two lots and the respective content is explained.

Afterwards the chapter focuses on the project second lot and presents the technical proposal of the call for tenders winner as a suitable approach for building and implementing a reference data access with a global extent. Furthermore the project end users and the tender contractor are introduced.

In chapter 3 the project status in relation to all the activities is presented, following the tasks organization of the tender contractor. Firstly the data inventory phase is described and the data categorization is explained. Secondly the data quality assessment is introduced together with the concepts implemented during the definition of a suitable data quality protocol and the operations of protocol application. Thirdly the data repository collected during the inventory phase is analysed by mean of the data quality concepts, in order detect gaps in the information content, and technical proposals for gaps filling are reported. Finally a set of data access demonstrators is described: they are online tools developed on the basis of free and open source software, that comprehends WebGIS tools for data visualization on maps. Moreover demonstrators are based on real cases of natural disaster events such as floods and fires and are developed following the feedbacks of user identified on purpose.

Geospatial data quality is among the core activities for implementing a reference data access: it is the topic of chapter 4. Initially a bibliographic review allowed to provide a definition of data quality. A data quality protocol was conceived on the basis of the definition; in particular the protocol was targeted on the specific needs related to the GMES initiative, that is for making a choice among a set of candidate datasets in a certain category. The goal is to allow to establish a quality level that is used in a comparison among various data sources. Afterwards the protocol was also used for investigating datasets in order to find information gaps.

The data quality protocol follows a quantitative approach based on scores and weights; it is general enough to be adapted to different data categories in terms of content and different file and geometry types. Moreover the protocol was defined in order to take into account the end user's needs. In details the protocol is based on a set of quality indicators that describes the quality level on a complete set of point of views: topographic, geometric, semantic, content richness, spatial richness. It was conceived for a rapid and ready application.

During the project activities the protocol was applied to the set of datasets gathered in the inventory phase, which were gathered according to data categories; examples of protocol application are provided in chapter 5. The steps of the evaluation process and the results regarding two data categories are reported; they represent two different file types, vector and raster.

Afterwards the advantages and disadvantages that the quality assessment highlighted on the protocol are explained; such comments are useful for improving the protocol itself in order to increase the easiness of the application and change potential difficult steps. Indeed the data quality assessment allowed to collect information on the validity of the instruments, the processes and the techniques suggested on the protocol.

In chapter 6 a Web platform for geospatial data sharing is introduced and thoroughly described as one of the GMES RDA Lot 2 project demonstrator. It is a free and open source application

originated and funded mainly in the area of interest of the Global Facility for Disaster Recovery and Reduction of the World Bank, that covers natural disasters and environmental emergencies. In particular the software, called GeoNode, contains tools and applications such as metadata catalogue, data viewer and map composer all directly online by taking advantage of the Web 2.0 technology.

This software represents innovation in that it combines the open source approach together with the potentiality of the Web, the popularity and the collaborative side of the social networks and the phenomenon of the volunteer mapping that happened after some dramatic natural disasters all over the world.

Furthermore this platform is promoted by GFDRR as operational tool for building resilience in developing countries against natural disasters and as technical solution for supporting spatial data infrastructures. Agencies, NGOs, governments and volunteers are encouraged to collaborate and exchange data by mean of this platform.

In chapter 7 the GeoNode and the data quality assessment are brought together in order to exploit the advantage of a combination of the two. The GeoNode developers community encourages users to test, develop and improve the application; accordingly it was developed for the specific purposes of the data quality in the GMES RDA Lot 2 project.

Originally two separate goals were defined: on one side the availability of data quality assessment results required their publication. Therefore a GeoNode extension was developed, aimed at publishing the results together with the other metadata in the specific application data page. On the other side the quality protocol contained a tool for data choosing, based on quality level and user needs; the GeoNode was then extended in order to implement the data choosing tool.

In particular the document contains all the details and technical choices as well as an analysis of the possible alternative use and a first approach feedback from the users. Moreover the complete code written for the development is contained in the annexes.

In conclusion the thesis objective is describing the role of the geomatics in natural disasters and environmental emergency management: geomatics is the efficient combination of geo-referenced data and information technology. In particular the topic of the thesis is constituted by the activities required for implementing a set of services for the environmental variables observation, focusing on the framework of the European Union initiative GMES. The content is mainly related to a practical work of implementation of a reference geospatial data access by taking advantage of all the innovation and the research in the field. In particular significant efforts are devoted to the data quality issue; the implementation of the data quality definition into an operational protocol, the data quality assessment operations and the publication of the results as well as the provision of a tool for data choosing among valid alternative datasets, based on their quality level, constitute the core of the work.

I. Global Monitoring for Environment and Security

In everyday life the human being is in constant contact and continuously compares himself with the natural environment; human life conditions are strictly related to the environmental conditions. Many research centres and universities put continuous efforts in studying the relations between human and environment in various forms, such as investigating the land cover and the vegetation status for agricultural and forest exploitation as well as the status of the oceans and the regional seas for fishing and commercial routes. The status of the atmosphere is also studied due to its relationship with the climate change. The environment is also investigated for reason of managing the emergency caused by natural disasters and for human security from risks of other kind.

In this context at the international and European level there is a strong attention towards the development of this topic and especially on the relationship with technology; the awareness of the public towards these issues is also a key element. Indeed in the 7th Framework Programme huge investments were programmed and allocated in order to develop an initiative that is centred on the environmental monitoring.

In this chapter this initiative is introduced in order to represent the context in which the work of the thesis has been carried out: it is called Global Monitoring for Environment and Security (GMES) and is the European Programme for the establishment of a European capacity for Earth Observation.

I.1 GMES in brief

The well-being and security of future generations are more than ever dependent on everyone's actions and on the decisions being made today on environmental policies. To take the right actions, decision makers, businesses and citizens must be provided with reliable and up-to-date information on how our planet and its climate are changing. The European Earth monitoring programme GMES provides this information.

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 [GMES website].

The programme envisages the users to 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 related to floods, fires, technological accidents, humanitarian aid and security-related aspects such as maritime surveillance and border control. GMES services are all designed to meet common data and information requirements and have global dimension.

In practice, GMES consists in a complex set of systems which collects data from multiple sources like earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensor; then it processes these data and provides users with reliable and up-to-date information through the services mentioned above. Some of these systems and data sources already exist today, as well as prototype services but many developments are still required in all domains.

The GMES initiative is led by the European Union: the programme coordination and management is ensured by the European Commission. The setting up of initial versions of the GMES services have been assigned to several projects partly financed through the 7th Research and Development Framework Programme of the European Union, while the developments related to the observation infrastructure are performed under the aegis of the European Space Agency for the space component, such as the Sentinel missions, and of the European Environment Agency and the Member States for the in situ component. The sustainability of the GMES operational services are going to be ensured through public funding from EU, intergovernmental agencies, and Member States. These services should be accessible to any organisation or citizen, because they are considered as “public goods” [GMES website].

Based on the GMES services, many other value-added services tailored to more specific public or commercial needs are envisaged to be developed: for instance forecasting services with a local scope and others including socio-economic data. This is going to stimulate the downstream sector.

1.1.1 Access to the services

Primary users of GMES are public policy-makers, public bodies responsible for environment and security matters, and business operators. However, access will not be limited to these users and the data and information provided by GMES will be accessible in principle to any citizen.

A few exceptions will however exist in some specific situations where security is at stake. In such cases, a limited access should probably be envisaged. In practice GMES information will be provided to users through a set of services, most of them being available online.

1.1.2 Cost of the services

The main operational services will be free of charge for the end users.

Other policies are envisaged for the downstream services, which correspond to those services that are going to be implemented outside the scope of the GMES governance and without EU public funding. They will be developed by public or private operators with the objective to meet specific needs in terms of group of users or location.

Downstream services will provide their users with added-value by combining the information provided by the GMES services with additional data, for instance socio-economic data. Depending on the business model adopted by each service provider, these downstream services could be either free for the final user and funded through advertising or associated to a fee.

1.1.3 Services availability

The GMES services that are being developed through the 7th Framework Programme projects are currently in a pre-operational validation phase. They should enter into initial operations during the period 2011-2013, with the objective to be fully operational by 2014.

The availability of downstream services depends on the maturity of the GMES services and on the business plans of future downstream providers.

1.2 European Commission notes on GMES

The GMES service component depends on Earth observation data, collected from different areas, and instruments, examples of which are: space by mean of sensors mounted on satellites, air through airborne instruments and balloons to record stratosphere data, water by mean of floats and shipboard instruments or land with measuring stations and seismographs. These facilities are called the GMES infrastructure component; non-space based installations in the GMES infrastructure component are generally referred to as "in situ component".

By securing the sustainability of an information infrastructure necessary to produce output information in the form of maps, datasets, reports, targeted alerts and others, GMES helps people and organisations to take action, make appropriate policy decisions and decide on necessary investments. GMES also represents a great potential for businesses in the services market, which can then be able to make use of the data and information it provides according to a full an open access principle.

Earth observation-based services already exist in Europe, but they are dispersed at national or regional level and cannot rely on a sustainable observation capacity. With the exception of meteorological services, long-term availability and reliability of information is not guaranteed. This is why, in order to contribute to improve its response to ever growing challenges of global safety and climate change, Europe develops a sustained and reliable Earth observation system of its own [EC website, GMES pages].

I.3 Services specification

GMES services are dedicated to the monitoring and forecasting of the Earth's subsystems. They can be divided into:

- land, marine and atmosphere services; providing systematic monitoring and forecasting the state of the Earth's subsystems at regional and global levels;
- emergency and security services; providing support in the event of emergencies and humanitarian aid needs, in particular to civil protection authorities, also to produce accurate information on security related aspects, such as maritime surveillance, border control and global stability;
- climate change service; helping to monitor the effects of climate change and assessing mitigation measures.

I.3.1 Land Monitoring

Benefiting from Earth Observation satellite data, the GMES land monitoring service provides accurate and cross-border harmonised geo-information at global to local scales. The service provides geographical information on land cover including its seasonal and annual changes and monitors variables such as the vegetation state or the water cycle. It has a wide range of applications for use in land use and land cover change, soil sealing, water quality and availability, spatial planning, forest monitoring and global food security.

The pre-operational land monitoring service of GMES is currently provided through the EU-funded project geoland2: it develops a set of three mapping services that serve as base for the provision of value added information [GMES website, Land Monitoring page].

I.3.2 Marine Environment Monitoring

The GMES marine environment monitoring service provides regular and systematic reference information on the state of the oceans and regional seas. It addresses various main domains, such as marine safety which concerns marine operations, oil spill combat, ship routing, defense, search and rescue, and marine resources, for instance fish stock management. Other domains are: marine and coastal environment that concerns water quality, pollution and coastal activities; climate and seasonal forecasting, topics of which are climate monitoring, ice survey and seasonal forecasting.

The pre-operational marine service of GMES is currently provided through the EU-funded project MyOcean [My Ocean website].

I.3.3 Atmosphere Monitoring

The GMES atmosphere monitoring service provides data records on atmospheric composition for recent years, current data for monitoring present conditions and forecasting the distribution of key constituents for a few days ahead. The service addresses topics such as greenhouse gases, reactive gases, ozone layer and solar UV radiation, aerosols, which affect temperature, air quality and the transmission of solar radiation.

The pre-operational atmosphere service of GMES is currently provided through the EU-funded project MACC [MACC website].

1.3.4 Security

The GMES services for Security applications aim at supporting the related European Union policies in the priority areas of border surveillance, maritime surveillance and in support to EU External Action.

In the area of border surveillance, main objectives are to reduce the number of illegal immigrants entering the EU undetected, to reduce the death toll of illegal immigrants by rescuing more lives at sea and to increase internal security of the European Union as a whole by contributing to the prevention of cross-border crime.

Concerning maritime surveillance, the overall objective of the European Union is to ensure the safe use of the sea and to secure Europe's maritime borders. The corresponding challenges mainly relate to safety of navigation, marine pollution, law enforcement, and overall security.

As a global actor, Europe has a responsibility in promoting stable conditions for human and economic development, human rights, democracy and fundamental freedoms. In this context, a main objective of the EU is to assist third countries in a situation of crisis or emerging crisis and to prevent global and trans-regional threats having a destabilising effect. It can also contribute towards improving crisis prevention, preparedness and response capacities.

Pre-operational versions of GMES security services addressing border surveillance and support to EU External Action are currently provided through the EU-funded project G-MOSAIC [gMosaic website].

1.3.5 Climate Change

GMES closely monitors each of the Earth's different subsystems and help to better monitor and understand climate change. The climate change service takes advantage of the Atmosphere Monitoring service for what concerns greenhouse gases and aerosol concentrations; it also takes advantage of the Marine Environment Monitoring service that continually observes how oceans store carbon and transport heat from one region of the world to another; finally it uses the Land Monitoring service that provides data on the vegetation state and carbon fluxes in order to quantify terrestrial sinks and carbon sources.

1.3.6 Emergency Management

The GMES emergency management service addresses, with a worldwide coverage, a wide range of emergency situations resulting from natural or man-made disasters. Larger space will be given to this service because it constitutes the framework of the rest of the thesis.

This service covers in particular floods, forest fires, landslides, earthquakes, volcanic eruptions and humanitarian crises. The pre-operational emergency management service of GMES is currently provided through the EU-funded project SAFER [SAFER website].

Floods

The flood service platform covers both the plain flood and the flash flood risk analysis at European level. The Plain Flood Risk Management service provides decision-making tools which include the mapping of past floods, potential flood damage maps and flood information service. The Flash Flood service must be considered as a demonstrator service to be implemented within the framework of each local or national hydro-meteorological operator having specific responsibilities in a country, region or basin prone to flash floods [GMES website, Emergency Management page].

Forest Fires

The fire services include: Fire Monitoring at Middle Resolution that provides hot spots as well as a rapid mapping of burnt areas; Burnt Scar Mapping that provides burnt scars maps based on high resolution imagery from space, helps users in analysing fire impacts, performs vegetation damage assessment and designs recovery plans; Global Fire Risk service that provides forecast of danger maps, which is especially useful in supporting prevention activities, with special emphasis during the faire campaign, allowing the user to know the identification and evaluation of the risk area [GMES website, Emergency Management page].

Landslides

Landslide services are the following [SAFER website, Landslides page]:

- The Landslide Inventory Mapping (LIM) service reports on the spatial distribution of past landslides: it covers large areas, with an order of few thousand square kilometres;
- Landslide Monitoring (LM) service reports on the past ground movements of single large landslides affecting built-up areas with a high risk level;
- Rapid Landslide Mapping (RLSM) service offers rapid mapping capacities when a landslide has occurred;
- Real-Time Shallow Landslide Forecasting (RSLF) service provides assessment maps consisting of a near real-time prediction of the occurrence of shallow landslides at large scale.

Earthquakes and Volcanoes

The earthquake monitoring and damage assessment service provides soil displacement maps for the prevention phase. During crisis phases, it also monitors co-seismic displacements and performs damage detection in urban areas.

The volcanoes-related service addresses risk management during the whole management crisis cycle: pre, during and post crisis [GMES website, Emergency Management page].

Humanitarian Crisis

The service provides information capabilities to support early warning, urgent assistance, relief operations, humanitarian aid and reconstruction activities.

In particular, it provides a rapid mapping capacity when crisis occur and supports preparatory activities through the collection and mapping of spatial and socio-economic data [GMES website, Emergency Management page].

I.4 Projects Launch

Since the launch of the initial concept of GMES back in 1998, substantial investments for Research and Development have been made in Earth observation by the European Union, the European Space Agency and their respective Member States.

The "Forum GMES 2008" held in Lille, France, in September 2008 marked the launch of the first GMES services in pre-operational mode. The implementation of these first services is ensured through R&D projects funded under the 7th Framework Programme of the European Union:

- geoland2, pre-operational service for Land Monitoring;
- MyOcean, pre-operational service for Marine Environment Monitoring;
- MACC, pre-operational service for Atmosphere Monitoring;
- SAFER, pre-operational service for Emergency Management;
- G-MOSAIC, pre-operational service for Security.

Several other projects related to the GMES initiative are currently funded under the 7th Framework Programme and over the last decade, numerous R&D projects funded either by the European Commission through its successive Framework Programmes or by the European Space Agency through its "GMES Services Element" (GSE) programme have contributed to the implementation of GMES. The preparatory actions have to be mentioned in this projects category too; one of these will be presented in chapter 2.

I.5 7th Framework Programme

Conducting European research policies and implementing European research programmes is an obligation under the Amsterdam Treaty; moreover, since even entire Member States find it increasingly difficult to play a leading role in many important areas of scientific and technological advance, international cooperation is needed. Research and innovation are much profitable for the entire community on the long term: it was estimated in 2006 that the contribution of € 7 billion/year might generate a GDP increase of € 200 billion/year in the 2030s. Framework Programmes go in this direction promoting research and development in an environment of international cooperation.

Framework programme projects are generally funded through instruments. They are composed of multiple partners coming from different countries but can join several tens of partners; the typical duration of the projects is several years, usually from 3 to 5. These projects aim at fostering European competitiveness in research and applied science [Wikipedia, Framework Programmes for Research and Technological Development].

Proponents gave the intangible incentive to face the intrinsic complexity of international collaborations; diversity introduces additional costs, but it facilitates addressing competitors in an even more diverse world. Changes triggered by research policy directly affect people and enterprises, which experience broader horizons and experience the advantages of international collaboration.

GMES initiative is funded under the 7th Framework Programme; in particular an advisory group at international level was set up to advise the European Commission on the overall strategy to be followed in carrying out the Information and Communication Technology thematic priority. It reflects and advises on the definition and implementation of a coherent policy for research in and on ICT in Europe; the GMES project is part of this policy.

1.6 SAFER

SAFER is the project that currently provides the GMES Emergency Response Service in a pre-operational phase. It is funded by the European Union.

An example of the products provided by SAFER follows. The reader can get aware and appreciate the role of the geospatial information and of the reference spatial data in providing the service. Geospatial information in emergency management is the framework in which the activities of the thesis were performed.

1.6.1 Cambodia floods example

This example shows the results of the activation of SAFER after a real flood occurred with focus on Cambodia: they consisted in a short description of the event, a map, reported in Figure 1.1, that shows the affected areas and a table, reported as Table 1.1 in this document, that resumes the most important information. In particular the map is very useful in the first hours after the event, because it could help in saving lives and sending aids to the population hit by the flood, in performing a post disaster needs assessment and in taking decisions related to the recovery of the disaster.

| Cambodia Floods SAFER Activation | |
|---|------------------|
| Type | Flood |
| Location | Cambodia |
| Date of event | October 8, 2011 |
| SAFER triggered by | WFP |
| SAFER triggered on | October 11, 2011 |

Table 1.1 Information details on the SAFER activation on the Cambodia floods (source SAFER website)

Description. Strong typhoons and heavy rains caused severe floods in the South East Asia. Cambodia experienced the worst flooding in over a decade with an over flood of the Mekong and Tonle Sap rivers affecting 17 of 24 provinces. The flood water in many provinces has been raised both from heavy rain and from the impacts of flooding from neighbouring countries. Thailand and Vietnam also experienced heavy flooding. There are currently more than one million people affected, of which 172 death, and at least 100,000 people displaced, but numbers are expected to

increase. More than 10 percent of agricultural land has been inundated and the rice crop will most likely be lost [SAFER website, Cambodia Floods].

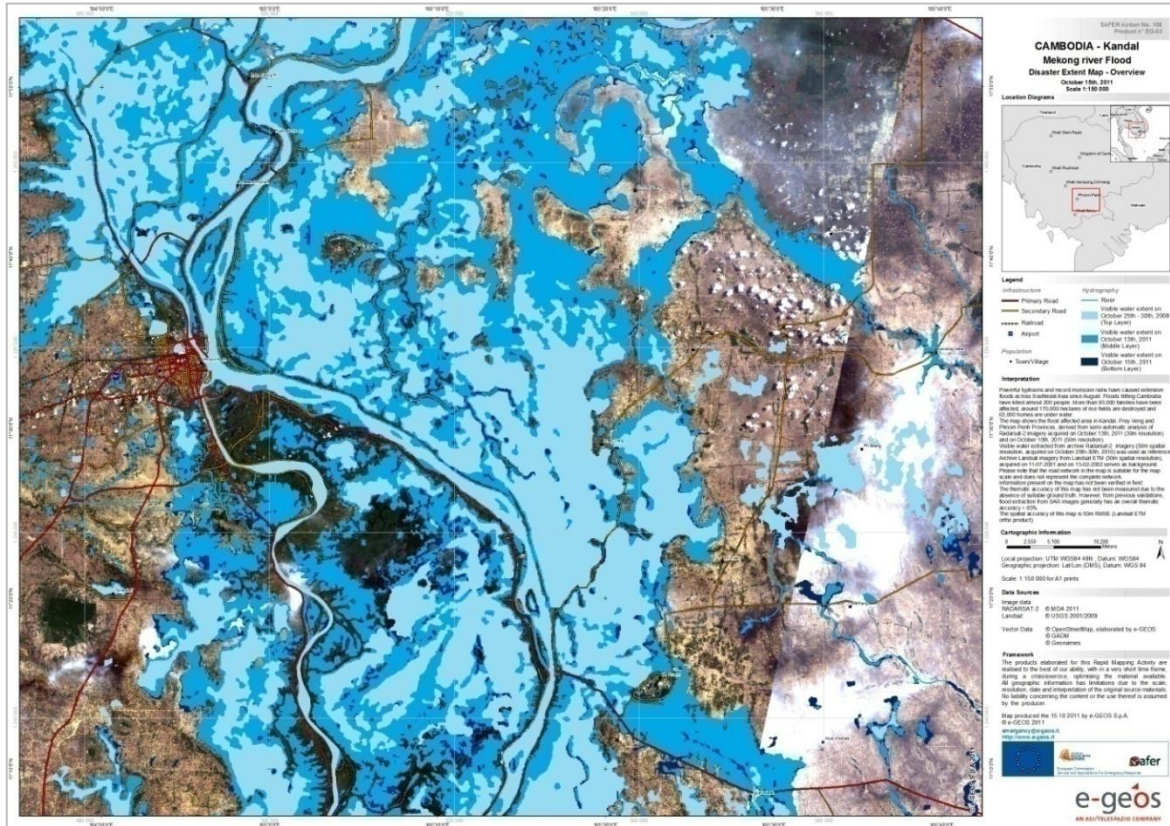


Figure 1.1 Map produced by the SAFER project after an activation with focus on floods in Cambodia. The map shows the flooded areas extracted from the classification of satellite imagery in two different days after the events, in comparison with the normal conditions of a pre-event satellite image; the base reference data is constituted by a satellite image in true colours and vector sources of transportation network and settlements (source SAFER website).

1.7 Final notes

In this chapter the EU GMES initiative has been introduced, the services it comprehends have been listed and the services accessibility have been explained. The project funding form was also introduced: GMES main services are considered “public goods” for the citizens’ welfare and security. The pre-operational services are already active and in a few months some of them are going to be replaced by the stable and operational versions.

The focus of the last part of the chapter was set to the emergency management topic and an output example of the respective pre-operational service SAFER was reported. From the example it was easy to remark the relevance of the geomatics, that is the conjunction of the information technology, such as computer processing and internet communication, with the geospatial data. Data can have different role such as base reference or theme of the map message; nowadays all data are digital already from their acquisition during measures or surveys, or are converted into a

digital format during the following steps of processing and analysis. In the next sections the role of the geomatics in one of the GMES preparatory actions is disclosed.

2. Initial GMES service for geospatial reference data access

In chapter 1 the Global Monitoring for Environment and Security initiative was resumed and its core projects were presented. It was also mentioned that beyond the core projects a series of preparatory actions and activities are promoted by the European Commission and financed in various forms, as a base or a support to them.

Most of the GMES core projects are strictly related to the geospatial data availability, in particular the ones that involve geospatial analysis, geoprocessing and cartographic representation of the results. In this sense geospatial data are needed in both the methodology application and the presentation of the activity results. For instance the SAFER project (see section 1.6) takes advantage of geospatial data both during the definition of the area hit by a natural disaster, through the comparison between the data collected in the first hours after the event occurs and base reference data, and during the presentation of the results in the form of a map.

Base reference data can be defined as collections of geospatial data that contain information useful as spatial terms of reference; for example elevation data, river network and road network can be imagined as fundamental elements. Anyway the requirements in terms of content vary in function of the activities that are going to be performed with the data and of the user who actually performs them. Therefore the data collection must come after a deep analysis of the user needs and of a technical assessment of the activities in which the data are going to play a fundamental role.

The data repository requires a physical space where data have to be collected: therefore a strategy for data storage must be prepared before the beginning of the collection. The data collection is performed with the aim of obtaining a data repository from which they can be retrieved easily on demand: therefore a strategy for data providing has to be prepared, which involves a smart data archiving and data cataloguing model and eventually an easy access. Moreover base reference data are used only if they provide a representation of the reality sufficiently precise and reliable, that is if they have a good quality: therefore, during data collection, activities aimed at defining minimum requirements in terms of data quality and then at verifying whether the collected data respect the given requirements have to be performed.

In data collection, as in most of the other activities, there is a gap between the expectations and the real implementation: the smaller is this gap the more successful is the activity. Therefore while performing the activities, sessions of reference data gaps analysis have to be performed and gap reduction measures have to be applied whenever possible. Base reference data collection comprehends challenging tasks that require significant efforts; the results are expected to be used by the widest community of users possible. For reaching this goal preventive activities have to be carried out in order to find similar ongoing projects; in positive cases collaboration is an option to be considered.

Technology is another important element that has to be mentioned; the disclosure of the digital world and of the network broadened the potential choices when it comes to geospatial data collection and exploitation. Available and suitable data formats, data storage techniques, data distribution protocols are numerous and have to be analysed carefully in order to take the right choices in relation to the needs.

The European Commission promoted the realization of a geospatial reference data access as a preparatory action of the GMES initiative: it implies data gathering activities, the definition of a protocol for data quality checking, the transformation of data into formats suitable for storage and then their distribution. All these steps must follow the specifications collected during the investigation and assessment of the end user needs.

The financial support is given by the publication of a public call for tenders dated back to 2009. In particular the call divides the activities into two lots on a spatial criterion: the first lot refers to data regarding the inside of the European borders while the second lot has a global view. In this chapter the call for tenders is presented and the content of the second lot is analysed; the tender contractor is introduced and the technical agreement between the parts is described as a solution of a correct and innovative way of implementing of a geospatial reference data access to be exploited also for emergency management purposes.

2.1 Tender Specifications and Monitoring

2.1.1 Objective

The European Commission defines reference data as information handled by data providers, both institutional and commercial, and recognised by a large community of users as basic spatial data to reference their thematic data upon, for example road networks, rivers and water bodies, administrative units and toponyms. Reference data fulfils three functional requirements: provide an unambiguous location for a user's information; enable the merging of data from various sources; provide a geographic context to allow others to better understand the spatial information that is being presented.

The EC promotes a positive management of geospatial data and points out good examples regarding the Northern European Member States, where efforts have been made to ensure that throughout the various levels of public authorities, the same spatial data are being used as reference data. This approach allows to avoid duplication of data collection, to facilitate any combination of thematic spatial data and eventually to ensure continuous updating of these

reference data by taking onboard the necessary changes as produced by authorities responsible for specific spatial object categories, forming subsets of the overall topographic set of reference data.

Reference data are considered vital to GMES services, in particular Emergency Response, Land and Security services, and to end-user applications in order to provide the basic geographic framework on top of which additional thematic spatial information can be produced and disclosed. Examples of this additional information are land use/land cover maps, asset maps and damage assessment maps in response to crisis and the set of relationships between the geographical components that will allow building the assessments, analyses and monitoring from combinations of data sets.

The EU introduces the GMES preparatory action with the main objective of supporting the uptake of GMES services. As such the activities promoted by the mentioned call for tenders fits into a step up of the R&D results and GMES precursor activities towards fully operational services. Therefore it has to take into account previously obtained results as well as ongoing developments in the field [EC Enterprise and Industry Directorate-General, 2009].

The EC defined also the main goals of the preparatory actions for Land Monitoring and Emergency Core services, that are:

- for Land Monitoring Core services the preparation includes evaluation of fitness for purpose of existing datasets and procurement, harmonisation and processing in view of making these data themes available;
- for Land Monitoring Core services the proposal of options for complementing the reference data sets with new data when there are gaps to be filled and for designing a future sustainable procurement strategy for spatial data as input to the GMES core services. Data policies and license conditions constitutes a key issue;
- for Emergency Core service the development of a data access strategy, including a data policy and actual access to remotely residing and existing reference data repositories held by various data providers at international, continental or national level. Prevailing data policies and license conditions which may apply are key issues also in this case.

The objective of the mentioned tender is to address the progressive operational implementation of GMES services and more specifically to start the implementation of an initial GMES service for a geospatial reference data service element cross-cutting the different GMES services, Land, Emergency and Security in particular.

Given the fact that the context and availability of data as well as the data sources are very different inside and outside the EU, their implementation have been considered in two separate lots of the tender:

- the first lot is dedicated to the implementation of initial GMES service for geospatial data access in Europe. It encompasses spatial data improvement, middleware services development and installation on the necessary hardware components on which the services will finally reside.
- the second lot is dedicated to the implementation of an initial GMES service for geospatial reference data access covering areas outside Europe. It will focus on the analysis of non-EU reference data availability, quality and consistency and gaps to be filled, as well as the status of existing initiatives, in particular UN-SDI to establish links

and possible complementarities. The development for this part is limited to demonstrators on real user cases.

2.1.2 General considerations

A reference data access service, in its adaptation towards the GMES initiative, is conceived by the European Commission representative as following some general and fundamental principles to which a broad validity can be given. These principles are recalled in this paragraph [EC Enterprise and Industry Directorate-General, 2009].

Lowest feasible access threshold

A public access to geospatial data, that originates from the citizen and has the citizens' security as a final goal, must have the widest possible accessibility. The focus of the project is on the uptake of GMES services by aiming at an as low as feasible access threshold to spatial data through middleware services, whilst paving the way for sustainable mid and long-term availability of some data themes. Thus a data service developed following this principle has to be able to reflect changes in the applicable schemes for provision of data in a flexible and continuous manner.

Reference data harmonisation and integration

Data collected under the activity of building a reference data access come from several data holders, are distributed by different data providers from various countries; due to this heterogeneity data types will be various and different among datasets, as well as the forms of data storage and database schemas. Therefore the necessity of data harmonization and integration among datasets with different sources is recognized as one of the core activities of such an initiative.

Integration with similar projects

An initiative with global extent requires strong efforts from all the agencies, companies and governments involved: sharing efforts is generally a wise approach in order to reach better results and potentially save resources. Therefore the investigation of possible similar projects has to be performed as well as an attempt to integrate them under a common goal.

At a European level the overall concept draws on various spatial data and services either already made available by the member states at a national or regional level in the framework of their compliance to the INSPIRE directive; alternatively and complementary some data which are to be integrated, complemented or improved to spatial data products, should be managed in an INSPIRE compliant way in the near future.

Other two cases have to be mentioned: on one side as part of the action on the EU territory, data of interest generated by other GMES services, such as land cover information or others, should be hold an processed as part of the reference data service; on the other side data bases comprising information on non-European areas, should be treated like additional data bases from any country.

Sustainability

The provision of a reference data access service, due to its condition of initial stage, has to be considered into the overall long-lasting of the GMES service portfolio. Indeed the service to be developed is not going to be discontinued at the end of the validity of the contract, but is going to be transferred into a long-term sustainable configuration: therefore an initial iteration of the sustainability concept should be conceived.

On the practical level this concept requires in general the portability of the S/W and database solution from the initial configuration towards the final target environment. Further, this way involves the procurement of dedicated hardware platforms, as well as system software fitting into the operational environment of the hosting entity. Finally, a complete set of transfer, installation and acceptance testing documentation has to be established.

2.2 Technical Proposal and Stakeholders

An Italian consortium, composed by a research institution and a no-profit association called respectively SiTI and ITHACA, won the call for tenders described in section 2.1: from now on the terms consortium or contractor will refer to the two. The project has duration equal to three years, starting from January 2010, in which the goal of building a reference data access must be completed. The list of activities to be performed along the duration of the contract was defined on the basis of the tender specification and the technical proposal of the consortium. These activities as a stack can be considered the correct approach to be followed when an access to geospatial reference data has to be built and implemented.

In the following paragraphs the components of the consortium are introduced and the activities agreed between the European Commission and the consortium are described. Finally the selected users of the reference data access and their needs are discussed.

2.2.1 Contractor

SiTI

SiTI – Istituto Superiore sui Sistemi Territoriali per l’Innovazione (Higher Institute on Territorial Systems for Innovation) is a non-profit association set up in 2002 between the Politecnico di Torino and the Compagnia di San Paolo, to carry out research and training oriented towards innovation and socio-economic growth.

Its activities are concentrated in the following sectors: Logistics and transport, Environmental heritage and urban redevelopment, Environmental protection. Moreover it engages the help of numerous researchers who complete the various specialist skills available within the Politecnico di Torino, the Universities and other research centres.

The Institute develops its researches mainly through participation in European bids for tender and the funding of projects by internationally acknowledged Entities and subjects. SiTI carries out useful monitoring and evaluation activity in various sectors in which the Compagnia di San Paolo is involved, with particular commitment to the “Social Housing Programme” [SiTI website].

ITHACA

ITHACA is the acronym for Information Technology for Humanitarian Assistance, Cooperation and Action; it is a non-profit association, founded in November 2006 by the Politecnico di Torino and SiTI.

The different competences characterizing ITHACA are related to the acquisition, management and elaboration of geographic and cartographic data. The operational team has deep experience in several of the domains concerning geomatics, such as remote sensing, photogrammetry and Geographic Information Systems (GIS). Therefore, ITHACA is able to produce basic and thematic cartography, through both consolidated and innovative procedures.

With a view to cooperate with the World Food Programme, the food aid arm of the United Nations and the world's largest operational humanitarian agency, the Association proposes itself as a centre of applied research and for the distribution of products and services related to Information Technology in support of humanitarian operations. It aims to serve as a bridge between advanced research science and humanitarian practitioners, delivering scientific methodologies, analytical services and technical tools to improve the capacity of the WFP and the broader humanitarian community in Early Warning and Early Impact phases.

ITHACA's activities are realized with the cooperation of Politecnico di Torino, specifically with the Land, Environment and Geo-Engineering Department and of other technology partners such as public and private research centres, companies operating in the aerospace and telecommunications fields and other suppliers of territorial data. ITHACA is governed by a Board of Directors and assisted by a Technical Committee [ITHACA website].

2.2.2 Fundamental activities

The fundamental activities to be performed for building a reference data access for emergency purposes are the following;

- Data inventory and modelling;
- Data quality assessment;
- Analysis of similar ongoing projects;
- Gaps identification and narrowing;
- Development of demonstrators.

The activities can be defined successful if they satisfy the end users needs: in order to better reach the goal, a preliminary identification of some of the users and an investigation on their strongest needs has to be performed. Therefore the process in its whole comes after and must follow the user needs assessment.

Data inventory and modelling

This stage consists in an investigation and searching of available geospatial data with a global extent, by mean of bibliographic review, technical documentation and also web search engines;

preliminary suggestions on suitable data sources can be found in the tender specification as supplemental information to the call.

A catalogue of the data found during the previous activities have to be created and updated continuously; the catalogue have to be filled with their metadata and exposed to the access of the users, preferably through the internet [EC Enterprise and Industry Directorate-General, 2009].

This stage comprehends also the implementation of a logical relational database model, using the UML modelling language, that has then to be integrated with specific geographic behaviour that cannot be expressed by mean of a UML design. The implementation of the model is conceived as an iterative and multi-versioned process that foresees amendments on previous versions and integration of new data according to the users' feedbacks: it consists in the physical implementation of a version on a development environment, sample dataset loading and usage testing by the final users [ITHACA 2010].

In this stage public domain datasets have a priority rank rather than commercial data; the use of commercial datasets is optionally postponed for gaps filling and quality increase purposes. This choice allows to get an operative product in shorter terms by avoiding the management of time consuming licensing issues.

Data quality assessment

The concept of data quality draws on the relationship between the reality and its representation contained in data itself. The closer to the reality data collected during the inventory phase are, the more useful they are; if data doesn't pass a threshold of closeness to the reality they cannot be considered useful reference data. Naturally similar datasets in terms of content can have much different gaps in comparison to the reality: other conditions being equal, data user is going to choose the one that better model the real conditions for his purpose.

Data quality has to be defined by identifying a set of quality indicators: then they have to be applied to the data in order to get a quality assessment.

Thus this stage comprehends: firstly the definition of data quality in terms of the identification of a set of indicators; secondly the definition of a protocol for applying this indicator to the dataset selected during the inventory phase; finally the publication of the results of the protocol application to the real datasets, that is a data quality statement.

Analysis of similar ongoing projects

The most important initiative regarding the collection and the distribution of geospatial data with a global coverage is the UN-SDI. The United Nations have a long tradition in handling geospatial data; many agencies have GIS capacity in their head quarters or in their country offices as well as specific field projects. But there are only few examples of spatial tools to provide analysis to the decision makers.

Although the UN agencies are many and their needs are much diversified, their needs in terms of geospatial data can be categorized in [ITHACA 2009]:

- agency related data such as assets, programmes and personnel that should be progressively georeferenced and shared to support the UN reform and to increase coordination;

- geographic features in developing countries, aimed at supporting emergency operations, should be updated, corrected where possible and georeferenced;
- thematic data according to the mandate of the agency, such as land cover, food security, poverty, climate change.

The UN made two moves recently in order to satisfy its internal needs: in the year 2000 a consultative forum that brings together the focal points of many UN entities, called UNGIWG, was established with the aim of providing advices and promoting to exchange geo-referenced data amongst the UN agencies. Then in the year 2005 this group endorsed the creation of a spatial data infrastructure for the UN agencies in order to support coordinated efforts in the development and management of geospatial information.

The tender contractor is represented since 2007 in the initiative as development partner with precise responsibilities and can provide an analysis of the status of it as well as field for collaboration with the GMES projects. GMES projects may become a data aggregator and a dissemination hub for the international community, receiving data related to small areas from the UN agencies and integrating them into sharable global layers; also GMES may support the creation of a UN federative network by sharing with its agencies data management services and infrastructures [ITHACA 2009].

Gaps identification and narrowing

At some points during the activities of building a reference data access service, a comparison between the current results and the needs highlighted by the users has to be performed in order to identify information gaps; accordingly technical solutions have to be proposed for gap narrowing and filling.

The availability, accessibility and possibility of integration of better quality data should be investigated as well as the opportunity of acquiring information from other sources, such as field campaigns, classifications and photo-interpretation of remote sensing data or setting up volunteered geographic information systems like Google Map Maker¹ and Open Street Map².

Development of demonstrators

The call for tender introduced in this chapter proposes the development of an initial data access service; in the first phase the access has to be provided only as a proof of concept. Therefore more importance is given on the software side of the development rather than on the hardware side and on the performance of the data access itself.

Demonstrators of the reference data access functionalities have to be developed on real user cases, by firstly enabling and brokering the access to individual non-European data. This real use demonstration study has to be focus on some non-EU areas of interest, in particular Africa, for the emergency and global land monitoring services.

¹ <http://www.google.com/mapmaker>

² <http://www.openstreetmap.org>

In this stage the technology in geospatial data management and in data access and communication in general plays a key role. Following the more recent trends in terms of information technology the focus should be on Open Source software and on Web 2.0.

2.3 Identification of users and needs assessment

The needs assessment is the first necessary step for the successful implementation of a reference data access service: it requires a systematic involvement of the possible end user and a thorough analysis of their requirements in terms of geospatial data [ITHACA 2010].

For what concerns the GMES initial reference data access service, after a preliminary analysis and various requests of availability, some users were identified and their representatives were interviewed in order to identify their needs.

The World Food Programme agency was considered the first user of the data access; it is the food aid branch of the United Nations and the largest humanitarian organization in the world addressing hunger worldwide. Its headquarters are in Rome, while more than 80 country offices are located around the world [WFP website].

Three WFP units are interested to the project:

- the Emergency Preparedness and Response unit offers rapid response to natural disasters in order to mitigate the impact on hit population. In doing so it makes use and draws cartographic products at a small scale in support to rescue operations; it is also involved in the UN WFP SDI;
- the Vulnerability Analysis and Mapping Unit monitors vulnerable areas before, during and after a catastrophic event or a crisis, with particular emphasis on the food security topic. It produces various reports and map products, often based on satellite derived data and indicators;
- the UN Joint Logistic Centre supplies the logistic support needed for the WFP operations during emergencies. Local offices are often equipped for acquiring data when required, mainly through GPS technology.

The main outcomes of the investigation of WFP needs are [ITHACA 2010]:

- Data reliability and integrity;
- Adoption of standards regarding data formats and projections, up-to-date sources, workflow and metadata;
- Identification of common base layers with a worldwide coverage, possibly updated on a regular basis and with large map reference scale;
- Development of WebGIS applications, including routing capabilities;
- Adoption of UML data modelling;
- Considering open source platforms for GIS data management;
- Possibility to update locally the data repository in case of emergency activation;

- Large community of potential users;
- Remote hosting of the server where the geodatabase is maintained can be a solution only if security over sensible datasets, 24/7 operability, periodical backup and fast recovery are granted.

The Global Facility for Disaster Recovery and Reduction Labs (GFDRR Labs) are another important user; founded in July 2010, the mission of the Labs is to use science, technology, and innovation to better empower decision-makers in the developing world to increase their resilience to disasters in a changing climate.

The Labs strategy is to build partnerships and communities that utilize open data and open source technology to assist decision-making. Labs activities are designed and implemented in partnership with the World Bank's regional Disaster Risk Management teams, which insures that these initiatives are developed to add value in operational environments [GFDRR website, Labs page].

An innovative initiative of the Labs is Open Data for Resilience (OpenDRI): in order to build resilient societies, policy-makers and the public must have access to the right data and information to inform good decisions.. The initiative aims to reduce the impact of disasters by empowering decisions-makers with better information and the tools to support their decisions [GFDRR website, OpenDRI page]. More information on this initiative is contained in section 6.3.

The main input received from GFDRR Labs is related to the activities of *Post Disaster Needs Assessment* in which data preparedness is considered a key role: mapping and web-mapping applications are specifically identified as operational tools. Therefore GFDRR needs concern the functionalities of the demonstrators are:

- the possibility to generate maps using a pre-defined map template, optimizing the map production process;
- the possibility to download not only maps in raster or *.pdf* formats but also the cartographic layers adopted to generate them;
- the possibility to upload geospatial information as a group rather than as individuals in order to increase the level of trust in the products.

Furthermore other interested users are the GMES related projects mentioned in section 1.4 as well as the Lot 1 of the reference data access introduced in this chapter.

Finally it is worth mentioning the GMES in-situ coordination, a FP7 funded project aimed at developing an innovative and sustainable framework for open access to in-situ data. One of its work packages consists in documenting needs and requirements specifically for in-situ data and the results are of interest due to the common tasks.

3. Project activities status

In chapter 2 the list of task necessary in order to set up a reference data access for emergency management purposes was introduced and details were given on a call for tenders published by the European Commission in the framework of the GMES initiative.

In the current chapter the status of the project activities is investigated in details, by keeping the last interim report produced by the contractor in September 2011, the so called deliverable M21, as reference.

3.1 Data inventory and related activities

The contractor organized the activities related to the data inventory phase into four groups that are focused on the collection of geospatial data and on the criteria to be respected in terms of data storage and distribution in order to match the tender requests.

3.1.1 Data inventory

The data inventory activity has been developed so far in two separated steps, in order to match the needs of the identified users, which can be here summarized in making use of a limited number of common geospatial data and several specific ones. Accordingly, useful data have been separated into two groups:

- Core data, that are the minimum geographic data baseline that constitutes a common essential background in order to allow the various application and activities to work properly;
- Thematic data, that are additional data, identified specifically for certain applications.

The comprehensive inventory report [ITHACA 2009], prepared by a UNGWIG technical group in conjunction with the FAO *Poverty Mapping Project Group*, was considered as a reference document for the identification of the core data in the first steps of the project [ITHACA D10 M21, 2011]. This report was produced in 2005 and contains an inventory of global spatial data

sources and the definition of a standard list of core global databases for use across all UN agencies in the field of emergency preparedness and response, food security and poverty mapping issues. The proposed list has been developed and modified in order to match the EC requests.

Afterwards core data and thematic data have been classified into topics and sub-topics as a choice of methodology and internal organisation as well as a push towards clearness for the end user: this classification took also into account proposal of spatial data themes needed for environmental applications addressed by the INSPIRE directives [INSPIRE website, Data Specifications page]. Topics and sub-topics can also be intended as data categories and sub-categories; specifications on the classification are contained in Table 3.1 and Table 3.2.

| TOPIC | SUB – TOPIC |
|---|--|
| Boundaries: costal, administrative and areas of special interest | Coastlines and Maritime Boundaries Country Political Level 0 Boundaries and Area of Dispute Boundaries Sub-national Boundaries (from 1st 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 Surface Water Bodies (SWB) Watersheds and River Basins Water Points |
| Human population: population centres and distributions | Population Centres Urban and Rural Population Density and Distribution |
| Transportation: roads, railways, airports, harbours and nautical routes | Roads Railway lines, Stations, and Marshalling Yards Airport Harbours Nautical Routes |
| Bathymetry and terrestrial elevation | Bathymetry Terrestrial Elevation |
| Points of Interest | |

Table 3.1 Topics and sub-topics definition for the Core datasets (courtesy of ITHACA)

After the initial activity of content classification and definition, the consortium performed a review of available geographic data sources, that has been planned to be a recursive activity in relation to the feedbacks of the EC representatives. The review started from an initial data list provided together with the tender specification and from queries performed in the metadata catalogues of the major organizations that deal with geographic data, such as UNEP GRID and FAO. Further review has been performed with Web search engines and using specific technical documentation.

The results of the review were further filtered with some restrictions; indeed particular attention was given to datasets with global geographic extent and relatively small scale as well as with reduced or absent constraints in access and use for the services. As a result global datasets free of

charge for access and use in academic, research and other non commercial uses constitutes most of the review.

| TOPIC | SUB – TOPIC |
|---|---|
| Geology | Geology Geo-morphology Hydro-geology Soils and Soil Properties |
| Satellite imagery, land cover/use and vegetation | Satellite Imagery Land Cover/Use and Forests Vegetation |
| Meteo-climatic | Rainfall Evapotranspiration Temperature |
| Socio-economic | Demography Economy Society Poverty Environment |
| Natural disasters | Earthquake data Volcanic eruptions data Flood data Fires data |

Table 3.2 Topics and sub-topics definition for the Thematic datasets (courtesy of ITHACA)

The review activities provided an inventory of the available and globally consistent data sources, grouped according to the predefined topics and sub-topics and organized as tables with useful information for the following steps of the project, such as: the identification of the institution, organization or company responsible for the production and maintenance of the dataset, the geometry data type, the reference scale, the date of the last edition as well as the procedures needed to obtain the datasets and potential access constraints.

3.1.2 Data modelling

The contractor is implementing a logical relational database model on the basis of the assumptions made on reference data subdivision above described, using UML as modelling language. UML model will be then integrated with specific geographic behaviour such as topology, network and raster catalogues that cannot be universally expressed in a UML design.

The definition of the model is an iterative process, involving several sequential steps:

- physical implementation of a draft version on a development environment;
- sample dataset loading;
- usage testing by the final users, in particular WFP units;
- feedback collection and model amendment.

The process will end with the definition of the first version of the data model; newer versions of the model itself will be developed and will include amendments on previous versions and integration of new reference and thematic data, according to emerging user needs [ITHACA D10 M21, 2011].

The contractor is also performing a fundamental activity of reference and thematic data harmonization following a three-step approach in parallel and in conformity with current data model version: evaluation of input dataset characteristics, evaluation of the effort requested for their transformation and development of harmonization procedure [ITHACA 2009].

3.1.3 Metadata and OGC compliant on-line catalogue

The consortium created metadata for all the record found during the review of the data sources, according to the current standards concerning geographic information. The INSPIRE Implementing Rules for metadata were followed, in particular the Dublin Core, that collects together the major existing international standards in an integrated way: EN ISO 19115, EN ISO 19119 and ISO 15836 [INSPIRE 2007]. These standards define an extensive set of mandatory and conditional metadata elements, that constitutes a basic information on datasets to be maintained for catalogue purposes for example data discovery, data access, data transfer and use.

Afterwards the consortium collected all the metadata into a catalogue implemented following the standards defined by the Open Geospatial Consortium and available for the access from the Web. The software GeoNetwork Opensource³ has been adopted as technical solution for the catalogue on-line, although other similar products are available, such as ESRI GeoPortal. At the level of the United Nations, GeoNetwork Opensource was considered as a standard *de facto* for metadata sharing and archiving and it is cited as an example of interoperability because of the adoption of international and open standards for services and protocols [ITHACA D10 M21, 2011].

The software application allows a user to search for available geospatial data by mean of a search engine that works both with keywords and with spatial extent, such as country or region names. Metadata can be exploited on-line or downloaded; links for geospatial data downloading are integrated. The latest versions of the software permit also to preview the data in a map viewer as a further chance of getting information.

The GeoNetwork node set up for the purpose of the GMES RDA Lot 2 project is available at the following URL:

<http://stage.ithacaweb.org/geonetwork>

3.1.4 Demonstrator applications

The contractor defined mandatory an application able to show the information and services collection from multiple servers; in this way the end user can truly take advantage of the benefits of the OWS services, making simple assessments and a first approach to useful analysis.

A web application was adopted as a basis for the demonstrators technology, aimed at accessing the selected reference data outside EU and at allowing their use in conjunction with geospatial

³ <http://geonetwork-opensource.org/>

data and services from other parties. The application is the GeoExplorer, a map builder completely open source that works as a Web tool. The user interface is a simplification of the one adopted by the most common desktop GIS application in order to be easy and ready to use by the widest possible community of users. As a map builder GeoExplorer has a big canvas for showing maps and a table of contents with function of layer managing. Layer sources can be local or external, indeed the user can reach other sources by typing the URL of the respective server and select from the available layers there provided. Further technical details are provided in section 6.4.

3.2 Analysis of data quality and consistency

The consortium defined a list of requirements on the basis of the results of the user needs assessment and on the basis of a bibliographic review that takes into consideration international standards as well as practical concepts. These requirements allowed the definition of several quality indicators.

The contractor then defined a protocol to be followed for assessing the geospatial data quality level; the protocol constitutes a quantitative approach to the data quality assessment. In fact each indicator, when applied to the datasets, can assume values that correspond to a score in the range from 1 to 5. Not all the indicators are worth the same, so a weight is assigned to each indicator on the basis of its importance towards the others. The total score is then the sum of the partial scores, weighted according to the respective indicator.

The protocol is general in order to be applied to a wide variety of geospatial data and is made to establish a general approach to the quality assessment in order to perform assessments as objective as possible.

All the datasets resulted to be part of the previous data inventory phase were submitted to the data quality assessment: the protocol states that it has to be performed at a subtopic level. Therefore the datasets classified under a specific subtopic were assessed separately from the others. The dataset with the highest score among the group was then suggested to be part of the final data model.

During the assessment activities, two different control methodology were applied; on one side systematic controls were applied, that have been executed over the whole set of available records, on the other side sample controls were performed over randomly defined and distributed portion of the datasets. Details on the data quality protocol definition and on the quality assessment are reported respectively in chapter 4 and in chapter 5.

3.3 Analysis of the UNSDI initiative

Before and during the GMES project activities, the contractor took part of the UN-SDI initiative as a development partner with precise responsibilities stated into shared official documents. Furthermore duty of the GMES contractor was to join the UN-SDI initiative and provide an analysis in order to catch the strongest points of contact between the two initiatives and potentially establish collaborations and integration [ITHACA 2009].

The UN-SDI is an initiative of the UN Geographic Working Group, an assembly of 32 UN Secretariat Departments, Programmes and Specialized Agencies in which FAO, WFP, OCHA, UNDP, UNICEF, World Bank are among others. It was established in the year 2000 with the aim to address common geospatial issues such as maps, boundaries, data exchange, standards, naming conventions and location. The members consist of focal points and professionals working in the field of cartography and geographic information science within the UN system. Participation is only on a voluntary basis and UNGIWG does not control any budget, therefore cannot provide funds to its members [ITHACA M21 D12, 2011].

It is relevant that non members who have a specific expertise related to the topics being deliberated by the Group may also participate to the meetings by official invitation. The Group works on three distinct processes: annual plenary meetings, time bound ad-hoc Task Groups and longer term Special Interest Groups. The concept of establishing a UN Spatial Data Infrastructure was endorsed in 2005 because it would allow a stronger contribute of the geographic information to the mission of the United Nations. In the UNGIWG vision, the UN-SDI is a comprehensive, decentralized geospatial information network that facilitates efficient global and local access, exchange and utilization of geospatial information; the main goal is enhancing the decision making from the lowest to the highest level in the UN.

In 2009 the process of implementation was defined in its final form and it states the creation of a Centre of Excellence to lead the UNSDI generation and development. This choice has the advantage of pushing forward the adoption of new technology, being this an easier task rather than changing the workflow of big agencies. At first an indicative five-years Implementation Plan for the UNSDI was defined as a collection of four overlapping stages that are described below.

Finally it must be highlighted that since the UNSDI follows an evolutionary path, influenced by the availability of resources such as skilled people, technology and infrastructure, deadlines set for completion of some tasks remain tentative [ITHACA M21 D12, 2011].

Stage 1: Building UNSDI foundations

The aim of the first stage is to establish an effective UNSDI organizational architecture and formalize agreed policies and standards. Priorities include the facilitating of access to established core datasets, metadata development and distribution, initiation of critical new databases and development of visualization tools.

Stage 2: Building the infrastructure

The second stage of the UNSDI implementation focuses on further development of core data layers, including the datasets already identified by UNGIWG Task Groups for development but not yet implemented for lack of available resources. Metadata population and the development of catalogue services remain priorities throughout Stage, aided by the adoption of relevant policies, standards and tools by both internal and external partners. Numerous of these projects can be expected to continue well into Stage 3 of the UNSDI planning period as well.

The design, development, and testing of site interoperability, data access and maintenance procedures are important at this stage of infrastructure development. Nominated data custodians for example, should already be maintaining and updating data and publishing metadata associated

with the data archives for which they have accepted responsibility. Remedial action needs to be initiated during Stage 2 if any concerns are highlighted during the testing of the infrastructure.

Stage 3: Institutionalizing the infrastructure

Stage 3 sees the institutionalizing of the information infrastructure and the dawning of operational geospatial data services across the distributed matrix of data resources that comprise the UNSDI. By the time Stage 3 implementation begins, UNGIWG agencies and partners have significantly increased the number and capability of geospatial data portals and other data resources linked within the UNSDI. Improved access to data and distributed processing resources that leverage value from new data sharing agreements should result, but this will come at price. As the number of documented and interoperable geospatial data layers increases and access to the information infrastructure becomes more widespread, the visibility of UNSDI service levels and their responsiveness or otherwise to the user community's needs will increase. UNSDI portals therefore need to be adequately equipped, staffed, and maintained to assume operational responsibilities. Budgets available for Stage 3 must reflect this reality, before operational status is declared. Data custodians also need to be in a position to provide operational levels of data maintenance early in Stage 3 of implementation to ensure currency of archives [ITHACA M21 D12, 2011].

Stage 4: Continuous improvement of UNSDI processes

The final implementation stage involves UNSDI configuration management whereby the information architecture and technical infrastructure built up over the previous Stages is extended and refined over time, as demand for stakeholder services and finances dictate

3.3.1 Current status of the UNSDI

The UNSDI initiative experienced a long period of uncertainty; the lack of budget to be administered by the UNGIWG makes all the efforts expression of volunteers. The result is that the agencies involved in the initiative usually don't destine a budget to it, preferring to invest into internal similar projects which seem to be more helpful in the short term.

In this way actual tasks and responsibilities of each agency are not easy to be detected and understood. Therefore the potential integration between the activities of the UNSDI and the GMES activities is a difficult task: at the time of the last GMES deliverable the contractor couldn't take advantage or cooperate with the UNSDI working groups.

3.4 Identification of reference data information gaps

One of the tasks included in the agreement between the EC and the contractor was the identification of gaps in the reference data collected during the previous phases and the proposal of technical provision with the aim of gaps reduction.

The approach used by the contractor for identifying gaps is based on the analysis of the scores resulted in the previous activity of data quality assessment; the average values of the scores given

to all the datasets, separated for each indicator, were calculated and then put in comparison between each other. It emerged that two of the quality indicators got globally significant lower scores than the others and highlighted a gap in the reference data information: they are *Scale Denominator* and *Update*.

The contractor observed also that, according the European Environment Agency [EEA 2011], the tightest requirements for data to be used in rapid mapping and emergency support mapping are timeliness and accuracy/resolution/scale, together with availability and geographic coverage. Therefore scale and update are elements that require a particular focus, because they're critical both from data quality results and the user requirements perspective.

3.4.1 Gaps in the indicator *Scale Denominator*

Some of the possible applications related to GMES services require detailed and accurate dataset for extracting value added information on the basis of the Reference Data Access service, for example rapid mapping activities after an earthquake occurs. This kind of dataset is generally characterized by a limited spatial extent and usually is not free of charge.

A comparison between the mentioned EEA requirements and the results of the data quality assessment was performed by the contractor; it is resumed in Table 3.3. Huge gaps emerged between requirements and current status of the available datasets, often with several orders of magnitude. Data sources with an adequate reference scale can be found only for the category regarding ports [ITHACA M21 D13, 2011].

3.4.2 Gaps in the indicator *Update*

Again according to EEA, all data categories are requested with the latest available update. The contractor assumed that the terms *latest available* requires data to be updated on a continuous process.

In comparison with the previous case the situation seems to be less critical, as formally only data on population, Digital Elevation Models, Forests and soil do not meet the criterion: they are not systematically and continuously updated. Nevertheless it must be pointed out that most of the declared continuously updated data sources are not completely revised at each release; most of them are locally updated when new specific information is provided and the alignment between the dataset and the reality cannot be granted. Furthermore no control can be established on the data sources updated on a voluntary basis, such as OpenStreetMap and GeoNames⁴ [ITHACA D13 M21 2011].

3.4.3 Proposal of technical provision

The consortium, considered the gaps on reference data information raised during the analysis, proposes an integration between various data sources as technical provision to gaps filling; this approach is particularly effective for solving the issues related to the dataset scale.

⁴ www.geonames.org

Three different approaches of integration are taken into consideration [ITHACA D13 M21, 2011]:

- Multi-scale approach, aims at storing and managing multiple representations of the same object: it implies a more complex database model and the necessity of developing procedures for data management. For example automatic selection of the correct geometry to be represented in function of the map scale is necessary, that allows an optimized and complete geographic storage, analysis and representation;
- Substitution approach, aimed to keep only the most precise and reliable data. That implies a simpler data model but with a single dataset that may have different reference scales;
- Country specific datasets, aimed at the creation of local datasets and loading new data on countries with specific needs.

| EEA categories | RDA Lot2 sub-topics | EEA target | RDA Lot2 best available |
|---|--|-------------------|--------------------------------|
| Administrative boundaries | Country political level 0 boundaries Coastline and maritime boundaries Sub-nationals boundaries Areas of conflict | 1:5.000 | 1:250.000 |
| Settlement locations and toponyms | Population centres | 1:5.000 | 1:250.000 |
| Transport networks - roads | Transportation - roads | 1:5.000 | 1:100.000 |
| Transport networks - railways | Railways | 1:50.000 | 1:100.000 |
| Transport networks - airports | Airports | 1:35.000 | 1:250.000 |
| Transport networks - ports | Harbours | 1:100.000 | 1:100.000 |
| Population data | Urban and rural population density and distribution | 1:300.000 | 1:3.000.000 |
| Digital elevation model | Terrestrial elevation | 1:3.000 | 1:100.000 |
| Forest maps | Land cover/use, forests and vegetation | 1:50.000 | 1:1.000.000 |
| Protected areas | Parks, conservancies and protected areas | 1:50.000 | 1:1.000.000 |
| Soil information | Soils and soil properties | 1:100.000 | 1:3.000.000 |
| Geological maps | Geology Hydro-geology | 1:100.000 | 1:5.000.000 |
| Hydro graphic information – water bodies | Drainage, rivers and flow routing Surface water bodies | 1:25.000 | 1:100.000 |

Table 3.3 Scale Denominator: comparison between EEA requirements and results of the data inventory phase (courtesy of ITHACA)

3.5 Demonstrators and Validation

It has been said that setting up demonstrators is required by the good practises in building RDA service process, in order to validate the use of global reference data in emergency and global land monitoring services. Demonstrators are applications aimed at testing the functionalities and show the advantages of a product, in the present case a RDA.

Since the data access is required to be an online tool, consequently the demonstrators are Web applications; the objective of the demonstrators is providing access to the reference geospatial data coming from the data inventory phase and filtered on a quality basis.

Even if innovation in terms of technology is not required in the demonstrators, innovative applications were adopted and developed; the technology stack is composed by free and open source OGC-compliant tools. The technology solutions are described in details in chapter 6.

Three demonstrator applications were developed and are maintained by the consortium; each application was customized in order to match the needs of an identified user:

- Extreme Rainfall Detection System – ERDS;
- Fires Monitoring System;
- GeoNode for CoSA.

3.5.1 Fires Monitoring System

The Fires Monitoring System is a Web service aimed at providing value added information on the active fire impact on local resources; it is focused on the continent of Africa.

Fires can be caused either by nature, such as lightning and lava, or by humans in accidental way but also on purpose in agriculture: in all the cases fires can destroy natural resources and human structures. The tool introduced in this paragraph can offer basic information in order to better define the threshold between a fire event and a disaster.

The technology of this demonstrator is based on a custom instance of GeoExplorer capable to query a PostgreSQL database; an example screenshot is contained in Figure 3.1. It is accessible at the following URL:

<http://www.ithacaweb.org/geoexplorer/composer>

On the graphical point of view the application is made by a big map canvas where multiple geospatial data sources can be merged together thanks to the functionalities of GeoExplorer. A table of content with the layer tree is showed on the right side of the screen. A main toolbar is displayed on top of the screen and contains many of the common WebGIS commands such as pan, zoom and getting feature information.

The map is composed by base reference layers and by thematic layers; the base reference layers come from the data collection retrieved during the inventory phase and filtered on a quality basis.

Some of these are the administrative boundaries, the surface hydrology and the transportation network.

Land cover and population layers can be listed as thematic layer and come from the RDA repository as well. An external layer comes from a service that integrates remote sensing and GIS technologies to deliver global MODIS hotspot/fires locations and burned area information to natural resource managers and other stakeholders: it is the Fire Information for Resource Management System (FIRMS). For the purpose of the demonstrator the layer of the active fires over the past 48 hours is showed, but other time period layers are provided.

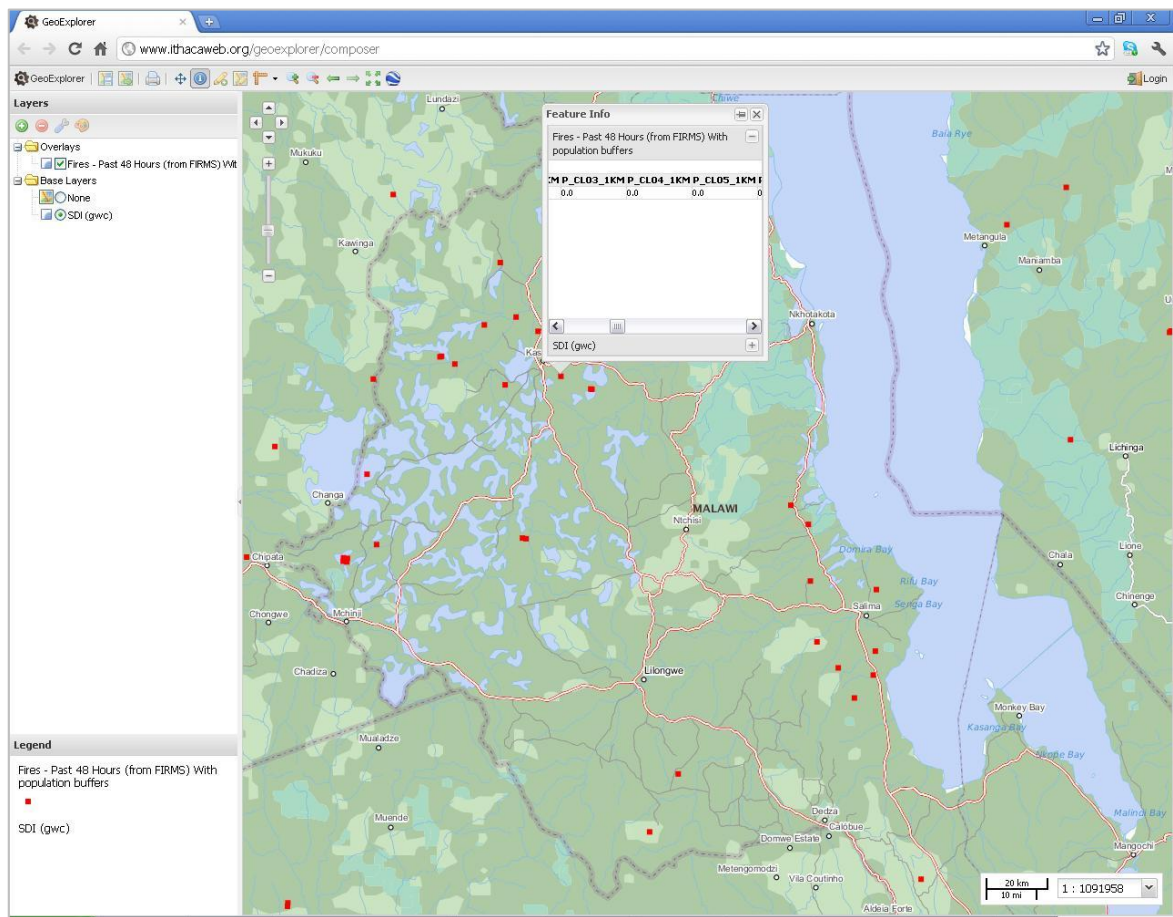


Figure 3.1 Screenshot taken from the demonstrator Fires Monitoring System

The application provides value added information other than the active recent fires on a base reference map of Africa. In particular a procedure based on PostGIS functions (see section 6.4.1) generates two circular buffers around the fire points, respectively with 1 and 10 km of radius. For each buffer the percentage of the land cover category contained in it and the comprehended population are calculated by mean of spatial intersections with the correspondent thematic layers. The application was then developed by mean of *Ext.js* and *GeoExt Javascript* libraries (see section 6.4.5) in order to show the calculated value added information on a table, when the user takes advantage of the *Get Feature Info* tool on a fire point [ITHACA M21 D14, 2011].

Possible users are organization responsible for monitoring emergencies worldwide and planning rapid response on the field because the application provides a first approach for estimating the population and the type of land cover affected by a fire.

3.5.2 ERDS

An Extreme Rainfall Detection System is a service that allows a near real time global flood monitoring: it is aimed to deliver early alerts on hydrological basins possibly affected by flood events.

The service is up and running at the following *URL*:

<http://erds.ithacaweb.org/gmes-rda>

It takes advantage of a WebGIS technology through a tailored instance of the GeoExplorer, similar in graphics to the Fires Monitoring System application. It is fed by base reference layer as well as thematic layers: base reference layers come from the GMES RDA data repository and comprehends administrative boundaries, surface hydrology and transportation layers. Thematic layers comprehend human population as well as land cover; furthermore rainfall data are used in the application, even if they're not displayed directly but they are considered as input data in a procedure capable of determining floods at catchment level as described below.

The system is based on Tropical Rainfall Measuring Mission satellite rainfall data processing with an update time equal to 3 hours; rainfall time series processing takes into consideration data from 1998 and allows to calculate the depth-duration-frequency curves required to have a complete historical database of heavy rainfalls with global coverage and single basin resolution. Heavy-rainfall alerts are triggered on specific basins as a result of the comparison between empirical curves estimated through the historical analysis and the near real time rainfall data [ITHACA M21 D14, 2011].

The user can see polygons corresponding to hydrological basins, displayed in different colours in order to indicate the three classes of magnitude of the alert according to the rainfall severity: see Figure 3.2 for an application screenshot. Countries currently under alert are displayed in a drop down dynamic list: selecting one of those countries triggers the application to zoom on the country extent and to create and display markers on alerted areas. The name of the main river of the watershed, the event beginning and ending dates and the population potentially involved, calculated by intersecting the global population layer with the footprint of each river basin, are displayed when the user selects a specific marker.

The UN WFP has been identified as the end user of the ERDS demonstrator; in particular the system is being tested in supporting the early warning activities of the Emergency Preparedness and Response Branch.

Furthermore a feedback tool was developed in order to allow the end user on the field to validate the system. When the feedback tool is triggered, a window is opened; it drives the user through selecting a point on the map and indicating whether it is flooded or not.

- storage of the damage assessment interpretation results, exploiting the SDI versioning capabilities.

CoSA demonstrator is based on an open source software able to provide these functionalities: the GeoNode. This software is thoroughly described in chapter 6.

4. Data Quality Assessment Protocol

In chapter 1 the role played by geo-referenced information in natural disasters and environmental emergencies management was described. Then in chapter 2 the GMES Reference Data Access call for tenders and the contractor were introduced and the best practises for building a RDA service were described in details. Finally in chapter 3 a complete description of the status of the GMES RDA Lot2 project was provided.

The data quality assessment is amongst the activities for setting up a RDA service; a protocol with a set of rules and procedures aimed at performing the assessment must be defined. Although various data quality protocols exist in the field of geo-referenced information, the majority of them is conceived for using during data acquisition and production and therefore isn't immediately applicable in the present case, where data checking is the leading activity.

The definition of a protocol for data quality assessment and its implementation are essential in the workflow of a team devoted to geospatial data collection for obtaining good quality products. The importance of data quality was felt at the international level and out of the edges of the scientific community, to such an extent that the EC and the ISO were charged with defining geospatial data quality protocols to be considered valid to a community of people as wide as possible, eventually a global scale.

A data quality protocol was defined in the framework and for the purposes of the GMES RDA Lot2 project, by bringing together: the protocols defined at the international level in terms of geometric and topographic approach; the user needs assessment in terms of specifications on geographic extent, dataset reference scale and updates; the call specifications in terms of licence conditions; the contractor's experience in the field of geo-referenced information for environmental emergency management; the choice of leaving a wide range of choices to the final users in selecting geo-spatial data because he is going to be informed on partial data quality results that can take into higher or lower account according to his specific needs.

Furthermore the contractor decided to apply a quantitative approach to the data quality assessment protocol in order to provide the communication and the publication of the results in the clearest possible way.

4.1 Data Quality definition

A bibliographic review allowed to come to a definition of geospatial data quality: two sides of the quality are taken into consideration, both the coherence of the data in comparison with the real objects that are represented and the coherence between the data content and the end user's requirements.

Geospatial data are a “model of reality”: therefore every map or database is a model. During the modelling process, a series of factors may induce errors in the data that are lately produced. All geospatial data are then, at different levels, imprecise, inaccurate, out-of-date and incomplete.

Several authors distinguished two definitions of spatial data quality: *internal quality* and *external quality*. On the one hand *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”. In practice 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 R. *et al.*, 2006].

On the other hand 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”. The evaluation of *external quality* can imply criteria that describe *internal quality*.

On the *internal quality* side description, the ISO 19113 standard on geospatial data quality recommends the use of the following criteria [ISO/TC 211 – 19113, 2002]:

- *Completeness*: presence and absence of features, their attributes and relationships;
- *Logical consistency*: degree of adherence to logical rules of data structure, attribution and relationships;
- *Positional accuracy*: accuracy of the position of features;
- *Temporal accuracy*: accuracy of the temporal attributes and temporal relationships of features;
- *Thematic accuracy*: accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships.

Finally data quality can be defined also by mean of the denotation, which is a radical quality aspect of geographical data: we use the word “denotation” to designate the original, radical function of any representation. What represents is a sign toward what happens to be represented, a sign which expresses both the presence and the intensity of the phenomenon at stake [Devillers R. *et al.*, 2006].

When geographical data are produced, the resulting denotation may depart from the expected denotation, and consequently from the geographic features which happen to be represented. The following are the sources for fluctuations in denotation:

- Modelling and realization of the model of the world: the construction of the model of the reality;

- Modelling and realization of the operations: the act of processing the initial data;
- Space: geometry and topology. While geometry is the numerical location of an object, the topology is the invariant in the geometrical properties of objects;
- Time: date and chronology. They're the analogous of geometry and topology in the time dimension;
- Value: semantic and semiology. They regard the expression of the thematic values of objects.

4.2 Data Quality Assessment procedure

An assessment on denotation quality requires some general principles: assessment is based on comparison of produced data to the world. Elementary measures on objects and statistics are made, validity of measured characteristics may be pronounced on their inclusion within confidence intervals. Eventually, reporting of the results is required [Devillers R. *et al.*, 2006].

A data quality assessment procedure is defined in order to be applied on the spatial datasets collected in the inventory phase [see section 3.1]: the process runs separately for each data sub-topic defined in the GMES RDA Lot2 project deliverable [ITHACA Interim Report M21 D10, 2011]. The procedure comprehends the following steps:

- twelve quality indicators are defined on the basis of the user needs assessment, the ISO specifications [ISO/TC 211 – 19113, 2002; ISO/TC 211 – 19115, 2004] and the call requirements;
- several confidence intervals are set for each indicator;
- each confidence interval is associated to a score between 1 and 5;
- a value for each indicator is obtained through measurements and statistics performed on each candidate dataset;
- a score is assigned to each dataset by comparing the value obtained with the fixed confidence intervals;
- a *Quality Matrix*, containing all the information on the *internal quality*, is created for every sub-topic as a result of the previous steps;
- a weight included in the range 0-1 is then assigned to each indicator at the sub-topic level, thus expressing the importance attributed by a hypothetical user;
- the defined weights are collected together into the so called *Vector of Weights*, that is the implementation of the *external quality* definition;
- 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*;
- finally a ranking of the datasets, based on their quality score, is set up.

In the following pages, all quality indicators are described and the requirements for each of them are thoroughly explained: finally the total score calculation is introduced.

4.3 Data Quality Indicators definition

4.3.1 Geographic Extent

The geographic extent of a dataset consists in the size of the bounding box containing all its features. Usually datasets with large nominal scale have a small extent and vice versa. Where a wide field of view is needed, the extent could be insufficient to cover the area of interest and therefore a time consuming operation of merging different neighbouring datasets is necessary. Consequently issues such as the integration and linking of different datasets arise in cross border situations.

On the purpose of creating a global service, the larger is the extent of a dataset the higher is the score. Table 4.1 resumes the requirements and the respective scores.

| Score | Requirements |
|-------|--------------|
| 1 | Local |
| 2 | Sub national |
| 3 | National |
| 4 | Continental |
| 5 | Global |

Table 4.1 Proposed values for the indicator Geographic Extent and respective scores

4.3.2 Licensing and Constraints

Different data holders deliver their spatial data under various license agreements. Two different elements are taken into consideration: the cost of the data and their possible limitation of use. Table 4.2 contains several models of data distributions, that are taken into consideration, and their respective scores.

| Score | Requirements |
|-------|----------------------------------|
| 1 | Commercial data with limitations |
| 2 | Commercial data |
| 4 | Free data with limitations |
| 5 | Public domain data |

Table 4.2 Proposed values for the indicator Licensing and Constraints and respective scores

4.3.3 Scale Denominator

In most of the applications and software, a digital map can be visualized at different scales using the zoom buttons: this doesn't mean that the information contained in the map is adequate at all visualization scales. Actually, geospatial data have their own nominal scale, that is the larger scale at which they can "lawfully" be visualized or printed. From this point of view the possibility of zooming the map at larger scale than the nominal one has to be interpreted as an occasion of reading better the information that it contains, but not an occasion of reading more.

In particular the concept of nominal scale is deeply linked with the degree of precision and accuracy of the content. The tolerance of the map or spatial data can be derived from the nominal scale: the Italian topographic tradition defines the accuracy as the value of 0.2 mm at the nominal scale, which is derived from the smallest thickness of a printed map line [Comoglio G., 2004]. Furthermore, at the international level, the National Map Accuracy Standards (NMAS) defines a less restrictive tolerance, that corresponds to the value of 1/50" at the nominal scale for map scale smaller than 1:20.000. In the framework of the RDA project NMAS thresholds have been revised according to the National Standard for Spatial Data Accuracy (NSSDA) guidelines, using formulas proposed in the Appendix 3-D of the "Geospatial Positioning Accuracy Standards" document [FGDC, 1998]. Specifically, NSSDA uses root-mean-square error to estimate positional accuracy. RMSE thresholds are reported in the "RMSE NSSDA" column in Table 4.3. Accuracy thresholds adopted for the quality evaluation are reported in ground distances at the 95% confidence level in the "Accuracy NSDDA" field in Table 4.3.

The indicator *Scale Denominator* can assume the values in Table 4.3: the evaluation is based on the metadata or on the information published by the data holders. In case this information is not provided, some geo-processing operations can be performed in order to give an estimation of the scale: for example as far as linear datasets are concerned, an analysis of the line segments length is carried out and a statistical distribution graph is drawn. A value of the dataset tolerance can be estimated as the smallest value of length with a statistical importance: then using the relations reported above, a value of reference scale is calculated.

Another approach to estimate the scale denominator is to carry out a relative comparison with datasets with a defined map scale. Similar operations can be performed in order to verify the information provided in metadata or published by the data holders.

| Score | Requirements | ACCURACY NMAS (1/50"*SD - CE 90%) (m) | RMSE NSSDA (NMAS/1.5175) (m) | ACCURACY NSDDA (RMSE NSSDA*1.7308 - CE 95%) (m) | Approximate GSD (m) |
|-------|------------------------|---|------------------------------------|--|---------------------------|
| 1 | > 1.000.000 | > 508 | > 335 | > 579 | > 320 |
| 2 | >500.000 - <=1.000.000 | 254-508 | 167-335 | 290-579 | 160-320 |
| 3 | >250.000 - <=500.000 | 127-254 | 84-167 | 145-290 | 80-160 |
| 4 | >100.000 - <=250.000 | 51-127 | 33-84 | 58-145 | 30-80 |
| 5 | <= 100.000 | 51 | 33 | 58 | < 30 |

Table 4.3 Proposed values for the indicator Scale Denominator and respective scores

As far as raster datasets are concerned, approximate reference values of spatial resolution, that is Ground Sample Distance (GSD) expressed in meters, are reported in the last column of Table 4.3, taking into account both achievable accuracy and proper viewing.

4.3.4 Update

Different dataset may have different update rate: the more recently the data has been updated, the more reliable the data is assumed to be.

The evaluation is performed using the information provided on the metadata or directly by the data holders. Random controls can be performed if the time reference is provided feature per feature, for instance in a field of the attribute table of the dataset.

The indicator *Update* can assume the values contained in Table 4.4.

| Score | Requirements |
|-------|-----------------------|
| 1 | More than 5 years ago |
| 2 | From 3 to 5 years ago |
| 3 | In the past 2 years |
| 4 | Annually (planned) |
| 5 | Continuous |

Table 4.4 Proposed values for the indicator Update and consequent scores assigned to the datasets

4.3.5 Temporal Extent

The indicator *Temporal Extent* is applicable only in case of time series data. It refers to the period of time included between the first available release of the data and its last update; details are contained in Table 4.5.

| Score | Requirements |
|-------|-------------------|
| 1 | Months |
| 2 | Years |
| 3 | 10 years |
| 4 | Fifty years |
| 5 | Hundreds of years |

Table 4.5 Proposed value for the indicator Temporal Extent and respective scores

4.3.6 Fitness for Use

The generic concept of fitness for use varies on the basis of the applications in which the data is exploited. Current indicator aims to evaluate pre-processing requirements in order to provide data in a format suitable and optimized for data maintenance, analysis and display, both for map preparation and for data display on desktop and WebGIS applications.

Table 4.6 contains the values that can be assumed by the indicator *Fitness for Use*: the evaluation is performed in a relative way, taking into consideration the characteristics of all the candidates for each sub-topic.

| Score | Requirements |
|-------|------------------------------------|
| 1 | Need of significant pre-processing |
| 2 | Need of solid pre-processing |
| 3 | Need of average pre-processing |
| 4 | Need of slight pre-processing |
| 5 | No need of pre-processing |

Table 4.6 Proposed values for the indicator Fitness for Use and respective scores

4.3.7 Thematic Richness

The indicator *Thematic Richness* corresponds to the amount of information provided by each candidate dataset. It can be referred and calculated proportionally both to the number of fields available providing relevant information for the analysed sub-topic and to the level of adherence of the theme represented by the dataset to the respective sub-topic.

Table 4.7 contains the values that can be assumed by the indicator *Thematic Richness*: the evaluation is performed in a relative way, taking into consideration the characteristics of all the candidates for each sub-topic.

| Score | Requirements |
|-------|--------------|
| 1 | Low |
| 2 | ... |
| 3 | Medium |
| 4 | ... |
| 5 | High |

Table 4.7 Proposed values for the indicator Thematic Richness and respective scores

4.3.8 Integration

One of the most challenging issues in the implementation of a geospatial data service is the need of a perfect integration among datasets, both in the same and in different sub-topics; the most significant example is the comparison between surface hydrology and elevation data. The issue is even bigger if the datasets have different nominal scales, extents, geometrical and thematic structures.

Based on this assumption, the datasets are considered more useful if they're already integrated in a thematic or multipurpose database, since the relationship issues between datasets are already solved by data holders; see Table 4.8 for details.

| Score | Requirements |
|-------|--------------------------------|
| 1 | Stand alone dataset |
| 3 | Part of thematic database |
| 5 | Part of comprehensive database |

Table 4.8 Proposed values for the indicator Integration and respective scores

4.3.9 Data Integrity

Data integrity 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.

The evaluation is performed by mean of processing operations with computer software, in the present case the *ESRI ArcGIS Desktop*. This software can manage several topological rules both in phase of data production and in phase of data quality controls: several controls can be applied to different geometry type of data, such as points, lines or polygons.

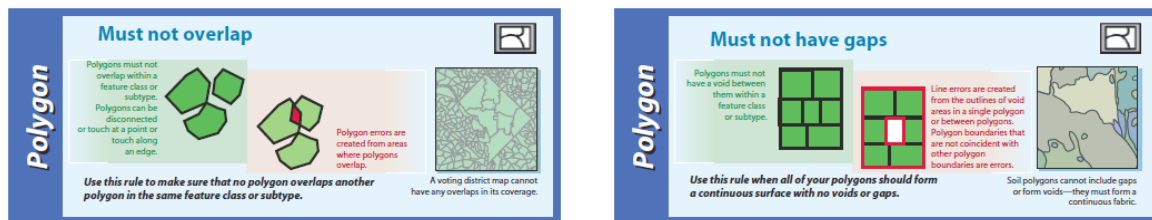


Figure 4.1 Description of two of the topological rules defined during the evaluation process (courtesy of esri)

The evaluation process consists in the following steps:

1. definition of the appropriate topological rules for each dataset: examples are in Figure 4.1;
2. validation of the topology: each feature is compared with the others involved in the topology, and their spatial relationship is controlled in order to find violations of the rules;
3. investigation of the results: the errors are studied in order to understand the nature of the violation and the effective presence of the violations themselves;
4. building of the resume table: for each dataset and for each topological rule, a table record is filled in with the results of the validation and with some notes and comments.

| Score | Requirements |
|-------|--------------|
| 1 | > 5 |
| 2 | 2 – 5 |
| 3 | 0.5 – 2 |
| 4 | 0.1 - 0.5 |
| 5 | < 0.1 |

Table 4.9 Proposed values for the indicator Data Integrity and consequent scores

The indicator *Data Integrity* consists in the percentage of the sum of the errors, for all controls performed on a dataset, over the number of features observed in that dataset: datasets are given a score in compatibility with the values reported in Table 4.9.

4.3.10 Positional Accuracy

It is represented by the mean value of the positional uncertainties for a set of positions, where the positional uncertainties are defined as the distance between a measured position and what is considered as the corresponding true position.

| Score | Requirements (% of errors) |
|-------|----------------------------|
| 1 | Coarse (80-100) |
| 2 | Low (60-80) |
| 3 | Intermediate (40-60) |
| 4 | High (20-40) |
| 5 | Fine (0-20) |

Table 4.10 Values that the indicator Positional Accuracy can assume

Free-released spatial data rarely come with the definition of a value of *Positional Accuracy* in their metadata: this value is deeply related with the reference scale and is derived from it, using an approach which is the opposite of the one presented in section 4.3.3.

In both the ways the value of the positional accuracy is obtained, it is used to draw a buffer around a set of sample features randomly extracted from the entire dataset under evaluation. These buffers are then put in visual comparison with a dataset that has a higher positional accuracy, of at least an order of magnitude. If a buffer doesn't contain a feature from the most accurate dataset, this is considered as an error of the dataset: the final percentage of errors in the dataset is used as indicator of the *Positional Accuracy* of the dataset itself. It can assume values from coarse to fine as summarized in Table 4.10.

| Score | Requirements (% of errors) |
|-------|----------------------------|
| 1 | Coarse (80-100) |
| 2 | Low (60-80) |
| 3 | Intermediate (40-60) |
| 4 | High (20-40) |
| 5 | Fine (0-20) |

Table 4.11 Proposed values for the indicator Thematic Accuracy and consequent scores

4.3.11 Thematic Accuracy

Thematic accuracy controls are random controls 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 quality indicator *Thematic Accuracy*.

The indicator *Thematic Accuracy* can assume values from coarse to fine as summarized in Table 4.11.

4.3.12 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 by mean of intercomparison with other datasets of the same sub-topic: the percentage of missing features over the number of features observed constitutes the indicator *Completeness*. The reference confidence intervals are reported in Table 4.12.

| Score | Requirements |
|-------|--------------|
| 1 | > 5 |
| 2 | 2 – 5 |
| 3 | 0.5 – 2 |
| 4 | 0.1 - 0.5 |
| 5 | < 0.1 |

Table 4.12 Proposed values for the indicator Completeness and respective scores

4.4 Overall Evaluation

4.4.1 Quality Matrix

After the quality assessment is performed on all datasets belonging to a specific sub-topic, the results can be conveniently resumed in a matrix where each row contains the scores assigned to a dataset for each indicator, while every column contains all the scores given to a specific indicator. The matrix described is called *Quality Matrix*: a template is provided in Table 4.13.

| Dataset | Geographic Extent | Licensing and Constraints | Scale Denominator | Update | Temporal Extent | Fitness For Use | Thematic Richness | Integration | Data Integrity | Positional Accuracy | Thematic Accuracy | Completeness |
|-----------|----------------------|---------------------------|-------------------|--------|-----------------|-----------------|-------------------|-------------|----------------|---------------------|-------------------|--------------|
| Dataset 1 | Scores (from 1 to 5) | | | | | | | | | | | |
| ... | | | | | | | | | | | | |
| Dataset n | | | | | | | | | | | | |

Table 4.13 Example of a Quality Matrix

4.4.2 Vector of Weights

The indicators previously defined may not have the same importance in absolute terms; furthermore, relative importance is normally significantly different according to the application in which data are used. For instance the end user may prefer to deal with spatial data that have high scores for some of the indicators, while he could not mind if the same data have low scores for the other indicators, according to the requirement of the application he is dealing with.

On the basis of the experience and sensibility of the data analysts, a weight for each quality indicator is established as a hypothesis of the needs of a potential user, working in the field of emergency management: it is a positive number between 0, if the indicator is not important at all, and 1, if it has the maximum importance. A *Vector of Weights* for the indicators can be set up with the chosen values (see Table 4.14).

The *Vector of Weights* constitutes the implementation of the general concept of *external quality* in the protocol. Although a first approach definition of the weights is given by the contractor, a real end user could define and apply his own weights according to the specific current needs. The contractor commits himself to publish all the partial quality scores in order to facilitate these operations.

It is important to highlight that the definition of the *Vector of Weights* must be done at a sub-topic level, in order to ease the process of best quality data choosing for each data category.

| Parameter | Weight |
|---------------------------|----------------------------|
| Geographic Extent | values in the range 0 to 1 |
| Licensing and Constraints | |
| Scale Denominator | |
| Update | |
| Temporal Extent | |
| Fitness For Use | |
| Thematic Richness | |
| Integration | |
| Data Integrity | |
| Positional Accuracy | |
| Thematic Accuracy | |
| Completeness | |

Table 4.14 Assignments of weights to the quality indicators: template of Vector of Weights

4.4.3 Final Scores and Ranking

The final step of the assessment process is the calculation of the total score for each dataset. The vector of total scores is the result of an ordinary matrix product between the *Quality Matrix* and the *Vector of Weights*; it consists in a weighted sum of the scores assigned to each dataset.

On the basis of the quality final scores, datasets can be ordered in a ranking that reveals, for each sub-topic, the best data able to match end user's need.

4.5 Notes on the protocol application

The process of data quality assessment is based on applying a set of indicators to a selection of candidate datasets for each sub-topic. The result of the analysis consists in a numeric value for each dataset which is derived from the contribution of each indicator, weighted on the basis of its importance in relation to the sub-topic itself. As mentioned before, the *Vector of Weights* have been identified by the contractor data analysts.

In relation to the definition of *Thematic* datasets reported in section 3.1, the user interest is more focused on the semantic content of the data rather than on the geometric and geographic content. Unfortunately, being the quality assessment of the semantic content a task that requires an in depth knowledge in a specific field, it is not possible to correctly evaluate the subjective indicators *Data Integrity*, *Positional Accuracy*, *Thematic Accuracy* and *Completeness* without having a proved topic-related expertise.

Therefore the protocol states the application of all the quality indicators to the *Core* datasets whilst the application of the first 8 quality indicators to the *Thematic* Datasets: so their overall assessment is based on the objective quality indicators that can be assessed analyzing the available metadata, without applying the weight vector.

4.6 Conclusions on the protocol

In this chapter the concept of geospatial data quality has been introduced and a protocol for data quality assessment has been proposed. It involves the definition of 12 indicators, each of them can assume values in different ranges: possible values are divided in 5 categories with the introduction of some confidence intervals. It also involves the assignment of a score to the datasets for each indicator, on the basis of the values assumed by the single indicator for every single dataset. Finally a total score of quality is calculated by mean of a sum of the partial scores and on the basis of the weights set for the indicators.

The step of setting the weights for the quality indicators can have a great impact on the final scores, so this operation has to be done carefully. During the GMES RDA project activities the definition of the weights was performed by the contractor data analysts, in the attempt of interpreting the end user's needs and translating them into numerical values. Instead in the ideal world this operation has to be done directly by the end user: since this requires further considerable efforts to the end user, such as performing the calculation himself maybe arranging a targeted spreadsheet, a Web application able to extract the dataset of best quality, on the basis of the defined weights, has been developed and is introduced in section 7.6 of this thesis.

During protocol application to datasets gathered in the inventory phase, data analysts experienced difficulties in exploiting the proposed range of values and requirements for some indicators: in these cases alternative values and requirements have been defined and applied. The modifications of the general protocol should be studied case by case in order to highlight limits of the protocol itself and to come to a better definition of the protocol itself as a recursive process.

The reporting of data quality assessment results as final scores in geospatial layer metadata was realized in a Web platform from which all datasets can also be previewed and downloaded (see section 7.5): the strategy of results publication has to be studied carefully because on one hand it

allows the best geospatial data choice but on the other hand there is a risk of misleading and incorrect interpretation to non-expert users.

5. Data Quality Protocol application

In section 3.1 the data classification in topics and sub-topics was introduced and in chapter 4 the data quality protocol adopted in the GMES RDA Lot2 project was described. The protocol was applied on a sub-topic base, that is the group of datasets belonging to a certain subtopic were assessed separately from the others: therefore for each sub-topic a ranking of the most suitable datasets was defined.

The quality assessment was performed by a team of analysts; each of them processed entire sub-topics for avoiding discrepancies in the evaluation of the same data category, as a guarantee of the correctness of the ranking.

Considering the high variability of the candidate datasets characteristics, for example in geometry types, the protocol was defined in order to be strict and straightforward to follow, but at the same time as general as possible in order to be easily adapted to the particular and specific situations. As a consequence several techniques were applied by the analysts for defining a quality score.

In the current chapter examples of the protocol application and results are presented; they regard two data sub-topics that are different in spatial type: vector data and raster data. Comments and suggestions for protocol improvement are reported in the final part of the chapter.

5.1 Country Political Level 0 Boundaries

The sub-topic *Country Political Level 0 Boundaries* belongs to the GMES RDA Lot2 topic *Boundaries: coastal, administrative and areas of special interest*. In particular it comprehends polygon datasets that represent the world countries border and shape: the related database table usually contains information about the country such as name, population, capital name etc. These datasets constitutes a good example of vector datasets.

The following datasets were selected during the inventory phase to be part of the sub-topic and submitted to the data quality assessment:

- Vector Map Level 0 (VMAP0);

- Vector Map Level 1 (VMAP1);
- Global Administrative Areas (GADM);
- Global Administrative Unit Layers (GAUL);
- World Vector Shoreline Plus (WVS+);
- Global Map, Administrative Areas.

Many of the mentioned data sources are databases containing information in many fields, other than country political boundaries: in these cases only the datasets related to country boundaries were extracted from the database and took part of the assessment.

Finally, concerning Global Map, only a sample of the entire dataset was analyzed.

5.1.1 Datasets introduction

The VMAP0 is a global database that provides worldwide coverage of vector geospatial data with a reference scale equal to 1:1,000,000. It is derived from the Digital Chart of the World, that is a global geospatial database produced by the US National Imagery and Mapping Agency. The VMAP0 database is entirely freely accessible and usable, composed by vector data and organized into 10 themes: *Boundaries* is one of the themes. The current and last edition of the VMAP0 was released in the year 2000; its use on maps should be accompanied by the presence of a disclaimer.

The VMAP1 is a global vector database at 1:250,000 scale with the same structure of VMAP0. VMAP1 data are divided into a mosaic made of 234 geographic zones. Only 57 selected zones are released free of charge; on the contrary many countries consider maps and cartography a state monopoly for business or security reasons, therefore data for those countries are not released.

The GADM is a spatial database of the location of the world's administrative areas or administrative boundaries. Administrative areas in this database are countries and lower level subdivisions such as province and departments, and for each area it provides some attributes, foremost being the name and variant names. It is available for download from the website and to use for academic and other non-commercial use [GADM website].

The GAUL is an initiative implemented by FAO and funded by the European Commission. It aims at disseminating and compiling the most reliable spatial information on administrative units for all the countries in the world, providing a contribution to the standardization of the spatial dataset representing administrative units. The GAUL always maintains global layers with a unified coding system at country, first and second administrative levels. It is released once a year and the target beneficiary is the UN, the Universities and other authorized international and national institutions and agencies; a disclaimer should always accompany any use of the GAUL data [GAUL metadata on FAO-GeoNetwork website].

The WVS+ is a database that contains global shorelines, international boundaries and country names. This database was created in 1995 by NIMA and is composed by several libraries with different reference scale, in the range between 1:250,000 and 1:120,000,000.

The Global Map is a database resulted from a global project promoted by the government of Japan with the aim of bringing all nations and concerned organizations together in order to collaboratively develop and provide easy and open access to worldwide geographic information at

a scale of 1:1,000,000. At present data for 71 countries and 4 regions are available. *Boundaries* is one out of the 8 themes of the database. The first version of the database was released in 2000 and is mainly based on existing global geographic data, such as VMAP0.

5.1.2 Indicators evaluation

Geographic Extent

The analysis was done through the reading of the metadata and of the information published by the respective data holders and an activity of visual control of the data with a desktop GIS software, in the present case ESRI ArcGIS Desktop. The visual control, being partially subjective, needs guidelines development and some training of the evaluators, in order to grant quality and homogeneity of the attributed scores.

It must be pointed out that the dataset Global Map, even if it is a global project, doesn't have actually a global coverage, since many countries in the world didn't apply to the project itself: this is the reason of the lower score assigned to it. All the results are resumed in Table 5.1.

| Dataset | Geographic Extent | Score |
|-------------------|-------------------|-------|
| VMAP0 | Global | 5 |
| VMAPI | Global | 5 |
| GADM | Global | 5 |
| GAUL | Global | 5 |
| WVS+ | Global | 5 |
| Global Map | Continental | 4 |

Table 5.1 Summary of scores given the datasets for the indicator Geographic Extent

Licensing and Constraints

| Dataset | Licensing and Constraints | Score |
|-------------------|----------------------------|-------|
| VMAP0 | Public Domain Data | 5 |
| VMAPI | Free data with limitations | 4 |
| GADM | Free data with limitations | 4 |
| GAUL | Free data with limitations | 4 |
| WVS+ | Commercial data | 2 |
| Global Map | Free data with limitations | 4 |

Table 5.2 Scores for the indicator Licensing and Constraints

Metadata and the information provided by data holders furnished enough elements to perform the evaluation; only the dataset VMAP0 can be considered of public domain while the VMAP1 is not free of charge for subsets regarding many countries around the world. Instead the datasets

GADM, GAUL and Global Map are free of charge only for non-commercial use. Table 5.2 contains the given scores.

Scale Denominator

The evaluation process was based initially on metadata and on other information provided by data holders. In case of lack of information, some geo-processing operations were performed in order to build and study the frequency distribution of the dataset segments length: from this graph a theoretical value of the dataset reference scale can be derived.

A reference scale was derived in the way described above for the datasets GADM and GAUL: Figure 5.1 and Figure 5.2 respectively report their segments length distribution, with values expressed in decimal degrees. All indicator scores are reported in Table 5.3.

| Dataset | Scale Denominator | Score |
|-------------------|-------------------|-------|
| VMAPO | 1.000.000 | 2 |
| VMAPI | 250.000 | 4 |
| GADM | 500.000 | 3* |
| GAUL | 1.000.000 | 2* |
| WVS+ | 250.000 | 4 |
| Global Map | 500.000 | 3 |

Table 5.3 Summary of scores given for the indicator Scale Denominator. The values with the star are derived from the segment length distribution

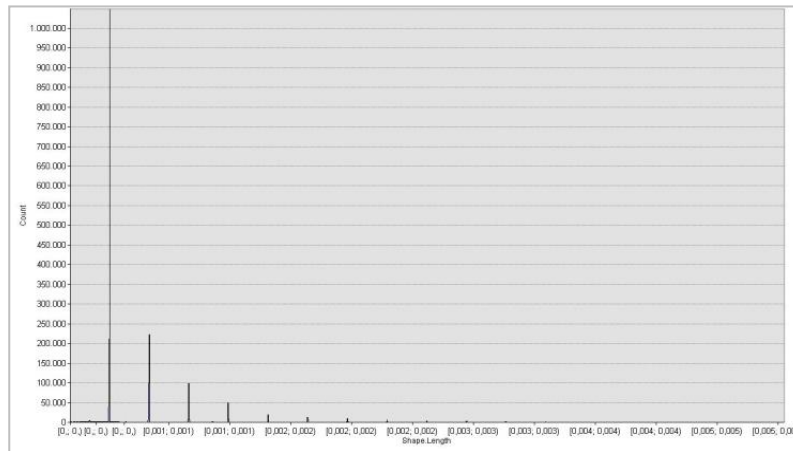


Figure 5.1 Distribution of the segments length in the GADM dataset: values are in decimal degrees. Segments length values are clearly divided in groups of defined length; this means that a sort of processing has been made before data publication. The highest peak, corresponding to the length 30 m, suggests that elements representing coastlines are obtained from the processing of SRTM raster data

Update

The evaluation regarding this parameter was carried out on the basis of the metadata and the information published by data holders: an assessment of the indicator *Update* based on single

features is not possible on this data since there aren't devoted fields in the attribute tables. The results are resumed in Table 5.4.

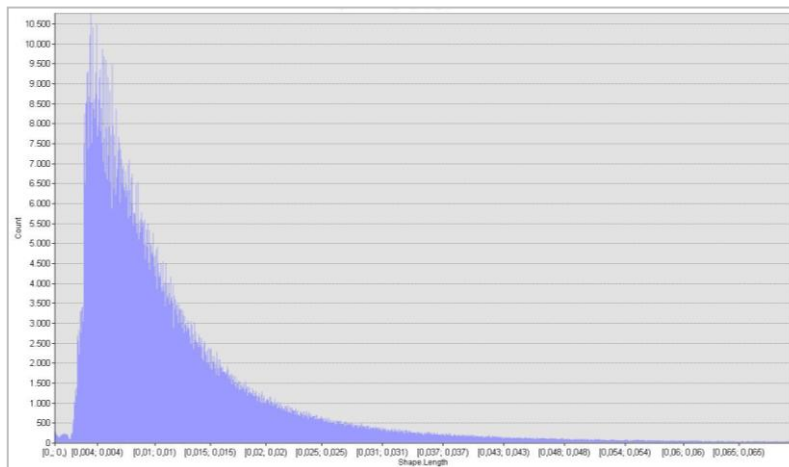


Figure 5.2 Distribution of the segments length in the GAUL dataset: values are in decimal degrees. The distribution in this case shows a clear value of the statistically meaningful minimum length: it corresponds to 300 m, which can be related to a scale denominator of approximately 1.000.000

| Dataset | Update | Score |
|-------------------|-----------------------|--------------|
| VMAPO | In the past 2 years | 3 |
| VMAPI | In the past 2 years | 3 |
| GADM | Continuous | 5 |
| GAUL | Annually (planned) | 4 |
| WVS+ | More than 5 years ago | 1 |
| Global Map | In the past 2 years | 3 |

Table 5.4 Summary of scores given for the indicator Update

Fitness for Use

The evaluation was performed by visually analysing the datasets using GIS software and by studying their metadata.

GAUL dataset obtained the maximum score because it is ready to use for map display and production. VMAPO, VMAPI, WVS+ and Global Map earned 4 points because often the features are divided in tiles: this means that an operation of geo-processing is necessary to derive datasets that can be visualized easily and correctly. Finally the GADM dataset got 3 points because the boundaries overlapping the African coastlines have a shape that follows the pixels edges of the raster dataset they derived from: this means that this dataset contains some nonessential graphical information that could decrease the graphic quality of maps and slow down the processing of visualization with GIS software and WebGIS browsers. The results of the evaluation are resumed in Table 5.5.

| Dataset | Fitness for Use | Score |
|-------------------|--------------------------------|--------------|
| VMAPO | Need of slight pre-processing | 4 |
| VMAPI | Need of slight pre-processing | 4 |
| GADM | Need of average pre-processing | 3 |
| GAUL | No need of pre-processing | 5 |
| WVS+ | Need of slight pre-processing | 4 |
| Global Map | Need of slight pre-processing | 4 |

Table 5.5 Summary of scores for the indicator Fitness for Use

Thematic Richness

The evaluation has been carried out with an operation of visual analysis on the datasets using GIS software and with the investigation of their attribute. A dataset belonging to the present sub-topic is expected to spatially describe the shape and the borders of all the countries in the world; the related alphanumeric information should contain at least the country name. The quality assessment was performed by comparison with these minimum requirements.

The datasets VMAPO, GADM, WVS+, and VMAPI contain the expected information in terms of geometric and alphanumeric. Indeed their attribute table contains the country name both in extended form and in standard coded form: consequently, they were given the maximum score.

The dataset GAUL misses the standard coded form of the country name, therefore scored 4 points. Finally, Global Map misses the extended version of country names. All the results are reported in Table 5.6.

| Dataset | Thematic Richness | Score |
|-------------------|--------------------------|--------------|
| VMAPO | High | 5 |
| VMAPI | High | 5 |
| GADM | High | 5 |
| GAUL | Medium-High | 4 |
| WVS+ | High | 5 |
| Global Map | Medium-High | 4 |

Table 5.6 Summary of scores for the indicator Thematic Richness

Integration

The information provided by data holders and collected during the inventory phase of the GMES RDA Lot2 project activities contains all relevant details for the indicator application; the relevant details are also recalled in the section 5.1.1: all the results are reported in Table 5.7.

Data Integrity

Several topological analyses were performed on the data, whose results are reported in Table 5.10. Summarizing the result, it can be stated that the datasets are excellent from the point of view of the internal integrity, that is the geometric relationship within polygon features included in the

same dataset: in fact, no gaps and no overlaps are present. Only the GAUL dataset has several topology errors but they're not severe, since the overlap and gaps magnitude is smaller than the tolerance derived from the dataset scale: therefore only 1 point was subtracted from the maximum admitted value, as can be noticed in Table 5.8.

| Dataset | Integration | Score |
|-------------------|--------------------------------|--------------|
| VMAP0 | Part of comprehensive database | 5 |
| VMAP1 | Part of comprehensive database | 5 |
| GADM | Stand alone dataset | 1 |
| GAUL | Stand alone dataset | 1 |
| WVS+ | Part of thematic database | 3 |
| Global Map | Part of comprehensive database | 5 |

Table 5.7 Scores in relation to the indicator Integration

| Dataset | Data integrity | Score |
|-------------------|-----------------------|--------------|
| VMAP0 | < 0.1 | 5 |
| VMAP1 | < 0.1 | 5 |
| GADM | < 0.1 | 5 |
| GAUL | > 5 | 4 |
| WVS+ | < 0.1 | 5 |
| Global Map | < 0.1 | 5 |

Table 5.8 Summary of scores for the indicator Data Integrity

| Dataset | Positional Accuracy | Score |
|-------------------|----------------------------|--------------|
| VMAP0 | Intermediate | 3 |
| VMAP1 | Fine | 5 |
| GADM | High | 4 |
| GAUL | Intermediate | 3 |
| WVS+ | Fine | 5 |
| Global Map | Fine | 5 |

Table 5.9 Summary of scores given the datasets for the indicator Positional Accuracy

| Dataset | Rule | Errors | Exceptions | Feature N. | Error % | Comment |
|-------------------|--------------------|--------|------------|------------|---------|---|
| VMAPO | Must Not Overlap | 0 | 0 | 25593 | 0 | |
| VMAPO | Must Not Have Gaps | 0 | 890 | 1833 | 0 | No significant errors |
| VMAPI | Must Not Overlap | 0 | 0 | 1780 | 0 | |
| VMAPI | Must Not Have Gaps | 0 | 1563 | 1780 | 0 | No significant errors |
| GADM | Must Not Overlap | 0 | 0 | 248 | 0 | |
| GADM | Must Not Have Gaps | 0 | 2660 | 57 | 0 | No significant errors |
| GAUL | Must Not Overlap | 1778 | 0 | 274 | 648 | Errors entity is lower than dataset tolerance |
| GAUL | Must Not Have Gaps | 1178 | 0 | 66 | 1784 | Errors entity is lower than dataset tolerance |
| WVS+ | Must Not Overlap | 0 | 0 | 3346 | 0 | |
| WVS+ | Must Not Have Gaps | 0 | 1 | 3346 | 0 | No significant error |
| Global Map | Must Not Overlap | 0 | 0 | 224 | 0 | |
| Global Map | Must Not Have Gaps | 0 | 138 | 224 | 0 | No significant errors |

Table 5.10 Summary of the results of topological controls performed on the datasets geometry elements

Positional Accuracy

The evaluation of this indicator still follows the protocol defined on the previous deliverable M9: a new assessment is being processed under the updated protocol [ITHACA, Deliverable 11.1 M9].

The data quality assessment according to the indicator *Positional Accuracy* was performed by mean of a visual comparison between every dataset under evaluation and some geospatial data having a higher accuracy of at least one order of magnitude. In particular, the following countries have been analyzed: Afghanistan, Bangladesh, Haiti, Colombia, Norway, Mexico and United States of America.

As a result the datasets VMAPO and GAUL appear to be very similar to each other in shape and are less accurate than the others at larger scales of visualization.

It must be pointed out that none of the datasets revealed evident errors in the geometric shape: the results are reported in Table 5.9.

Thematic Accuracy

The evaluation was performed on a randomly selected sample of features, each of them representing a country: the entries of the respective attribute table were inspected in order to find potential errors or typos. No errors were found as can be seen in Table 5.11.

| Dataset | Thematic Accuracy | Score |
|-------------------|-------------------|-------|
| VMAP0 | Fine | 5 |
| VMAPI | Fine | 5 |
| GADM | Fine | 5 |
| GAUL | Fine | 5 |
| WVS+ | Fine | 5 |
| Global Map | Fine | 5 |

Table 5.11 Summary of scores given the datasets for the indicator Thematic Accuracy

Completeness

The assessment of the indicator *Completeness* was performed partially while processing all the other indicators and in a separate way for exploring information gaps in terms of global coverage of the Earth's surface. No relevant information gaps were detected, therefore all datasets can be considered as complete, as can be noticed in Table 5.12.

| Dataset | Completeness | Score |
|-------------------|--------------|-------|
| VMAP0 | < 0.1 | 5 |
| VMAPI | < 0.1 | 5 |
| GADM | < 0.1 | 5 |
| GAUL | < 0.1 | 5 |
| WVS+ | < 0.1 | 5 |
| Global Map | < 0.1 | 5 |

Table 5.12 Summary of scores for the indicator Completeness

5.1.3 Overall evaluation

The results for all the indicators are resumed in the Quality Matrix reported in Table 5.13. The overall evaluation doesn't take into account the indicator *Temporal Extent* because this particular sub-topic is not meant to contain time-series datasets.

For the indicators of quality and consistency of datasets representing *Country Political Level 0 Boundaries*, the user needs assessment brought to the definition of weights contained in Table 5.14. The maximum weight is assigned to all the indicators except two: since the scale of spatial data is

strongly related to their positional accuracy the contractor decided to assign 0.5 points to both of them.

The final scores are calculated as a result of the product between the quality matrix and the vector of weights; then the datasets are ordered in descending order in Table 5.15. The datasets VMAP1 and VMAP0 obtained the highest score: they're the closest to the end user requirements.

Quality Matrix

| Dataset | Geographic Extent | Licensing and Constraints | Scale Denominator | Update | Fitness For Use | Thematic Richness | Integration | Data Integrity | Positional Accuracy | Thematic Accuracy | Completeness |
|-------------------|-------------------|---------------------------|-------------------|--------|-----------------|-------------------|-------------|----------------|---------------------|-------------------|--------------|
| VMAP0 | 5 | 5 | 2 | 3 | 4 | 5 | 5 | 5 | 3 | 5 | 5 |
| VMAP1 | 5 | 4 | 4 | 3 | 4 | 5 | 5 | 5 | 5 | 5 | 5 |
| GADM | 5 | 4 | 3 | 5 | 3 | 5 | 1 | 5 | 4 | 5 | 5 |
| GAUL | 5 | 4 | 2 | 4 | 5 | 4 | 1 | 4 | 3 | 5 | 5 |
| WVS+ | 5 | 2 | 4 | 1 | 4 | 5 | 3 | 5 | 5 | 5 | 5 |
| Global Map | 4 | 4 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 5 | 5 |

Table 5.13 Quality matrix for the data sub-topic Country Political Level 0 Boundaries

Vector of Weights

| Parameter | Weight (0 to 1) |
|---------------------------------|-----------------|
| Geographic Extent | 1 |
| Licensing and Constraint | 1 |
| Scale Denominator | 0.5 |
| Update | 1 |
| Fitness For Use | 1 |
| Thematic Richness | 1 |
| Integration | 1 |
| Data Integrity | 1 |
| Positional Accuracy | 0.5 |
| Thematic Accuracy | 1 |
| Completeness | 1 |

Table 5.14 Weights assigned to the quality indicators for the sub-topic: Country Political Level 0 Boundaries

Final Scores and Ranking

| Dataset | Final Score |
|-------------------|-------------|
| VMAPI | 45.5 |
| VMAPO | 44.5 |
| Global Map | 43 |
| GADM | 41.5 |
| GAUL | 39.5 |
| WVS+ | 39.5 |

Table 5.15 Results of the quality assessment: total scores and ranked datasets

5.2 Urban and Rural Population Density and Distribution

The sub-topic *Urban and Rural Population Density and Distribution* belongs to the topic *Human Population: Population Centres and Distribution*. This contains also the sub-topic *Population Centres*. The first of the two sub-topics represents the human population with a quantitative approach and a continuous coverage; the datasets belonging to this sub-topic are all raster because the raster data type is the most suitable for storing information with these characteristics. Instead the second of the sub-topics represents the population centres that pass a threshold of minimum population or area. The typical dataset belonging to it is a vector dataset that describes either only the position of the centre with a point feature or also the shape of the centre itself with a polygon feature; ancillary information such as the population contained in the centre or the centre name.

The datasets gathered during the inventory phase in the sub-topic *Urban and Rural Population Density and Distribution* are the following:

- Food Insecurity, Poverty and Environment Global GIS Database (FGGD);
- Gridded Population of the World (GPW);
- Global Rural-Urban Mapping Project (GRUMP);
- World Water Development Report II (WWDRII);
- World Resources Institute (EOS);
- LandScan.

The candidate's data sources are not all specific to population density and distribution. Population data have been extracted according to the following rules:

- FGGD. Rural population density accessed through FAO GeoNetwork portal⁵;
- WWDRII. Population density and distribution raster accessed through specific website.

⁵ <http://www.fao.org/geonetwork/>

All the datasets listed are raster with a global coverage: they show the higher or lower presence of population in a certain area of the world. The cell values usually represent an estimation of the number of people present in the correspondent area on the ground. The spatial coordinates of all of them are expressed in the Geographic Coordinate System WGS 84.

An improvement to the data information content is provided by some of the datasets, which divide the earth's surface in rural and urban areas: the population distribution and the population density are in this case provided for urban areas and for rural areas in separate files.

5.2.1 Datasets introduction

FGGD is a global GIS database prepared by the FAO in order to support global analyses carried out in the food security and poverty fields. It comprises six themes: Boundaries and topography, Population, Socio-economic and nutrition indicators, Environmental Conditions, Land productivity potential, Land use patterns and Land cover. Datasets are provided either in vector format with a scale of 1:5,000,000 or in raster data format with a resolution of 5 arc-minutes.

GPW and GRUMP are two gridded global population maps produced using different methods at different spatial resolution by the US Centre for International Earth Science Information Network based in Columbia University.

The UN World Water Development Report is the UN's flagship report on water, is released every three years in conjunction with the World Water Forum. It is a comprehensive review that gives an overall picture of the state of the world's freshwater resources and aims to provide decision-makers with the tools to implement sustainable use of water [UNESCO website]. A global database is released together with the report; it contains data on several topics and the human population is one of them [WWDRII download website].

LandScan is a global population distribution data available at a resolution equal to 30x30 arc-seconds; it represents an ambient population, averaged over 24 hours. The LandScan uses spatial data and imagery analysis technologies and a multi-variable dasymetric modelling approach to disaggregate census counts within an administrative boundary. It is produced by the Oak Ridge National Laboratory and the United States Department of Energy [LandScan website].

5.2.2 Indicators evaluation

Geographic Extent

The analysis was done through the reading of the metadata and of the information published by the respective data holders and an activity of visual control of the data with a desktop GIS software, in the present case ESRI ArcGIS Desktop. The latter part has been repeated few times during the analysis of the other indicators in order to potentially modify the previously given score with more experience.

Some datasets have a real global extent, in the sense that they cover the earth's surface from -180 to 180 degrees in longitude and from -90 to 90 degrees in latitude. Many others don't cover the area of Arctic and Antarctica: since the population in those areas can be considered as null, these datasets are retained as with global extent. With this assumption all the datasets have a global

extent, so the score assigned for the respective indicator is the maximum for all, as can be seen in Table 5.16.

| Dataset | Geographic Extent | Score |
|-----------------|--------------------------|--------------|
| FGGD | Global | 5 |
| GPW | Global | 5 |
| GRUMP | Global | 5 |
| WWDRII | Global | 5 |
| EOS | Global | 5 |
| Landscan | Global | 5 |

Table 5.16 Summary of scores given the datasets for the indicator Geographic Extent

Licensing and Constraints

All the dataset listed for the present subtopic are free and available through the internet. They are protected by copyright and there are constraints related to specific use. The indicator has a score equal to 4 for all datasets, as can be seen in Table 5.17.

In particular the dataset Landscan is not free for all purposes; the registration to the website⁶ is necessary and the option of downloading the dataset is granted only after a declaration of purposes.

| Dataset | Licensing and Constraints | Score |
|-----------------|----------------------------------|--------------|
| FGGD | Free data with limitations | 4 |
| GPW | Free data with limitations | 4 |
| GRUMP | Free data with limitations | 4 |
| WWDRII | Free data with limitations | 4 |
| EOS | Free data with limitations | 4 |
| Landscan | Free data with limitations | 4 |

Table 5.17 Scores for the indicator Licensing and Constraints

Scale Denominator

Rarely the metadata related to raster data indicate a value of the reference scale because of the nature of the data itself: the indication of the pixel size is more direct and clear.

In literature several ways of determining the scale denominator from the pixel size are described (see section 4.3.3). All the datasets have pixel size included in the range 30 arc-seconds – 30 arc-minutes: even using the most restrictive approach in determining the value of the current indicator, all of them reach values that are superior than the maximum threshold taken into consideration, that is 1,000,000. In this way all the datasets would have the minimum score for

⁶ <http://www.ornl.gov/sci/landscan/>

this indicator even if the difference between the maximum and the minimum pixel size is 60 times the minimum.

Therefore in order to take into account the great difference in pixel size, a procedure specific for this case was used in order to assign more proper scores. The pixel size values and the scores assigned are showed in Table 5.18.

| Dataset | Pixel Size | Score |
|-----------------|-------------------|--------------|
| FGGD | 5' | 2 |
| GPW | 2.5' | 3 |
| GRUMP | 30" | 5 |
| WWDRII | 30' | 1 |
| EOS | 30' | 1 |
| LandScan | 30" | 5 |

Table 5.18 Summary of scores given for the indicator Scale Denominator

Update

The evaluation in relation to this indicator was carried out on the basis of the metadata and the information published by data holders: the results are resumed in Table 5.19.

The current Sub-topic contains datasets which represent information related to a different epoch than the publication one. Presumably this is due to the approach taken during the production phase, to the input data and to the algorithm applied to them. Population census data are the most important ones amongst the others. The possibility of finding population census data is related to the availability of the data themselves in countries where investigations on the population are performed rarely: they constitute the bottle neck of the global population census update rate.

The datasets FGGD, GRUMP, WWDRII are referred to the year 2000, while the GPW is referred to the year 2005. The most updated one is the LandScan, which was released in 2008.

The less updated is the EOS; since the date of release is 1995, the indicator assumes a value out of scale. A score equal to 0.5 was assigned in order to take this fact into consideration.

| Dataset | Update | Score |
|-----------------|-----------------------|--------------|
| FGGD | More than 5 years ago | 1 |
| GPW | More than 5 years ago | 1 |
| GRUMP | More than 5 years ago | 1 |
| WWDRII | More than 5 years ago | 1 |
| EOS | Out of scale | 0.5 |
| LandScan | From 3 to 5 years ago | 2 |

Table 5.19 Summary of scores given for the indicator Update

Temporal extent

Some of the datasets listed, in particular GPW, GRUMP and EOS, provide data related to different epochs in the past and a forecast for the future. Anyway this approach cannot be fully considered as time series data: for this reason the indicator *Temporal Extent* is not significant for the current sub-topic.

Fitness for Use

The evaluation was carried out taking into consideration the effort needed to exploit different candidate datasets in a GIS environment and through the analysis of their metadata. Results are shown in Table 5.20.

The raster data type itself of the dataset listed for this sub-topic makes them ready and optimized for data maintenance, analysis and display, on desktop and WebGIS applications. Some of the dataset are also provided by data holders with a complete symbolization. In general operations of pre-processing are not required.

| Dataset | Fitness for Use | Score |
|-----------------|---------------------------|-------|
| FGGD | No need of pre-processing | 5 |
| GPW | No need of pre-processing | 5 |
| GRUMP | No need of pre-processing | 5 |
| WWDRII | No need of pre-processing | 5 |
| EOS | No need of pre-processing | 5 |
| LandScan | No need of pre-processing | 5 |

Table 5.20 Summary of scores for the indicator Fitness for Use

Thematic Richness

The current subtopic requires a dataset to give detailed information about the global distribution of the population, both in urban and rural areas. It should cover the Earth's surface with a series of features and an estimation of the number of people present in the area covered by each feature should be reported. From this point of view the choice of using the raster data type is particularly advantageous because it covers the Earth with a grid of features with a fixed dimension and gives the number of people present in the area covered by the pixel as a digital number.

The dataset listed for the present sub-topic are provided with geographic coordinates, referenced to an ellipsoidal datum; this means that the pixel dimension varies in a latitude dependent way. The horizontal dimension decreases as the latitude increases, while the vertical one increases but with an unimportant entity. The consequent behaviour is that the digital number isn't exactly proportional to the population density and there can be an incorrect data interpretation by a non-expert end user while visualizing them with a global extent. The data holders GPW, GRUMP, EOS provide also a dataset of population density in order to prevent this issue.

The data holders of GPW, GRUMP and WWDRII divide the Earth's surface in two categories: rural areas and urban areas. The population distribution in each of the two categories is provided

separately with two different datasets: urban and rural population distribution. In both cases data can be provided in the form of people raw count or population density.

In terms of *Thematic Richness* all the datasets can be considered with high richness but FGGD, because it is focused only on rural areas and doesn't contain the information related to urban areas.

The result of the evaluation of the indicator *Thematic Richness* is summarized in Table 5.21.

| Dataset | Thematic Richness | Score |
|-----------------|--------------------------|--------------|
| FGGD | Medium-Low | 2 |
| GPW | High | 5 |
| GRUMP | High | 5 |
| WWDRII | High | 5 |
| EOS | High | 5 |
| LandScan | High | 5 |

Table 5.21 Summary of scores for the indicator Thematic Richness

Integration

The information collected on the data holdings and reported in section 5.2.1 was enough to face the evaluation for this indicator: the results are reported in Table 5.22.

In the framework of the GMES RDA Lot2 project and in particular in relation with the data quality assessment activities, a database can be considered as comprehensive when it contains data interesting for different topics of the project. Instead it is considered thematic when it contains data interesting for at least more than one sub-topic. With this assumption the presence of more than one dataset in a database is not strictly related to the possibility of covering themes of different sub-topics. For example the databases named GPW and GRUMP contain at least a dataset of population distribution and one of population density with analysis repeated many times and spaced by 5 years. The EOS database contains time-series datasets: every dataset has a different time reference. All these databases are considered as stand-alone datasets because, even if in a strict approach they contain many datasets, they cover the information needed only for one sub-topic.

| Dataset | Integration | Score |
|-----------------|---------------------------|--------------|
| FGGD | Part of thematic database | 3 |
| GPW | Stand-alone dataset | 1 |
| GRUMP | Stand-alone dataset | 1 |
| WWDRII | Part of thematic database | 3 |
| EOS | Stand-alone dataset | 1 |
| LandScan | Stand-alone dataset | 1 |

Table 5.22 Scores for the indicator Integration

Instead the databases FGGD and WWDRII are thematic with the aim of supporting global analyses carried out in the food security and poverty fields or in water resources field: many datasets with information interesting for other topics of the GMES RDA Lot2 project are contained.

Data Integrity

The evaluation process as it was defined in section 4.3.9 can't be applied to raster data. In the present case only the data machine-readability can be evaluated: all the datasets are perfectly machine-readable.

Positional Accuracy

In the present case the positional accuracy is completely driven by the large gap in terms of pixel size among the listed datasets, as it is showed in Figure 5.3. Since pixel sizes have already been taken into consideration during the evaluation of the quality indicator *Scale Denominator*, the current indicator wasn't evaluated.

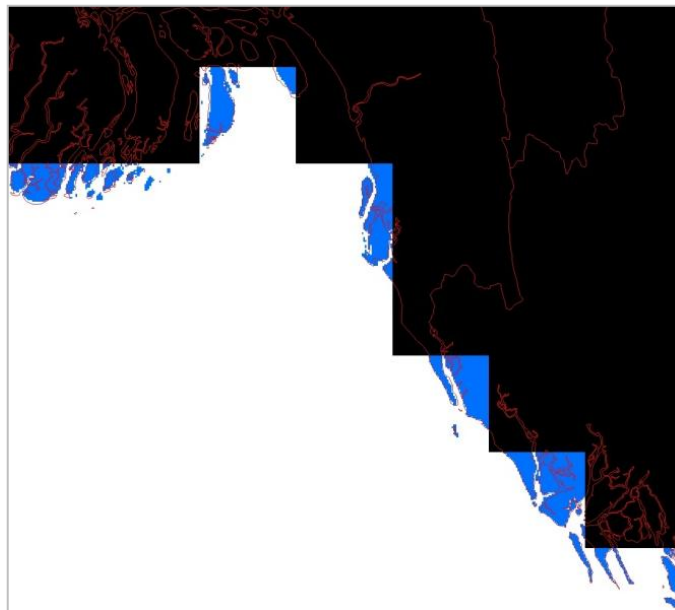


Figure 5.3 The figure shows a dataset from WWDRII database in black and the Landscan in blue. It can be noted how the pixel size affects the positional accuracy of the coastline: a level 0 boundary dataset with a higher accuracy (GAUL) is reported in red as reference data

Thematic Accuracy

Raster datasets are not provided with an attribute table. The information content is stored as pixel value; for the current sub-topic it is expressed in terms of number of people or number of people per square kilometer. Moreover checking the accuracy of pixel values requires a comparison with information on population which is not transferable to the same spatial framework of the raster without bringing a high component of uncertainty in the value itself.

Therefore the indicator was evaluated on the basis of the information provided by data holders, as follows. All dataset were given the maximum score except the dataset GRUMP because the data holder declares that the data is provided in *alpha* version: see Table 5.23 for details.

| Dataset | Thematic Accuracy | Score |
|-----------------|-------------------|-------|
| FGGD | Fine | 5 |
| GPW | Fine | 5 |
| GRUMP | Intermediate | 3 |
| WWDR11 | Fine | 5 |
| EOS | Fine | 5 |
| LandScan | Fine | 5 |

Table 5.23 Results of the analysis on Thematic Accuracy

Completeness

For all datasets a frequency histogram of pixel values was calculated in order to detect the presence of *NoData* values, that correspond to areas for which the population distribution is not provided. These areas could comprehend also seas and oceans: a visual comparison between the datasets in evaluation and a dataset representing the coastlines permitted to extract only the real data gaps. In these cases the ratio between the number of *NoData* pixels over the raster size was calculated in order to evaluate the current indicator.

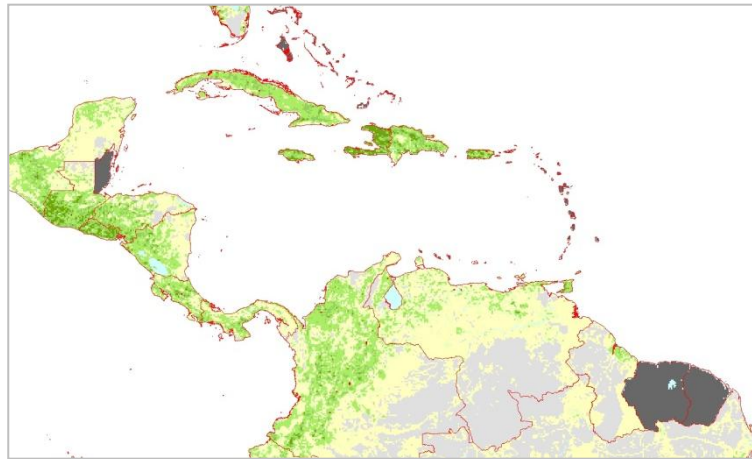


Figure 5.4 A sample of the dataset of Rural Population Density from FGGD database. The dark-grey areas are the ones with NoData values

Only one dataset shows clearly some gaps, it is the FGGD one; see Table 5.24 for details on results. Figure 5.4 shows an example of data gaps: in Central America the countries Suriname, French Guiana and Belize haven't population nor population density values.

| Dataset | Completeness | Score |
|-----------------|--------------|-------|
| FGGD | 1.19% | 3 |
| GPW | 0 | 5 |
| GRUMP | 0 | 5 |
| WWDRII | 0 | 5 |
| EOS | 0 | 5 |
| LandScan | 0 | 5 |

Table 5.24 Summary of scores for the indicator Completeness

5.2.3 Overall evaluation

A complete and concise view of the *internal quality* assessment results is contained in the *Quality Matrix* for the sub-topic *Urban and Rural Population Density and Distribution* of Table 5.25. The abbreviation *NA* stands for “not available” and is put in correspondence to the indicators that were not applied.

The contractor proposed the *Vector of Weights* of Table 5.26 for the assessment of the *external quality*: the indicators *Licensing and Constraints*, *Fitness for Use*, *Integration* were considered less important than the others. Clearly a weight equal to 0 is put in correspondence to the indicators that weren't applied in the assessment.

Quality Matrix

| Dataset | Geographic Extent | Licensing and Constraints | Scale Denominator | Update | Temporal Extent | Fitness for Use | Thematic Richness | Integration | Data Integrity | Positional Accuracy | Thematic Accuracy | Completeness |
|-----------------|-------------------|---------------------------|-------------------|--------|-----------------|-----------------|-------------------|-------------|----------------|---------------------|-------------------|--------------|
| FGGD | 5 | 4 | 2 | 1 | NA | 5 | 2 | 3 | NA | NA | 5 | 3 |
| GPW | 5 | 4 | 3 | 1 | NA | 5 | 5 | 1 | NA | NA | 5 | 5 |
| GRUMP | 5 | 4 | 5 | 1 | NA | 5 | 5 | 1 | NA | NA | 3 | 5 |
| WWDRII | 5 | 4 | 1 | 1 | NA | 5 | 5 | 3 | NA | NA | 5 | 5 |
| EOS | 5 | 4 | 1 | 0.5 | NA | 5 | 5 | 1 | NA | NA | 5 | 5 |
| Landscan | 5 | 4 | 5 | 2 | NA | 5 | 5 | 1 | NA | NA | 5 | 5 |

Table 5.25 Quality matrix for the data sub-topic: Urban and Rural Population Density and Distribution

The final results declare the dataset Landscan as the best quality dataset for the current sub-topic. The dataset GRUMP took the second place and should be taken into consideration because is a product similar to the Landscan in terms of geometry and information content. The biggest difference between the two is the population estimation process that takes into consideration different base data: the consequence is that Landscan represents better the daytime ambient population because it takes into consideration land use and transportation layers amongst others whilst GRUMP represents better the night-time population because information on night-time

light is inserted in the estimation [SEDAC website]. Therefore the suggestion is to use one of these two datasets in function of the daytime or night-time focus of the analysis.

Vector of Weights

| Indicator | Weight (0 to 1) |
|---------------------------------|-----------------|
| Geographic Extent | 1 |
| Licensing and Constraint | 0.5 |
| Scale Denominator | 1 |
| Update | 1 |
| Temporal Extent | 0 |
| Fitness For Use | 0.5 |
| Thematic Richness | 1 |
| Integration | 0.5 |
| Data Integrity | 0 |
| Positional Accuracy | 0 |
| Thematic Accuracy | 1 |
| Completeness | 1 |

Table 5.26 Weights assigned to the indicators of quality and consistency for the sub-topic: Urban and Rural Population Density and Distribution

Final Scores and Ranking

| Dataset | Final Score |
|-----------------|-------------|
| Landscan | 32 |
| GPW | 29 |
| GRUMP | 29 |
| WWDR11 | 28 |
| EOS | 26.5 |
| FGGD | 24 |

Table 5.27 Results of the quality assessment: total scores for each dataset

5.3 Conclusion and comments

The contractor designated a group of data analysts for working on the data quality assessment of the GMES RDA Lot2 project; all the data analysts had good skills in GIS and remote sensing and previous experience in managing and processing spatial data.

In particular all the data analysts were able to apply the data quality protocol, after having read the protocol documentation [ITHACA, Deliverable 11.1 M9] and some application examples. Moreover the entire data collection was assessed and results for all the datasets were obtained.

Therefore the protocol can be defined useful and successful in determining the best quality data for implementing a RDA service. Nevertheless some improvements can be applied.

Firstly, as it has been told, the data quality protocol defines procedures and thresholds aimed at assigning quality scores to the datasets; in some cases the analysts found difficulties in applying the thresholds due to their scale range and to the specific data situation. In these cases new scale ranges were proposed and applied by the data analysts themselves in order to take advantage of the full score range; the changes were made in conformity with the protocol general principles and at the entire sub-topic level allowing correctness and uniformity in the evaluation. All the cases of changes in the protocol should be studied in order to determine whether or not the protocol needs stable and definitive modifications.

Secondly the data quality assessment activities are much time and resource consuming; the application of every indicator requires a specific procedure and reading carefully the entire documentation and metadata available. The high variability of data formats both in terms of file formats and spatial types, such as raster, polygon vector, line vector and point vector, makes it difficult to generate automatic procedures. Therefore strong efforts are required for writing computer procedures able to simplify and speed up the operations of the data analysts.

Thirdly the technical procedure necessary during the quality assessment requires GIS specialists to perform the assessment itself due to the technical knowledge needed in the operations. Moreover decision making based on GIS knowledge and previous experience in mapping, spatial analysis and geo-processing is necessary in many cases. These characteristics make it difficult the application of the protocol to a non-experienced GIS user. Therefore the development of a simplified version of the quality protocol could be useful to extend the possibility of the protocol application also to less-experienced GIS user: this would require writing computer code able to perform part of the data processing and guided procedures with multiple choices for score giving.

Lastly most of the data processing was performed by mean of proprietary software, in particular *esri ArcGIS for Desktop*⁷; the use of proprietary software can be an obstacle for a third party that wants to apply the data quality protocol. Therefore the open source software available in the GIS field should be studied in order to understand its potential exploitation in alternative to the proprietary.

In conclusion the geospatial data quality protocol defined specifically for the GMES RDA Lot2 project and reported in chapter 4 of this thesis can be considered a valid instrument for every situation that requires the quality level ranking of a number of geospatial datasets grouped into categories. Although the data quality assessment concluded successfully, the protocol can be improved in order to perform the assessment faster and broaden the potential community of users of the protocol itself. Firstly the modification of some indicator thresholds should be considered for exploiting the entire score range. Secondly some scripts in a suitable programming language should be written in order to simplify and speed up the work of the analysts. Thirdly the development of guided procedures in a simplified version of the protocol can allow non-experienced GIS users to take advantage of the protocol. Finally the exploitation of open source software should be considered for broadening the community of the protocol potential users.

⁷ www.esri.com/software/arcgis/arcgis-for-desktop/index.html

6. GeoNode: a Web Application for Data Sharing

In the field of natural disasters and emergency management there is a strong interest in geospatial data sharing among different agencies or communities, with the aim of avoiding duplication of efforts and allowing a quicker operational response.

The traditional schema of activities considers the work of map producers and map users as separated. First aid agencies and other authorities act as map users and define the characteristics of the maps, while map producers apply the users' needs and release hard copies of their products or digital copies saved in blocked file formats such as *PDF*. Therefore in this schema map users don't have the possibility to modify their maps in case of need but they have to ask back the map producers even for simple modifications such as colour schemes, symbol schemes, data styling, scale factor and bounding box. Moreover map users cannot reuse the data that are part of the map for other representations nor spatial analysis unless performing an annoying operation of re-digitizing.

In very recent years, especially after the severe disaster that was the earthquake in Haiti on January 12th, 2010, humanitarian agencies, civil protection departments and other aid structures started expressing the need of having data beyond the maps, in order for them to be possible to reuse the data themselves. In that particular case there was a strong interest in damage assessment data, such as closed roads and collapsed buildings, but also base reference data such as elevation, transportation and ground cover.

At the same time initiatives for promoting the importance of geospatial data sharing in order to better face natural disasters were born. At that time most of these activities were felt as direct answers to needs input and happened mostly without a complete consciousness of the whole process. Only afterwards they were conceived, promoted and financed as proper development projects; one of them is the GFDRR Open Data for Resilience Initiative. Furthermore such an initiative was accompanied by the development of an operational tool that allows data sharing over the internet, the GeoNode.

In this chapter the GeoNode platform is introduced, together with the respective community of developers, and the OpenDRI initiative. Moreover technical details on the main functionalities

and an overview of the main software components are provided. The choice of giving space to the GeoNode software is justified by its adoption as demonstrator application of the reference data access introduced in section 3.5; furthermore it was also chosen for the data quality results publication, as described in chapter 7.

6.1 GeoNode introduction

GeoNode is an open source platform that facilitates the creation, sharing and collaborative use of geospatial data. The project aims to surpass existing spatial data infrastructure solutions by integrating robust social and cartographic tools.

At its core, the GeoNode is based on open source components GeoServer, GeoNetwork, Django, and GeoExt that provide a platform for sophisticated web browser spatial visualization and analysis. Atop this stack, the project has built a map composer and viewer, tools for analysis and reporting.

To promote collaboration, the GeoNode is designed on Web 2.0 principles to:

- make it extremely simple to share data;
- easily add comments, ratings, tags;
- connect between GeoNode and existing GIS tools.

To secure distribution, the GeoNode enables:

- simple installation and distribution;
- automatic metadata creation;
- search via catalogues and search engines, such as Google and others.

Moreover, to promote data collection, the GeoNode is aimed to align incentives to create a sustainable Spatial Data Infrastructure to:

- align efforts so that amateur, commercial, NGO and governmental creators all naturally collaborate;
- figure out workflows, tools and licenses that work to assure data quality;
- to promote living, constantly evolving, authoritative and always up to date data.

The idea is to create a full featured platform for helping decision makers easily compose and share stories told with spatial data [GeoNode website].

6.2 GeoNode functionalities

The GeoNode provides the user with a series of functionalities in the field of geo-spatial data which are typical of a desktop GIS application, such as creating and printing maps. The capacity of accessing from the web functionalities which have been given by desktop applications for many years is provided by many companies nowadays, consider for example the availability of

common office applications which are offered by Google⁸. This technical choice, also called Software as a Service, is expected to increase in the next future and to overcome the common strategy of desktop applications [Patrignani N., 2009]. The aim is to simplify the management of desktop computers, moving workload towards computer servers, and to promote sharing among users since data in this way are centralized and can be easily accessed from many computer clients even in concurrent mode.

6.2.1 Data sharing

The promotion of data sharing is among the first objectives of the community of GeoNode developers. On the one hand data sharing means that a user is put in condition of being able to upload data into the system and to give other users the possibility of exploring these data. On the other hand the user is provided with the capacity of taking advantage of data shared by others either through online tools or by downloading them locally in formats suitable for desktop applications.

6.2.2 Data uploading

Data can be uploaded by registered users through a simple web interface; with the last official release, GeoNode permits to upload geo-spatial data in the following formats:

- *geoTiff*, it is a file format for storing georeferenced raster data;
- *shapefile*, it is the *esri* format for georeferenced vector data, a standard *de facto* among the GIS community.

Figure 6.1 Screenshot of the webpage designed for data uploading, when a vector file is being uploaded. Note that permissions on data viewing and editing can be assigned to users and groups by means of the commands implemented on the right-hand side of the page

Moreover for each data the web interface permits also to upload the description of the style to be used during graphical rendering of the data themselves, encoded into an open format which is called *SLD*, that is *Styled Layer Descriptor*. Furthermore during the upload process data are

⁸ <http://support.google.com/docs/bin/answer.py?hl=en&answer=49008>

converted and then stored into the server in formats suitable for efficient storing and future retrieving: this step is invisible to the user. A screenshot of the web page developed for data uploading is showed in Figure 6.1.

6.2.3 Data downloading

Data can be found on the platform by exploiting the embedded search engine, which allows to combine classical search on a keyword basis with spatial search, by graphically defining a bounding box on a computer screen.

Once interesting data have been found, the user is prompted to the so called *data page*: it is a web page where data are graphically rendered in a frame, over a base layer that is retrieved dynamically from available free web services, for example Google Maps, Open Street Map and NASA Blue Marble. This frame contains also tools for zooming and panning the spatial content as well as a querying tool for data semantic content.

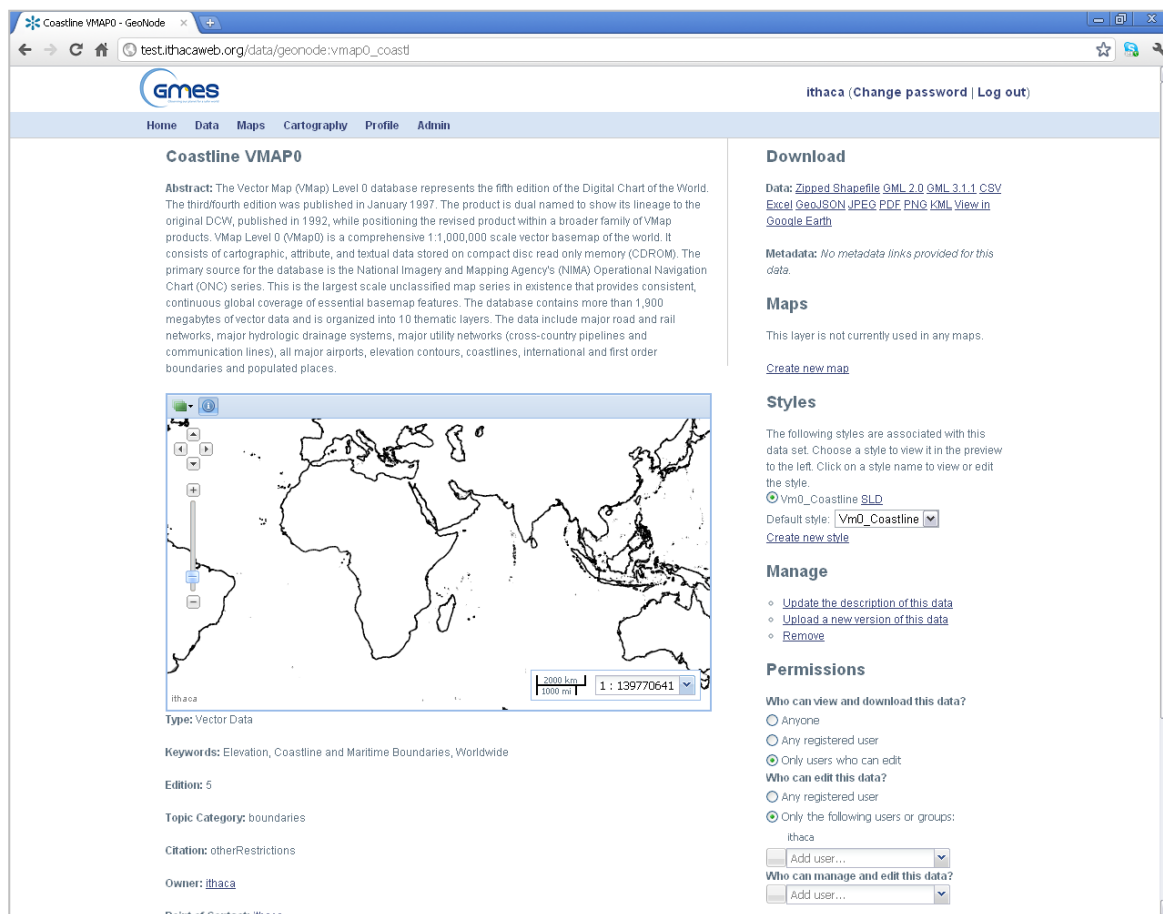


Figure 6.2 Screenshot of the GeoNode data page for data representing the coastlines and produced along with the VMAP0 project.

The graphical rendering is devolved upon GeoServer, that is a Web Mapping Server able to serve data and maps in an efficient and powerful way and can convert data into different formats upon

request. The user can take advantage of this technology and download data by simply clicking on the specific link on the right-hand side of the *data page*. With the last release of the GeoNode raster data can be downloaded as *geoTiff*, *JPEG*, *PDF*, *PNG*, *KML*. Instead vector data can be downloaded as *shapefile*, *GML*, *CSW*, *Excel*, *GeoJSON*, *JPEG*, *PDF*, *PNG*, *KML*.

Other interesting elements displayed in the *data page* are the metadata, that are showed partially on top and partially under the above mentioned frame, and the style manager application that allows to change the way data are displayed. Furthermore, if the current user has been given the required permissions by the data owner, several links for data management and permission assignment are displayed on the right-hand side. A screenshot of the mentioned *data page* is showed in Figure 6.2.

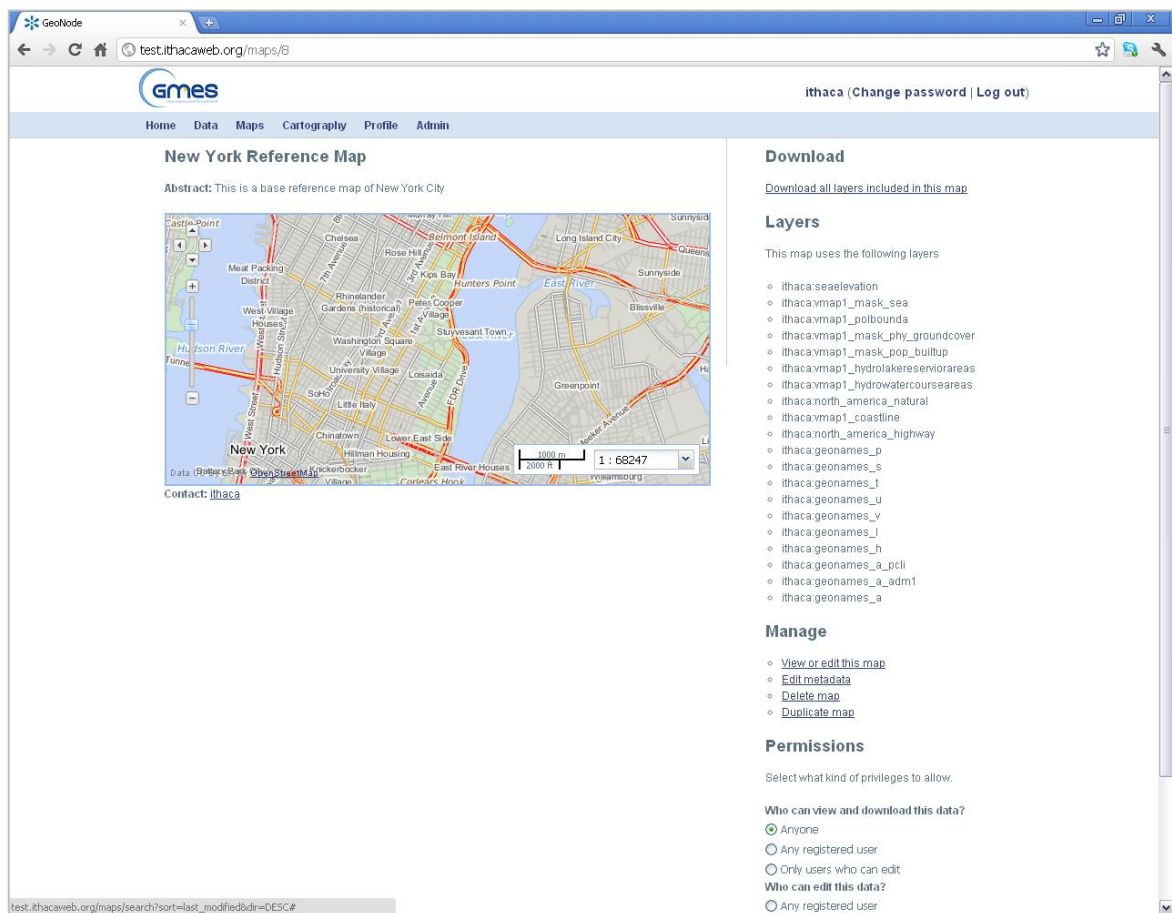


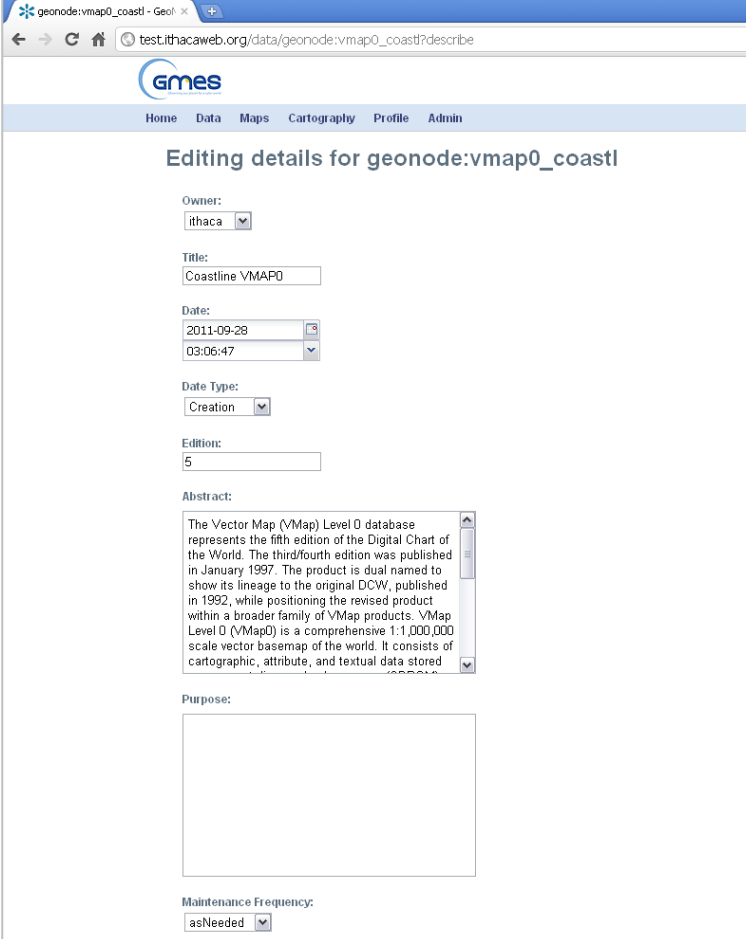
Figure 6.3 Screenshot of the GeoNode map page: it contains preview of the map with the respective title and abstract on the left-hand side of the page; while the list of layers comprehended in the map and some management tools are implemented on the right-hand side

6.2.4 Map creating

Registered users can create web maps by collecting together a number of data of interest and styling them in the favourite way. Maps in GeoNode are ordered collection of layers which are rendered to a protocol suitable for visualization in a web browser. The web interface for map creation make this operation similar to the most common desktop GIS applications: its name is

GeoExplorer. Once a map is created it can be saved and stored into the platform, in order to be easily retrieved, explored and potentially modified in the future. Maps can also be easily published on blogs and other websites and they can be printed to a *pdf* format.

Moreover maps existent on the platform can be searched by means of a search engine based on keywords; when a map of interest is found, the user is prompted to the so called *map page*. This is a web page that shows a preview of the map itself in a frame on the left-hand side of the screen; on top of it the map title and an abstract are displayed. Furthermore the list of layers contained in the map as well as the link for data downloading are displayed on the right-hand side of the screen. Finally, similarly to what happens with data, special commands for map management and permission assignment are showed in case the current user has been enabled by the map owner. A screenshot of the GeoNode *map page*, that represents New York City map, is reported in Figure 6.3.



The screenshot shows a web browser window with the URL `test.ithacaweb.org/data/geonode:vmap0_coastl?describe`. The page header features the Gmes logo and navigation links: Home, Data, Maps, Cartography, Profile, Admin. The main heading is "Editing details for geonode:vmap0_coastl". The form contains the following fields:

- Owner:
- Title:
- Date:
- Date Type:
- Edition:
- Abstract:
- Purpose:
- Maintenance Frequency:

Figure 6.4 Screenshot of the GeoNode online tool for metadata editing

6.2.5 Metadata management

When uploading data the user is prompted to fill in a metadata form that will be showed to the users who will look for those data in the future. The same form can be used in every following time to update the metadata content in case of need: a screenshot of the on-line form for metadata editing is contained in Figure 6.4.

The last release of GeoNode can manage a metadata form which contains a subset of the collection of fields suggested in the ISO standards for geo-spatial data. The metadata form can be developed in order to match the requirements defined in INSPIRE or other requirements specifically defined for this project. Moreover metadata can also be developed for maps.

6.2.6 Data sensitivity

Data sensitivity is a key issue in every environment that involves data exchange on a web platform. It has to be guaranteed especially in case of spatial data related to emergency management.

The GeoNode implements a framework of users, groups and permissions typical of a DBMS; this is due to the Django framework that ensures fast, direct and reliable reading and writing to a DBMS which runs in support of the web platform.

Every user can access and potentially edit only data which he's allowed to see and manage. On the other hand the user who uploads data can assign permissions on his data to other users and groups. The system administrator has the possibility and the responsibility of making changes to the permissions in case of need.

6.3 GeoNode and the OpenDRI initiative

In order to build resilient societies, policy-makers and the public must have access to the right data and information to inform good decisions. Examples of such decisions include where and how to build safer schools, how to insure farmers against drought, and how to protect coastal cities against future climate impacts. Sharing data and creating open systems promotes transparency, accountability, and ensures a wide range of actors are able to participate in the challenge of building resilience.

The Open Data for Resilience Initiative (OpenDRI) aims to reduce the impact of disasters by empowering decisions-makers with better information and the tools to support their decisions.

OpenDRI implements the policy recommendation of the World Bank which establishes the importance of data sharing to reduce vulnerability to disasters. OpenDRI is currently implementing these ideas in 25 countries around the world to improve disaster and climate change resiliency [GFDRR website: OpenDRI].

GeoNode is the operational platform that supports the implementation of the OpenDRI initiative; it is currently being adopted in many developing countries at risk from natural disasters for data sharing among various agencies, scientific and governmental institutions at a country level. This action has two advantages: on one hand duplication of efforts is avoided in data collection, validation and management, while on the other hand high reliability is given to the data

since the responsibility of data related to specific areas of information is assigned to experts in the respective sector.

Examples of OpenDRI in action can be seen on several Web platforms running: one of them is the *Open Data for the Horn*⁹ platform, that was set up in order to better answer to the drought crisis opened in 2011 in the countries of the Horn of Africa. World Bank, the UN WFP, ITHACA and other agencies collaborated to the platform implementation in order to inform for better decisions.

Another example is the *haitidata.org* platform that collects data produced following the 2010 Haiti earthquake and makes them available for anyone to download and use. The platform enables individuals and organizations to contribute data and download data, thereby enabling people to collaborate and utilize these data required to rebuild Haiti.

A technical proposal, written in response to a call for tenders promoted by GFDRR, regarding the implementation of the OpenDRI initiative in the country of Malawi is contained in the Annex B of this thesis.

6.4 GeoNode components

When the GeoNode initiative was conceived other useful projects in the field of geo-spatial data were already been developed and were in production serving many companies around the world. On one hand several companies had been developing software for decades in geospatial data management, with proprietary formats and procedures. On the other hand the Open Geospatial Consortium had been defining several open standards for geospatial data for years, on top of which open source softwares both on the desktop and on the server side had been developed by developer communities and software houses. The GeoNode initiative builds upon the useful open source softwares already developed and combines them together in order to take advantage of their functionalities in an integrated manner.

In particular the GeoNode is made of the following components:

- PostgreSQL;
- PostGIS;
- GeoServer;
- GeoNetwork;
- GeoExplorer;
- OpenLayers.

In the following sections the mentioned components are introduced and their role in the platform is described.

⁹ <http://horn.rcmrd.org/>

6.4.1 PostgreSQL and PostGIS

PostgreSQL is a free and open source object-relational database management system available for many computer platforms. It is developed by a global development group, consisting of a handful of community volunteers employed and supervised by large companies in the software development field. It implements the majority of the *SQL:2008* standard, is fully transactional, includes all *Data Definition Language* statements, has extensible data types, operators and indexes, and has a large number of extensions written by third parties [Wikipedia : PostgreSQL].

PostGIS is an extension to PostgreSQL, developed by a spatial database technology research project, which allows GIS objects to be stored into the database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems. It includes support for *GiST-based R-Tree* spatial indexes and functions for analysis and processing of GIS objects. Last development have added user interface tools, basic topology support, data validation, coordinate transformation and programming among other functionalities [PostGIS website : Home].

In GeoNode PostgreSQL is used with a double function: on one side it works in support of the GeoNode website, being the place where all the information about users and groups is stored and the instrument that assure the adherence to the user permissions. Moreover it contains also all the information that regards specifically the dynamic content of the website such as sentence translations in various languages. On the other side it works as data repository for geospatial data by mean of the PostGIS extension, in particular for vector data. Relevant advantages can be detected when storing vector data in PostGIS, rather than storing them in filesystem, both in terms of performances in data selection and loading and in data management. On the contrary, even if the last PostGIS version allows raster storage, there isn't a relevant advantage for raster data, therefore they are stored in the filesystem.

6.4.2 GeoServer

GeoServer is an open source software server written in Java that allows users to publish geospatial data. It is designed for interoperability and publishes data from any major spatial data source using open standards. Being a community-driven project, GeoServer is developed, tested, and supported by a diverse group of individuals and organizations from around the world. Furthermore it is the reference implementation of the Open Geospatial Consortium Web Feature Service (WFS) and Web Coverage Service (WCS) standards, as well as a high performance certified compliant Web Map Service (WMS). [GeoServer website : What is GeoServer]: these standards are introduced in section 6.4.3. GeoServer can be controlled remotely by mean of a web browser: in Figure 6.5 the index page of the administration site is contained.

GeoServer is the map engine in support of the GeoNode, it works on the server side of the system and performs three main activities:

- it renders all the geospatial content shown by the GeoNode in a graphic format, in particular in the data page, in the map page and in the application that makes it available the map creation and editing, called GeoExplorer;
- it works as linkage between the geospatial data sources and the GeoNode layers, indeed it reads the content of the DBMS provided through the PostGIS extension and projects it

into a graphic format, in order to allow the GeoNode users to see the graphic content of the data and not their actual format;

- it performs all the data format transformations when it comes to data downloading in formats that are different from the original PostGIS source. When performing these operations it takes advantage of libraries specifically developed and collected together in the so called GeoTools¹⁰.

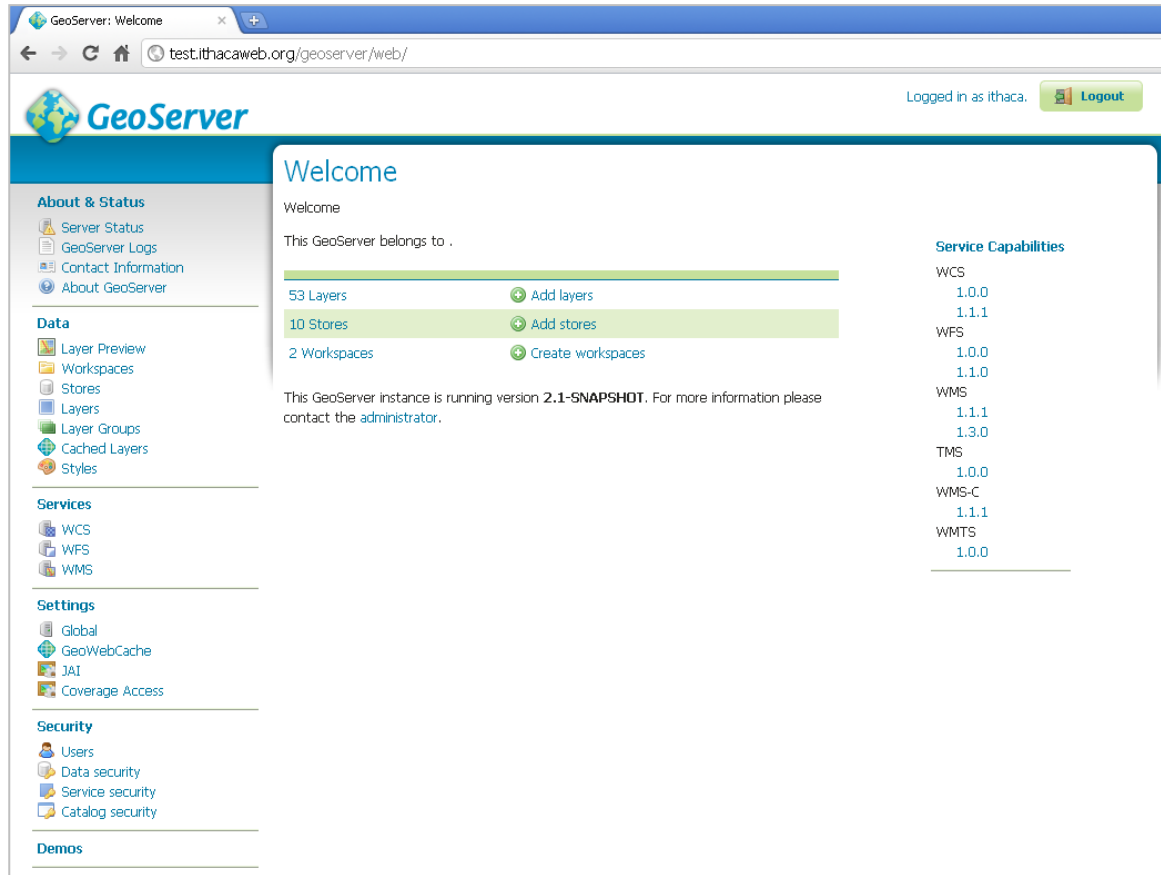


Figure 6.5 Index page of the GeoServer administration site; the software can be controlled from remote by mean of a web browser

6.4.3 Open Geospatial Consortium and OGC standards

The Open Geospatial Consortium is a consortium of over 390 companies, NGOs, research organizations, agencies and universities with a common vision where everyone can take advantage of geographic information and services made available across any network, application, or platform. The mission of the OGC is to promote the development and use of advanced open systems standards and techniques in the area of geoprocessing and related information

¹⁰ <http://geotools.org/>

technologies. OGC is supported by Consortium membership fees and, to a lesser extent, development partnerships and publicly funded cooperative programs.

Moreover OGC manages a global consensus process aimed at resulting in approved interface and encoding standards that enable interoperability among and between diverse geospatial data stores, services, and applications. Therefore systems implementing OGC standards can interoperate, whether those systems are running on the same computer or the same network or are distributed across the globe. Furthermore OGC standards provide essential infrastructure for the Spatial Web, a network of geospatial resources that is thoroughly integrated into Web [OGC Website : FAQs].

Three important standards are implemented by GeoServer: WMS, WFS and WCS. In the following sections these standards are introduced.

Web Mapping Service

The Web Map Service Interface Standard provides a simple *HTTP* interface for requesting geo-registered map images from one or more distributed geospatial databases. A WMS request defines the geographic layers and area of interest to be processed. The response to the request is one or more geo-registered map images, returned as *JPEG*, *PNG* and others, that can be displayed in a browser application. The interface also supports the ability to specify whether the returned images should be transparent so that layers from multiple servers can be combined or not.

Furthermore a WMS produces maps of spatially referenced data dynamically from geographic information. It defines a map to be a portrayal of geographic information as a digital image file suitable for display on a computer screen. It should be noted that a map is not the data itself, therefore what is published through a WMS is a representation of the data from which the data cannot be derived.

Moreover a WMS instance publishes its ability to produce maps rather than its ability to access specific data holdings and classifies its geographic information holdings into *layers*; it also offers a finite number of predefined styles in which to display those layers [OGC, 2006].

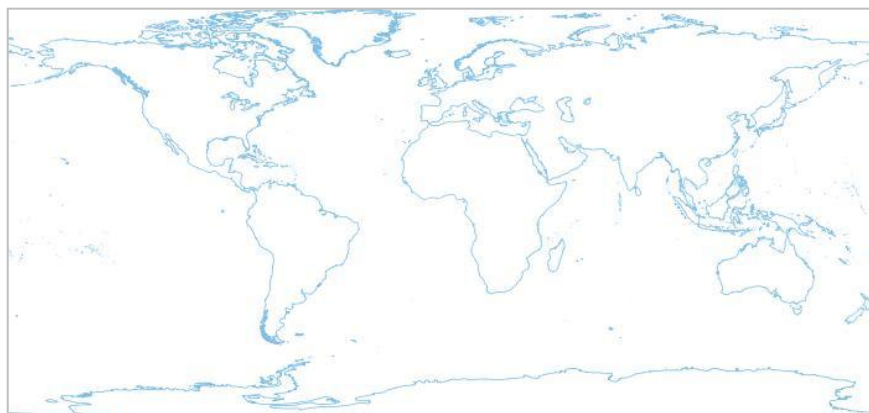


Figure 6.6 Response of the GeoServer instance set up for the GMES RDA Lot 2 project when a layer representing the VMAP0 coastline is requested in a JPG file format through a web browser

WMS operations can be invoked using a standard web browser by submitting requests in the form of *URLs*; the content of such *URLs* depends on which operation is requested. In particular, when requesting a map the *URL* indicates what information is to be shown on the map, what portion of the Earth is to be mapped, the desired coordinate reference system, and the output image width and height. When two or more maps are produced with the same geographic parameters and output size, the results can be accurately overlaid to produce a composite map. For instance in Figure 6.6 the image response of the GeoServer instance set up for the GMES RDA Lot 2 project is reported, when submitting a request through the following URL:

http://test.ithacaweb.org/geoserver/geonode/wms?service=WMS&version=1.1.0&request=GetMap&layers=geonode:vmap0_coastl&styles=&bbox=-180.0,-85.583,180.0,83.627&width=702&height=330&srs=EPSG:4326&format=image%2Fjpeg

Web Feature Service

The Web Feature Service (WFS) represents a change in the way geographic information is created, modified and exchanged on the Internet. Rather than sharing geographic information at the file level using FTP, for example, the WFS offers direct fine-grained access to geographic information at the feature and feature property level. Web feature services allow clients to only retrieve or modify the data they are seeking, rather than retrieving a file that contains the data they are seeking and possibly much more.

Moreover WFS specifies standards in several operations such as discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions. Discovery operations allow the service to be interrogated to determine its capabilities and to retrieve the application schema that defines the feature types that the service offers. Query operations allow features or values of feature properties to be retrieved from the underlying data store based upon constraints, defined by the client, on feature properties. Locking operations allow exclusive access to features for the purpose of modifying or deleting features. Transaction operations allow features to be created, changed, replaced and deleted from the underlying data store. Stored query operations allow clients to create, drop, list and describes parameterized query expressions that are stored by the server and can be repeatedly invoked using different parameter values [OGC, 2010].

Web Coverage Service

The OGC Web Coverage Service (WCS) supports electronic retrieval of geospatial data as coverages: a coverage is a feature that associates positions within a bounded space, that is its spatiotemporal domain to feature attribute values, that is its range. GIS coverages, including the special case of Earth images, are two-or higher-dimensional metaphors for phenomena found on or near a portion of the Earth's surface. A coverage can consist of a set of features or feature collections; Earth images are seen as Grid Coverages that contain features whose geometries are of type "set of cells" or "set of pixels" [OGC - Topic 6, 2006].

A WCS provides access to coverage data in forms that are useful for client-side rendering, as input into scientific models and for other clients. The WCS may be compared to the WMS, WFS service; as these service instances, a WCS allows clients to choose portions of a server information holdings based on spatial constraints and other query criteria.

Unlike WMS, which portrays spatial data to return static maps, rendered as pictures by the server, the WCS provides available data together with their detailed descriptions; defines a rich syntax for requests against these data; and returns data with its original semantics which may be interpreted, and extrapolated, other than portrayed.

Unlike WFS, which returns discrete geospatial features, the WCS returns coverages representing space and time-varying phenomena that relate a spatio-temporal domain to a range of properties [OGC WCS - Core, 2010].

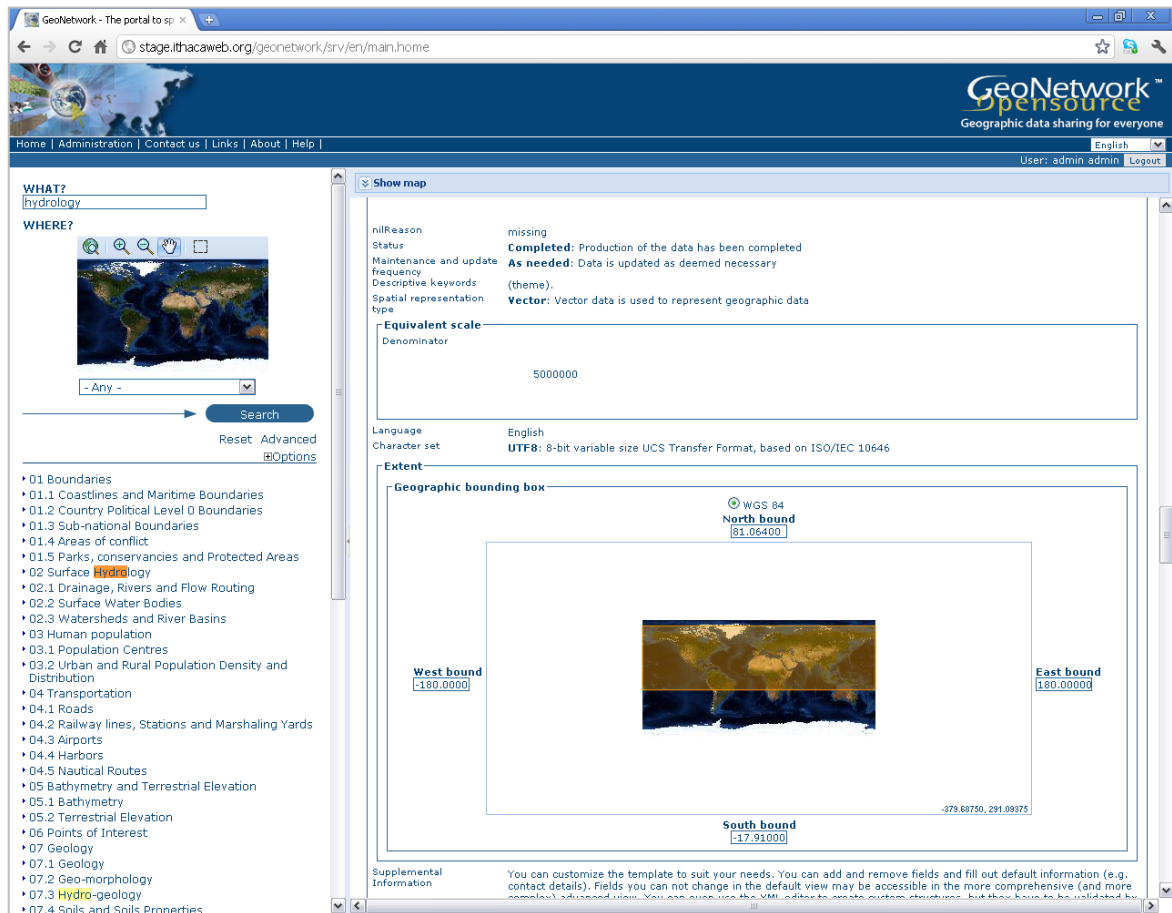


Figure 6.7 Screenshot of an active instance of GeoNetwork OpenSource that shows the metadata of a surface hydrology dataset

6.4.4 GeoNetwork

GeoNetwork OpenSource is a standard based and decentralised spatial information management system, designed to enable access to geo-referenced databases and cartographic products from a variety of data providers through descriptive metadata, enhancing the spatial information exchange and sharing between organisations and their audience, using the capacities and the power of the Internet. The system provides a broad community of users with easy and timely access to available spatial data and thematic maps from multidisciplinary sources, that may in the

end support informed decision making. The main goal of the software is to increase collaboration within and between organisations for reducing duplication and enhancing information consistency and quality and to improve the accessibility of a wide variety of geographic information along with the associated information, organised and documented in a standard and consistent way. A screenshot of an archive instance of GeoNetwork Opensource is contained in Figure 6.7.

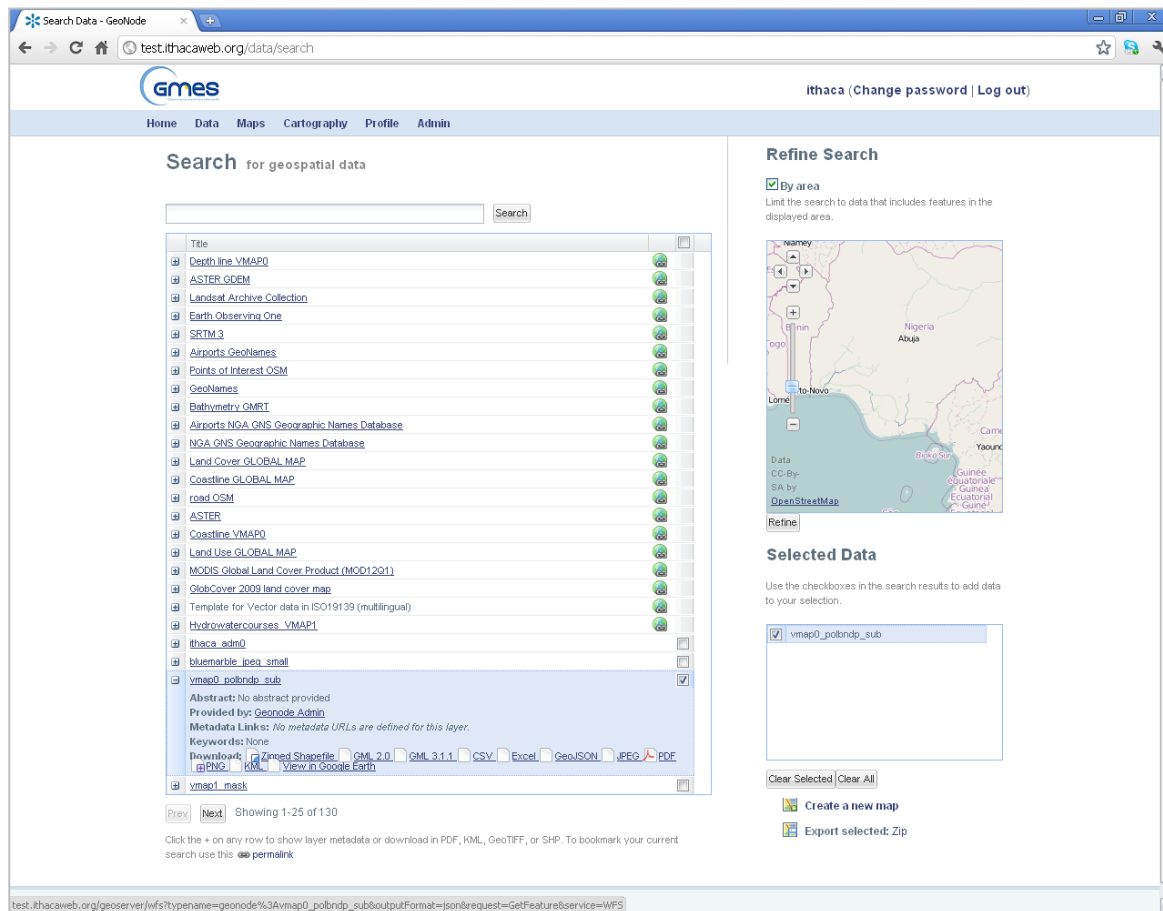


Figure 6.8 Screenshot taken from the GeoNode instance set up specifically for the GMES RDA Lot 2 project, that shows the page containing the two search engines; on the left the search engine based on keywords is shown, while on the right the search engine based on data spatial location is located

The main GeoNetwork features are:

- instant search on local and distributed geospatial catalogues;
- uploading and downloading of data, documents, *PDFs* and any other content;
- an interactive Web map viewer that combines WMS services from distributed servers around the world;
- online map layout generation and export in *PDF* format;
- online metadata editing with a powerful template system;

- scheduled metadata harvesting and synchronisation between distributed catalogues;
- groups and users management;
- fine grained access control.

The first prototype of the GeoNetwork catalogue was developed by the Food and Agriculture organisation of the United Nations in 2001 to systematically archive and publish the geographic datasets produced within the organisation. GeoNetwork Opensource is the result of the collaborative development of many contributors. It is compliant with the metadata standard ISO19115:2003, using the ISO19139:2007 implementation specification schema published in May 2007. The release also serves as the open source reference implementation of the OGC Catalogue Service for the Web (CSW 2.0.2) specification [GeoNetwork Opensource, 2012].

Specifically the GeoNode takes advantage of two main functionalities that are implemented in GeoNetwork: the two search engines. The first is a typical web search engine and works on the basis of some keywords that are assigned to each dataset and then retrieved when submitted in the search string. The second is a spatial search engine by mean of which it is possible to search for data covering a determined area of interest, on the basis of its geographic extent. An example screenshot of the two search engines, implemented into a GeoNode page, is contained in Figure 6.8.

6.4.5 GeoExplorer

One of the key GeoNode functionalities is the possibility to create and edit Web maps, by taking advantage via a Web browser of some capabilities typically offered by desktop GIS applications. This is allowed thanks to an open source application called GeoExplorer, developed for this purpose by an American software house called OpenGeo [OpenGeo website].

The goal of GeoExplorer is to make it easy for anyone to assemble a browser based mapping application; it includes basic layer browsing capabilities and can be placed in front of any compliant WMS. GeoExplorer takes advantage of the GeoExt toolkit and of the user interface power of ExtJS [OpenGeo website : GeoExplorer preview].

In particular, from the user perspective, GeoExplorer is composed by a large map canvas where geospatial data are rendered as maps and a table of content on the left-hand side of the screen where the list of layers present in the map is showed and where tools for controlling each layer representation and styling are embedded; furthermore a set of tools, aimed at controlling the zoom and pan functionality as well as the printing and publishing modules and a get-info function, is integrated in buttons contained in the upper part of the window. An example of the GeoExplorer user interface is showed in Figure 6.9. The GeoExplorer can access all the layers stored in the PostGIS database connected to the same GeoNode instance; moreover it can access every layer provided through WMS services by a network-connected server.

GeoExt

GeoExt is an open source Javascript toolkit aimed at building rich web mapping applications. It brings together the geospatial know how of OpenLayers with the user interface savvy of ExtJS to help build powerful desktop style GIS apps on the web with JavaScript.

On one hand GeoExt extends Ext JS, a rich library of web UI widgets and helper classes: in particular it is a Javascript framework for building rich application in web browsers, that enables developers to use a model-view-controller architecture when building their application, and offers a wide range of user interface widgets such as scalable grids, trees and menus [Sencha website].

On the other hand the OpenLayers mapping library is the key component of GeoExt, performing the core map-related functions of every GeoExt-based application [GeoExt website : Documentation, 2012].

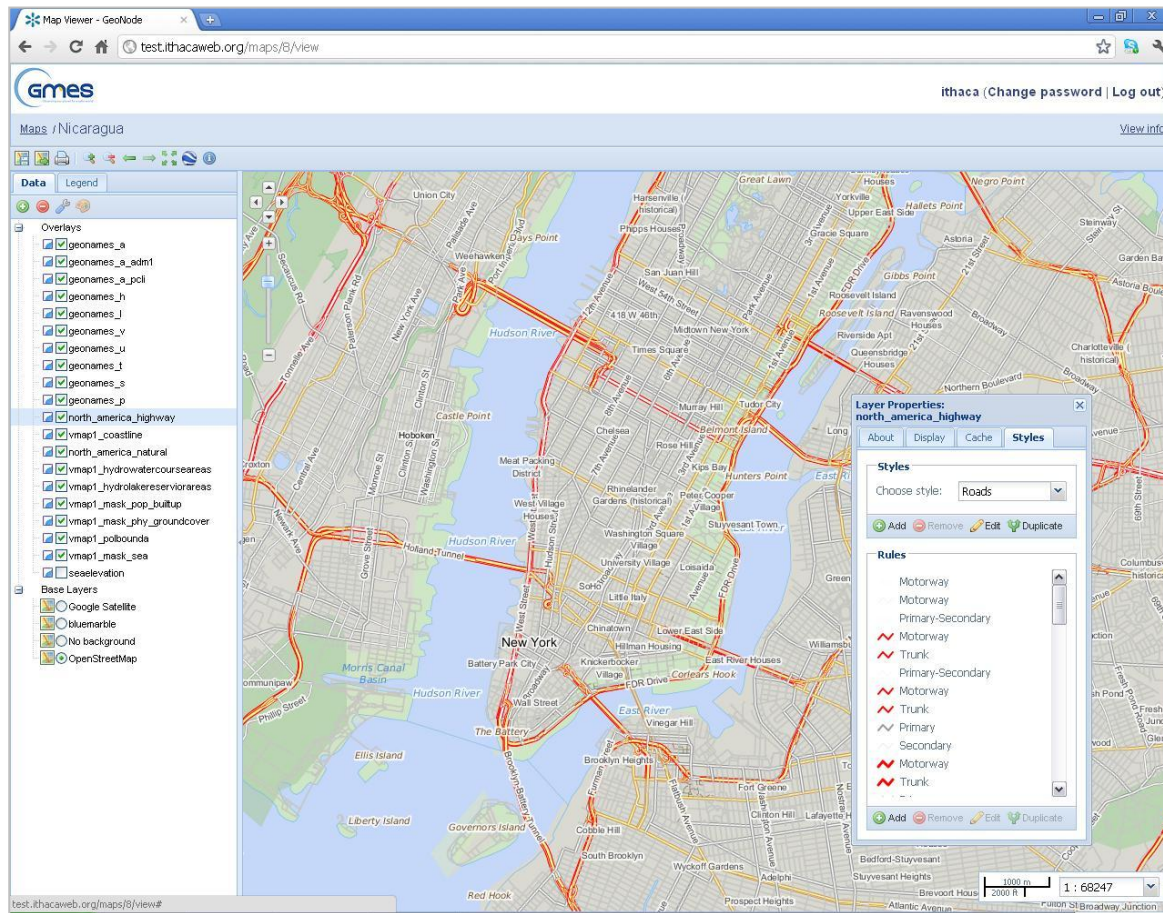


Figure 6.9 A GeoExplorer screenshot shows the main components: on the right is a map canvas where geospatial data are displayed; on the left a table of contents allows the layer management; on the top buttons implement functions related to zoom, pan, printing and saving. The smaller window on the screen allows to manage layer properties such as the representation style

6.4.6 OpenLayers

OpenLayers is a pure JavaScript library for displaying map data in most modern web browsers; it hasn't any server-side dependencies, and is developed and supported by a number of organizations around the world. In particular it implements a JavaScript API for building rich web-based geographic applications and is completely free and open source [OpenLayers website].

OpenLayers has been developed to further the use of geographic information of all kinds; it makes it easy to put a dynamic map in any web page and can display map tiles and markers loaded from any source.

OpenLayers is used in GeoNode on the client side, in particular in the web pages that show cartographic representation of data and maps; in respect to what said previously and in section 6.2, OpenLayers is used in the GeoNode data and map page and in GeoExplorer that is the map builder application. Examples of OpenLayers functions are the zoom and map options in the map visualization.

6.4.7 Django

As stated previously, GeoNode combines together many software products which are already existing and sufficiently stable to be used by a wide community of users. Being a web platform the main GeoNode component is the code that describes the website logic. This code is written in Python programming language and takes advantage of a framework for web development called Django.

The DRY principle

Django encourages rapid development and clean, pragmatic design and applies the DRY principle. Duplication, both inadvertent and purposeful, can lead to maintenance difficulties, poor factoring and logical contradictions. It can arise in architecture, requirements, code or documentation; with effects that can range from mis-implemented code and developer confusion to complete system failure. The DRY principle, that is an acronym of Don't Repeat Yourself states that “every piece of knowledge must have a single, unambiguous, authoritative representation within a system”. A mechanical, textual duplication is accepted as long as the authoritative source is well known; the user tries to identify the single, definitive source of every piece of knowledge used in your system and then use that source to generate applicable instances of that knowledge, that can be code, documentation and tests [c2 wiki].

The MVC pattern and respective modules

Moreover Django adheres to the MVC software architecture, that is the acronym of Model-View-Controller, an architectural pattern used in software engineering that isolates the application logic for the user from the user interface, permitting independent development, testing and maintenance of each part.

The model manages the behaviour and data of the application domain, responds to requests for information about its state from the view and responds to instructions to change state from the controller. A model is the single, definitive source of data about the data contained into a web site; it contains the essential fields and behaviours of the data stored into the DBMS that supports the website.

The view renders the model into a form suitable for interaction, that is a user interface element; multiple views can exist for a single model, that responds to different purposes.

The controller receives user input and initiates a response by making calls on model objects; a controller accepts input from the user and instructs the model and a view port to perform actions based on that input.

An example of control flow is the following [Wikipedia : Model-view-controller]:

1. the user interacts with the user interface in some way, typically by pressing a button or issuing a command;
2. the controller handles the input event from the user interface and converts the event into an appropriate user action, understandable for the model;
3. the controller notifies the model of the user action, possibly resulting in a change in the model state;
4. a view queries the model in order to generate an appropriate user interface. The view gets its own data from the model;
5. the user interface waits for further user interactions, which restarts the control from cycle.

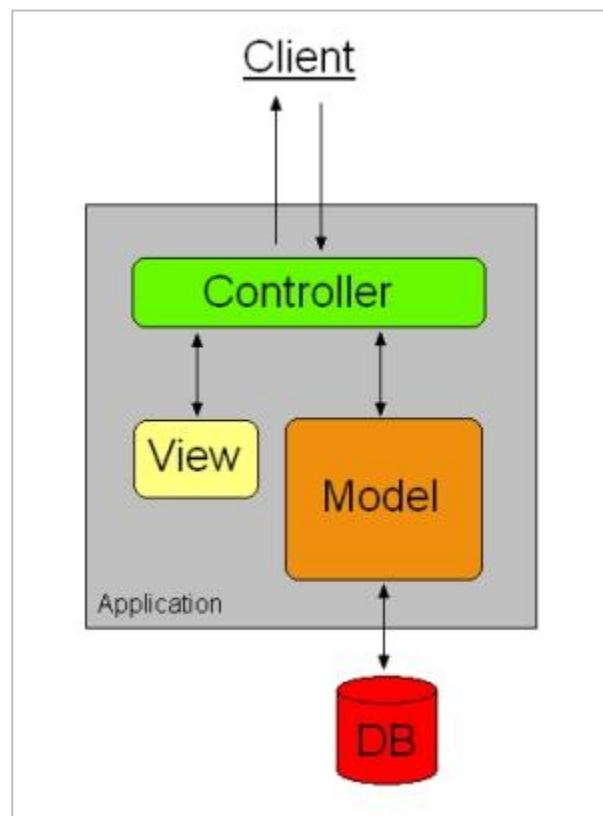


Figure 6.10 Logical schema of the Model-View-Controller architectural pattern for software development (source: Wikipedia)

The implementation of the MVC model in Django is realized in three Python modules that are automatically generated by the software when a new application is issued: `models.py` and `views.py` contain the homonymous components, while the controller is implemented into the template files. These are template web pages written partially with HTML language and partially with

Python code: when the template is called by a view, the missing part of the HTML itself is generated by the Python code and then the web page is ready for being sent to the client.

As previously mentioned, Django provides support and API for managing a DBMS without the need of dealing directly with SQL language. Many different DBMS can be used, two examples of which are *PostgreSQL* and *sqlite*.

Important Django modules

Other two key Django modules are the *settings* and the *URLconf* module. The first contains all the important configuration of the Web site and the Django installation. Since it is a Python module, the configurations are expressed as variables at the module level and it must not contains syntax errors. Among the variables, some of the most important are:

- *DEBUG*: it is the condition of the site whether in debugging mode or production mode; if in debugging mode the software communicates the stack trace of the errors when in site development phase;
- *SITEURL*: the *URL* of the website we are working on;
- *ADMINS*: it contains the website administrator name;
- *DATABASE_ENGINE*, *DATABASE_NAME*, *DATABASE_USER*, *DATABASE_PASSWORD*, *DATABASE_HOST*, *DATABASE_PORT*: are a set of variables that describe the DBMS in support to the website and the keywords to get data from and to it, such as name, user, password, connection host and connection port;
- *TIME_ZONE*: it is the time zone of reference of the website;
- *MEDIA*: describes the path that holds the media component;
- *TEMPLATE_DIRS*: the path to the template directory;
- *INSTALLED_APPS*: the list of the Django applications installed;
- *GEOSERVER_BASE_URL*: specific of the GeoNode platform, it defines the *URL* followed by Django in order to interoperate with the GeoServer;
- *GEONETWORK_BASE_URL*: specific of the GeoNode platform, it defines the *URL* followed by Django in order to interoperate with the GeoNetwork.

The *settings* module contains also the path of the *URLconf* module: this contains the *URL* configuration and is responsible for correctly linking the requests submitted by the client machine against the functions that have to be run by the server machine. It is a Python module that contains a simple mapping between *URL* patterns to Python callback functions: that mapping is stored in a variable called *urlpatterns*, in the form of a list of tuples. The *URL* patterns are stored as regular expressions while the callback functions are Python methods stored in the views.

When a user types an *URL* address into the address bar of a Web browser, he sends a request to the server. In case the respective Web site is powered by Django, it loads the Python module that contains the variable *urlpatterns*, runs through each *URL* pattern following the order and stops at the first one that matches the request. Once the regular expression has matched, the view contained in the tuple is called and the Python function executed. The view gets passed a request

for returning an *HTTP* response as first argument and any values captured in the regular expression as remaining arguments.

To sum up, the Django framework can be considered as the core of the GeoNode platform because it is responsible for keeping together all the other components and allows them to be interoperable and efficient. Furthermore it is the engine of the entire Web platform because it supports all the operations called by the users from remote. Therefore the Python programming language and the Django framework are the instruments of the GeoNode development: they are a requirement for every modification, improvement, configuration and customization of the Web platform itself.

7. GeoNode Development for Quality Assessment Results

In chapter 4 a protocol for the quality assessment of geospatial data was proposed and in chapter 5 some examples of the protocol application were introduced. The application of the protocol on all the datasets collected during the inventory phase of the GMES RDA Lot 2 project was performed by a team of data analysts, due to its nature of time consuming activity; for each dataset the results of the data quality assessment step consist in a list of scores given to each *Quality Indicator*. They were used for choosing the best datasets to be part of the initial Reference Data Access that is the final product of the project.

The Reference Data Access in the current configuration contains more than one dataset for each of the data categories defined in section 3.1.1; therefore if a hypothetic user needs a dataset for producing a map, developing a WebGIS application or perform a geospatial analysis, he can choose among various candidates. The results of the data quality assessment can be crucial for a conscious choice of the best dataset in relationship with his needs. The key issue of the present chapter is the distribution of these results in the most profitable way to the user.

7.1 Development goals

The main goal of the activities described in this chapter is the distribution of the data quality assessment results. In order to facilitate the process, the distribution should be provided on the same medium on which the geospatial data are accessed. Other than pure information, the user should be provided with tools for data choosing on the basis of this information.

In the following section the development of the medium for data sharing, performed in the framework of the GMES RDA Lot 2 project and in relation to the data quality assessment is thoroughly described. In particular the development has been addressed in order to meet the following two main goals:

- the publication of the quality assessment results in a suitable format in order to be associated with the respective dataset and easily accessible by the end users from the Web;

- the implementation of a tool for extracting the most suitable dataset for the end user on a quality basis among the datasets contained in a sub-topic, with the same approach described in the data quality protocol of section 4.4. Such a tool should be interactive and allows the user to define both the sub-topic of interest and his needs, by filling the *Vector of Weights*, introduced in section 4.4.2.

7.2 Medium choice

The Web is considered one of the best media for geospatial data sharing and it has been defined as a requirement for the distribution of quality information. Another key for the medium choice was that it had to be comprehended among the software adopted as demonstrators in the GMES RDA Lot 2 project.

In particular the GeoNode was chosen due to the rise of recent investments in the development, the adoption on various countries and agencies around the world, the efforts of the community in writing documentation and providing help in various forms. Moreover it is addressed towards a large community of users, including the GIS illiterate people, and has all the advantages of the Free and Open Source Software, not last the easiness in development and the possibility of releasing the results of the development back to the community.

The main characteristics of an open source software is that the source code of the software itself is distributed together with it and can be freely modified in order to add or change any of its functionality and behaviour. This is particularly true for every software written in Python code, since it is an interpreted programming language and the software is always distributed as source code; instead software written in other programming languages, such as Java for example, is distributed as compiled binaries that cannot be immediately modified without the source code.

7.3 Development approach

As a general requirement the developer must have a high degree of confidence with the software to be developed and the scope of the development has to be defined clearly by the clients or in some cases by the developer itself.

The technique used for the development consists in two main steps; the first one is the investigation of the modules and the lines of code that define the behaviour of the software itself that has to be modified. As described earlier in details(see section 6.4.7), every Web site powered by Django produces a response to a client machine when it receives a request in the form of a URL; the reception of the request itself in practice produces the execution of a Python function and the return of a response in the HTTP protocol. The developer must identify which are the URLs in the URLconf module and the function in the views that produces the HTTP response that needs to be changed or implemented.

As a second step the lines of code in the *URL* configuration and in the Python function have to be modified in a way that produces the effects described in the specification previously defined. The development must respect the DRY principle specified in chapter 6: only the essential information have to be put in the code and any repetition of information must be avoided.

As a matter of fact the steps of the development require that the software is installed in the machine used for the development: in details best practices suggest to install a particular instance of the GeoNode, when used for development purposes. This instance requires a lower amount of resources than the production one and runs on a so called *virtual environment* that allows the developer to take advantage of a series of environmental variable settings without the need of set this variable system-wide and of a series of dependency frameworks without the need of installing them on the machine.

In practice the code is usually modified with a simple text editor, rarely an integrated development environment is used. Testing the development effects is made by running a Web browser on the machine and pointing in the address bar to the same machine, *localhost*, followed by the rest of the URL of interest. If any error occurs, the response of the Django framework is a debug HTML page that contains traceback of the error in the Python modules in order to easily identify and correct the code syntax; this behaviour is followed only when the variable called *DEBUG* is set to *True* in the Django settings. When the Web site is put in production that variable has to be set to *False* in order to avoid intrusion in the server by potential malicious crackers.

Errors in the HTML code produced and sent to the client can be explored with the use of software specifically devoted to the analysis of the HTML and *css* content of Web-sites. An example of this kind of software is the open source Firebug¹¹, that is distributed as an extension of the FireFox open source Web browser.

The operations described above constitute the simplest approach to the GeoNode development: thus an effective way for reaching the objectives is guaranteed, but the fact that a community of GeoNode developers exists and that it continues to develop new solutions and functionalities is not taken into account. In practice day after day new lines of code are written, many are deleted and others are updated. There is a high risk that after a period of time has passed, the newest version of the GeoNode doesn't work with the personal development; the result is the need of adjusting the personal development or re-writing it in order to match with the last available release of the official repository.

In order to avoid this consequence another development approach is suggested, that is producing modules and lines of code aside from the main GeoNode repository with the aim of modifying the original code as little as possible. A practical way to adopt this solution is creating a new Django application and writing new *models*, *views* and *templates*; anyway some little changes to the main repository are needed at least for hooking to it the new portion. Two solutions are feasible: the first requires adding new URLs to the *URLconf*, the ones needed by the site extension module; the second is adding the path of the new templates and the name of the new application in the Django *settings* module.

The second approach was used for the GeoNode development described in the present chapter: a new Django application was created, called *quality*, that can be launched by the user by clicking on the tab link called "Quality" in the site index page, as can be seen in Figure 7.1. Inside this application several code elements were generated: in particular new data types were added to the original GeoNode data schema; therefore the structure for these data storage was defined in the *models.py* module. Furthermore some new function and methods were defined for implementing the logical approach to the functionalities and were written in the module *views.py*. Finally new Web pages were prepared for improving the interaction with the users, therefore the

¹¹ <http://getfirebug.com>

correspondent HTML templates were created and stored in the *template* folder. Therefore new URLs were defined in order to trigger the new functionalities: they were stored in the variable *urlpatterns* of the *urls.py* module.

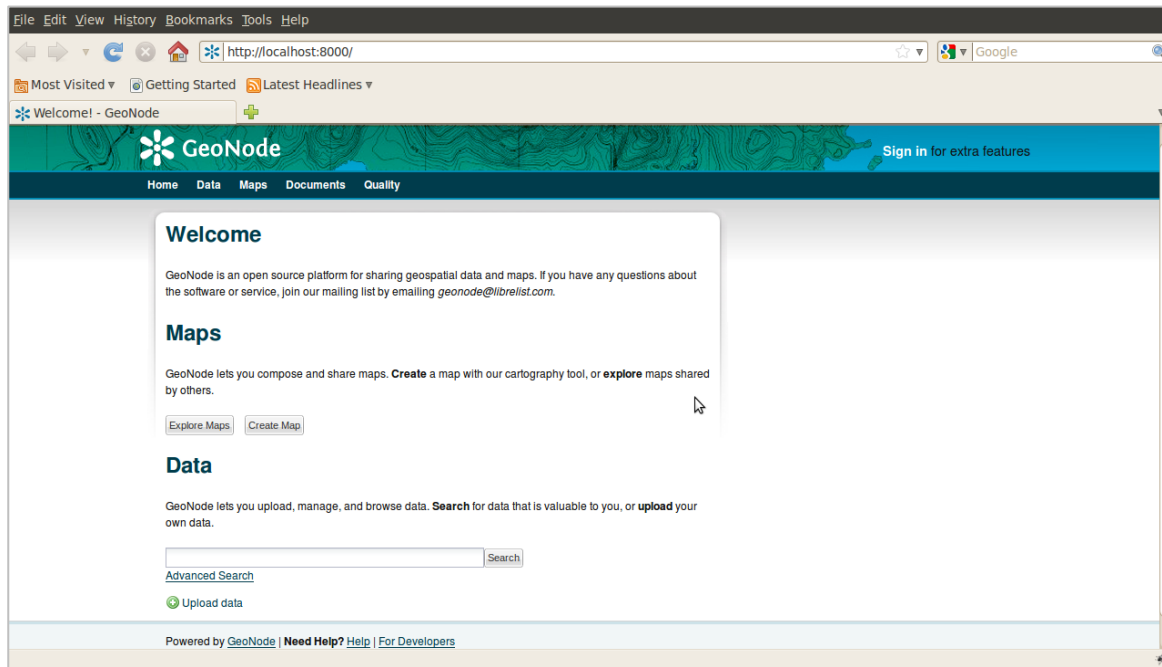


Figure 7.1 GeoNode homepage screenshot taken after the development. The new application “Quality” can be launched by clicking on the respective tab link.

Furthermore few modifications of the main repository were necessary for hooking the new application to the original ones, in particular: adding the path to the new HTML templates in the variable *TEMPLATE_DIRS* and the name of the new application in the variable *INSTALLED_APPS* of the module *settings.py*. Moreover the module *urls.py* of the GeoNode had to be modified by substituting the original module *layerController* with the new one stored in the application *quality* as well as adding the tuples identified in the module *urls.py* of the same application to the variable *urlpatterns*.

7.4 Data model

Before going into the development details it should be noted that the data of interest must be stored in the production server in a way that allows GeoNode to access them. Three types of information are needed: geospatial data, usually called also layers, their metadata and the results of the data quality assessment. Each of them requires a data model for their storage and access.

In details layers are stored in the usual way already described in chapter 6, in order for them to be easily accessed by the GeoServer that has the duty of their rendering. Instead metadata are stored using the standards implemented in the GeoNetwork and retrieved by the same application or by the Django components.

Finally the results of the quality assessment don't influence in any way the aspects of data representation, therefore they have not to be retrieved by the GeoServer. They don't have to be retrieved either by the GeoNetwork because they stand out of the common metadata standards.

Therefore the simplest and most suitable way to store them is in the database managed by the Django component; the respective data model is written in the Python module called *models.py*, following the Django rules, and is DBMS-independent.

In particular the data model comprehends two entities and a relationship; each one is implemented by a database table. The first entity is what can be called the *Quality Vector*, following the definition of *Quality Matrix* introduced in section 4.4.1. Each *Quality Vector* is a record of this matrix and therefore contains the scores given to the Quality Indicators and a reference to the dataset they describe. The correspondent data table contains as many integer fields as the quality indicators for storing values in the range 0÷5, where 0 is a code that has to be read as "not available"; an integer field for storing an identification code for every single record; a field for storing the foreign key in order to correctly refer the *Quality Vector* to the correspondent data layer.

The second entity is the *subtopic* in the way it was defined in section 3.1.1: briefly, it can be intended as a data category, a layer class. It was introduced in order to classify data layer retrieved during the data inventory phase. The correspondent database table contains a primary key field, which in this case is an integer field, and a text field for storing the sub-topic title; other fields related to the subtopic can then be added, such as a text field for a longer description of the *subtopic*.

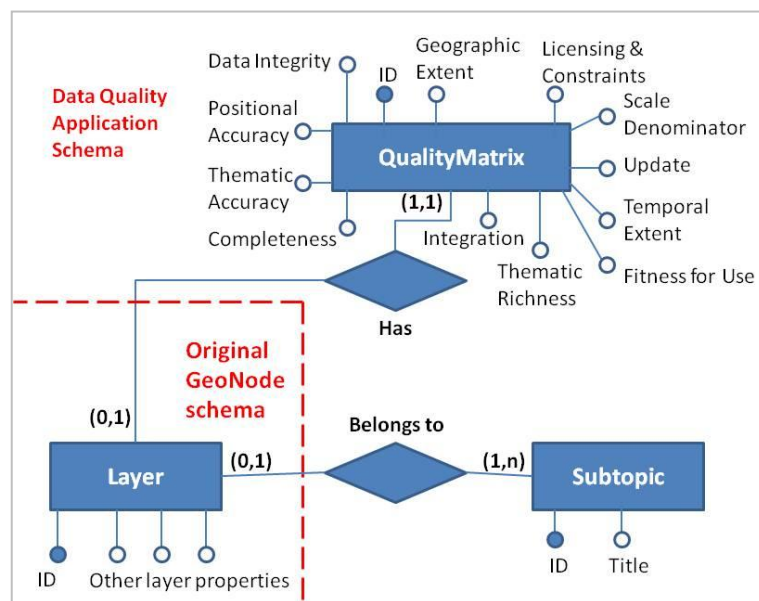


Figure 7.2 Entity-relationship schema of the Quality application of GeoNode

The relationship between the data layer and the sub-topic must be implemented in order to classify the layers; the relationship is implemented in the DBMS as a table that contains two foreign keys, one that references the layer and the other that references the correspondent *subtopic*.

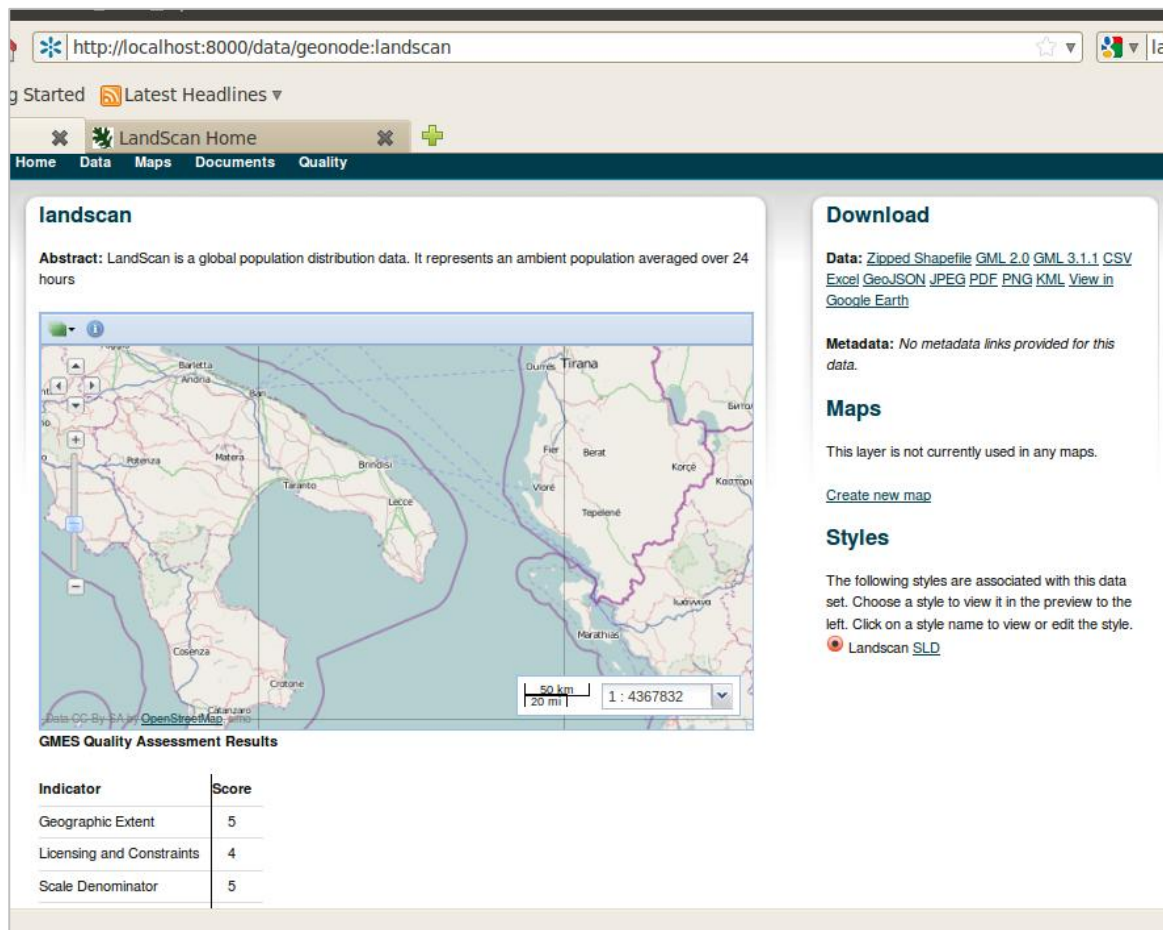
The Entity-Relationship schema of the Quality application data model is contained in Figure 7.2

7.5 Publication of the quality assessment results

The user of a GMES emergency management or land use service will have the availability of a geospatial data catalogue and several datasets will be valid alternative with similar content; therefore a choice of the dataset to use has to be made on the basis of some quality criteria.

In order to accomplish it, the user can perform a quality assessment if he is in possess of the required skills and the time necessary to the analysis; otherwise he can rely on the information provided by third parties. In this context the data quality assessment described in chapter 5 is central and the results are fundamental for the user to compare concurrent datasets and make a conscious choice.

The results of the data quality assessment are metadata by definition, therefore the natural place for them to be published is the GeoNode data page described in section 6.2, where all the metadata are shown together with the preview of the data they belong to.



The screenshot shows a web browser window displaying a GeoNode data page for a dataset named 'landscan'. The browser address bar shows 'http://localhost:8000/data/geonode:landscan'. The page has a navigation menu with 'Home', 'Data', 'Maps', 'Documents', and 'Quality'. The main content area is divided into two columns. The left column features a map of Italy with a red outline indicating the dataset's geographic extent. Below the map is a table titled 'GMES Quality Assessment Results' with the following data:

| Indicator | Score |
|---------------------------|-------|
| Geographic Extent | 5 |
| Licensing and Constraints | 4 |
| Scale Denominator | 5 |

The right column contains a 'Download' section with links for 'Zipped Shapefile', 'GML 2.0', 'GML 3.1.1', 'CSV', 'Excel', 'GeoJSON', 'JPEG', 'PDF', 'PNG', and 'KML'. Below this is a 'Metadata' section stating 'No metadata links provided for this data.', a 'Maps' section stating 'This layer is not currently used in any maps.', and a 'Styles' section with a radio button selected for 'Landscan SLD'.

Figure 7.3 A GeoNode example data page shows the results of the development: a table containing the internal data quality results is showed below the map preview

The data page is generated dynamically by Django on the basis of a template written part in HTML and part in Python language and called *layer.html*. The template has to be modified in order

to contain the information about data quality together with the other metadata. This isn't the only change needed, indeed the chain of changes follows below.

Firstly the URL request from the client to the server must be intercepted before it follows the usual path: this is done by modifying the variable *urlpatterns* in the *URLconf* module, especially the element in the list that contains the method *layerController*. In particular this method manages the requests related to data description and their metadata update; the *urlpatterns* entry has to be modified in a way that a new method substitutes the old one. Furthermore the mentioned method contains a reference to the new *layer.html* template and sends a new variable to it, that is the record of the table *QualityMatrix* linked to the objective layer through its foreign key. Secondly a table in the HTML template is going to appear and show each of the score obtained during the quality assessment.

Finally an *if...then, else* statement isolates the layers for which a quality assessment was performed in order to allow layers that have not been evaluated yet to be stored in the same GeoNode instance and published without any issues introduced by the development. An example data page that shows the results of the development to the user is reported in Figure 7.3.

7.6 Selection of the best quality dataset

It has been previously said that the applications described in this chapter are aimed at providing support to users who want to choose geospatial datasets using quality as criterion. In case the hypothetical user needs a dataset with a certain content, the best option at this point is investigating the results of the data quality assessment for the available datasets that belong to the correspondent data category, a sub-topic in the GMES terminology. A comparison of the mentioned results amongst the available datasets is needed in order to determine the best quality one: potentially only part of the results can be considered, for instance the ones related to some quality indicators of greater interest.

Although this approach is simple and direct, it requires the user to open a data page for each layer and either perform a visual and mental comparison amongst the layers or copy the quality results on another support in order to simplify the process. An application with the task of performing this process was developed on top of the GeoNode: it allows the extraction of the best quality dataset from the server, when the user provides the name of the sub-topic of interest and a series of weights for the quality indicators, as it was described in section 4.4.2.

The entire application was developed in Python, on top of the Django component of GeoNode; it is composed by two HTML templates and three methods in the views of the MCV pattern. It is available to be launched by the user from the main toolbar, either by clicking the link placed on the tab *Quality* or by typing the following URL after having replaced the word *sitename* with the actual name of the website:

<http://sitename/quality/subtopics>

In details the Django application intercepts the URL because it is defined in the variable *urlpatterns* and executes the method *listSubtopic*: the core of the method is a query over the database on the table *Subtopic* in order to get the list of sub-topics available on the server, filtered by checking for each one if at least one layer on the server is assigned to it. An *HttpResponse* is then returned by generating a web page from the template *subtopic.html*.

Consequently the user is going to see a Web page with a list of the subtopic retrieved by the query; next to each element of the list is a radio button. The user can select the subtopic of interest by clicking on the correspondent radio button and then submit the choice by clicking the appropriate button placed at the end of the page. A screenshot of the mentioned Web page is contained in Figure 7.4.

As a consequence this action is translated into a URL that contains a *GET* request with the identification code of the selected sub-topic. The application intercepts this new URL and executes the method called *ask4weights*. This method substantially performs a check on the nature, either Core or Thematic (see section 3.1.1), of the selected sub-topic, in relation to which it generates a web page respectively with 12 or 8 text boxes, referencing the *Quality Indicators*, on the basis of the template called *ask4weights.html*.

In this new page, an example of which is reported in refxyz, the user is asked to fill in the text boxes next to the *Quality Indicators* that he considers interesting for the dataset choice, with a weight value included in the range 1 to 10. By clicking on the button placed at the bottom of the page, the user can submit the weights for the Quality Indicators; the submission is translated into a URL that contains a *GET* request with all the numeric values.

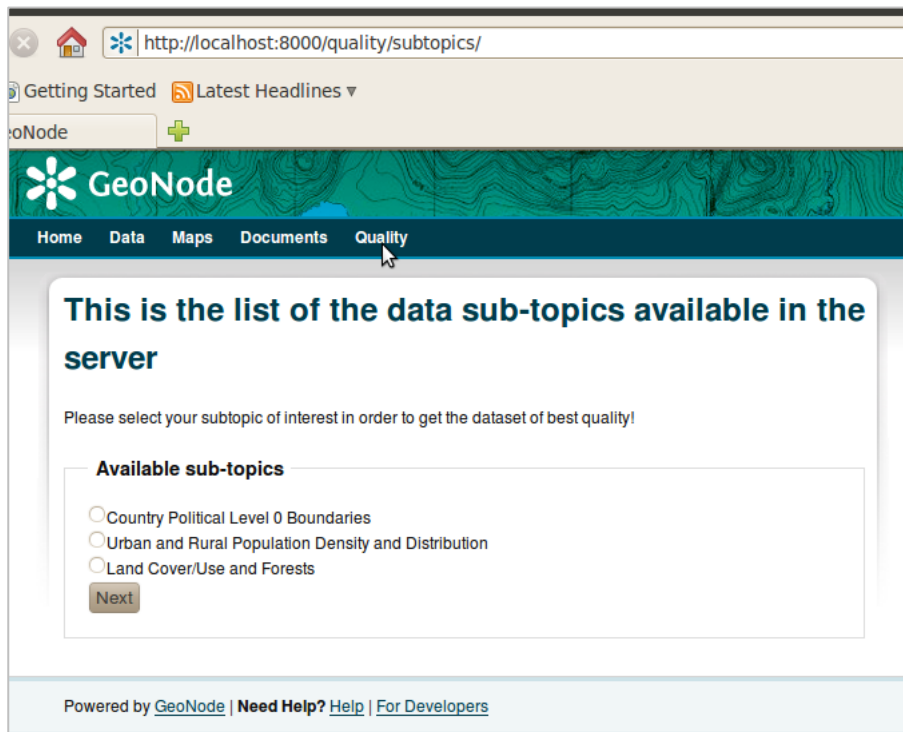


Figure 7.4 Screenshot of the Web page that is showed to the user when the “Quality” application is launched. The user can select here the sub-topic of interest and submit his choice back to the server

Once again the application intercepts this last *URL* and executes the method *calculateBest*, that is the core of the application. In details a query is performed over the database in order to get all the available datasets belonging to the subtopic of interest of the user. The system then calculates, in loop over the selected datasets, the total quality score in the way described in section 4.4; the GeoNode data page of the layer correspondent to the highest quality score is then returned to the

user, who can either use it in the GeoNode itself or download it for using it in another application.

Resuming, with the application described above, a hypothetical user can extract the dataset with best quality by submitting his needs in terms of numeric weights of the *Quality Indicators* and a category in terms of data content. The efforts required on user side is selecting a subtopic by mean of a radio button on a web page and filling the text boxes with the weights for the indicators of interest in the web page displayed afterwards.

This application is clearly related to the one of section 7.5 because at the end of the process the data page is shown, that contains the results of the data quality assessment. Moreover the Django views access the same database table.

http://localhost:8000/quality/ask4weights/?subtopic=2%2F

Latest Headlines ▾

GeoNode

Home Data Maps Documents Quality

Urban and Rural Population Density and Distribution

Please define your own weights between 0 and 10!

When you're done, press the "Calculate" button in order to get the dataset with best quality

Quality Indicators:

| | |
|---------------------------|----|
| Geographic Extent | 10 |
| Licensing and Constraints | 8 |
| Scale Denominator | 8 |
| Update | 5 |
| Temporal Extent | 3 |
| Fitness for Use | 10 |
| Thematic Richness | 8 |
| Integration | 6 |
| Data Integrity | 8 |
| Positional Accuracy | 9 |
| Thematic Accuracy | 9 |
| Completeness | 10 |
| The subtopic code is | 2 |

Calculate!

Figure 7.5 Screenshot from an example GeoNode page that allows the user to insert his needs as weights for each Quality Indicator. The weights submission results in the calculation of the best quality data in the chosen data category and the re-direction to the respective data page

7.7 Open issues and further development

The GeoNode application described in the previous paragraphs, even if constitutes a first approach to the publication of the data quality assessment results and a simple tool for the selection of best quality data, is efficient, reliable and easy to install and use.

Anyway improvements can be conceived and developed in order to either simplify further on the application or to get different or more complex results. Below some of the potential issues of the application and the proposals aimed at providing the right solutions are introduced.

The first identified issue is that the highest score calculated by the application can correspond to a dataset for which the user doesn't have the privileges for the visualization and downloading. The solution for this case can be double: on the one hand a control on the user permission can be inserted in the method that calculates the total scores for the datasets, in order to extract the best quality dataset amongst the ones the user has permissions on visualization.

On the other hand as a result of the calculation a table with all the datasets belonging to the subtopic of interest together with the total score for each one can be provided. Next to each dataset a link to the correspondent data page can be shown: in this way the permission issue is managed by the normal GeoNode procedures.

A more complex and interesting application would be the option of creating a template map with some layer selected from specific data categories. A proposal for obtaining these results can be enabling the possibility of choosing more than one subtopic at a time when displaying the *subtopic.html* template; instead the submission of the weights for the *Quality Indicators* wouldn't change. The application should be modified then in order to extract the best quality dataset for each of the subtopic enabled and to generate a first general map with them. The layer order in the map can be defined on the basis of the geometry types of the datasets, while the default style would be assigned.

Finally the following situation was faced so far: the data quality assessment was performed before the publication of the results for all datasets. Therefore the insertion of the results was performed directly on the server either using the Django management commands or the *SQL* language through the database interface. Anyway the possibility that skilled users can upload new datasets on the server and perform data quality assessment on them has to be taken into consideration. In order to permit this type of collaboration from the users, new user forms and new web pages have to be developed as well as the respect of user permissions; submitting values of data quality assessment has to be allowed by mean of *POST* requests. The development of this application has to be still designed.

Conclusion

Data and technology are the two main topics of the present dissertation: they're among the core components of European environmental monitoring services, in particular of the emergency management services, which have huge importance for citizen's security.

In chapter 1 the engagement of the European Union towards the global environmental monitoring for reason of citizens security was discussed. Proof is the GMES initiative that aims at implementing monitoring services in air, water and soil systems, as well as citizens' security services when natural or man-made disasters and humanitarian crises occur. At the time being all the services have already reached at least the pre-operational phase whereas some are now in the operational phase.

In particular all the services are meant to be online tools, that gives a key role to the internet communication. Moreover all the services are going to have an important spatial component, indeed the communication message contains mostly geospatial data conveyed by mean of traditional maps or WebGIS applications. In this context both data and technology are key issues; data can be thematic, for instance the results of specific analyses, and reference data, that are useful for comparison and for the detection of relative long term changes. On the other hand technology is also a key issue because nowadays data are digital or are converted into a digital format, therefore it is required in many steps of data management, from data collection to data visualization, from data storage to data processing, from data checking to data publishing. Geomatics is the science that studies the connection and the relationship between geospatial data and information technology; it has a strong importance in the GMES initiative and in all the projects, which involve environmental monitoring.

In chapter 2 the pre-operational steps of the GMES initiative were introduced and a particular focus was given to base reference data. A call for tenders for the implementation of an initial service for reference data access was presented as well as the division into two lots. The tender contractor and the service primary users were introduced. The project second lot was taken into consideration because of the global extent and the objective that is strongly related to the emergency management and land monitoring.

Moreover the main tasks and activities required for building and implementing a reference data access were presented; firstly an assessment of the end user needs is required in order to define

the service specifications. Secondly an inventory of the available and suitable data for building the spatial data repository at the basis of the data access should be carried out. Thirdly a quality assessment of the data collection is compulsory and aimed at assure the minimum requirements in terms of data correctness, richness and completeness. Finally demonstrators of the reference data access functionality and characteristics should be developed following the principles of the final GMES services, therefore by taking advantage of the internet communication and of WebGIS tools.

Again data and technology are the fundamentals for building a reference data access; the technical proposal presented by the project contractor and accepted by the European Commission technical personnel contains the steps presented above. To sum up, these can be considered as general key tasks for building a reference data access in many other cases.

In chapter 3 the status of the GMES RDA Lot 2 project was disclosed and an overview of all the project tasks was provided. At the time being the end user needs assessment is considered complete. End users were required to cooperate in giving feedbacks and define service specifications: several international agencies offered to participate to the project and give their help.

The data inventory task is considered complete as well. It was divided into two main activities, firstly the definition of the minimum content of the spatial data repository by proposing a data categorization into topics and subtopics. Secondly an activity of data searching was required in order to provide at least one dataset for each subtopic: usually more than one candidate dataset was provided. The data inventory phase was a high time-consuming activity; it was carried out with a profitable method and gave good results.

The data quality assessment task is also complete and it was divided into two steps as well. First a protocol for data quality assessment needed to be defined in order to fix a common and efficient procedure for quality evaluation. Afterwards the protocol was applied to the data repository collected during the previous inventory phase in order to obtain the final results and make comparisons among datasets.

Instead the demonstrators are still in development phase even if their first release is already up and running; they're the result of the combination between WebGIS technologies and free and open source software. They allow the base reference data visualization on a map canvas as well as the exploitation of some simple and effective GIS tools related to zoom and layer management; they allow data downloading and metadata checking; they're based on real natural disasters occurrence and allow to take advantage of some value-added information by mean of visual comparison with the base reference data. The demonstrators were mostly considered useful and significant by the European Commission representatives.

To sum up, in general the technical approach followed by the contractor matched the project requirements; the reference data access, objective of the project, is ready for exploitation. Only the development of some demonstrator modifications, following the end users feedback, is still a work in progress.

In chapter 4 the data quality protocol adopted for the GMES RDA Lot 2 project was presented; it was defined specifically for the project after a bibliographic review, in order to be efficient enough for direct application. The protocol required also to be general enough in order to be adapted to all the potential data types, both in terms of content and in terms of file and geometry type of the

wide geospatial data range. First a definition of data quality was given, on the basis of which an assessment method was defined. The method was conceived with a quantitative approach based on quality indicators that can assume certain values; then numerical scores were given by fixing thresholds on indicators values. The method allows also to take into account the user needs in terms of quality by assigning weights to the contribute of each indicator on the final result. Moreover the quality protocol takes into account various aspects of geospatial data such as data content richness, license conditions, coverage extent, update, integration among data sources, geometry conditions and completeness.

The data quality protocol was applied to all the datasets collected during the inventory phase of the project and was useful not only for giving data a quality level but also for ranking on a quality basis the datasets grouped into the same data category, allowing to perform a conscious data choice among valid alternatives.

In chapter 5 examples of data quality protocol application were reported in order to show the assessment activities and demonstrate some protocol characteristics. In particular the application on two data file types, namely raster and vector, and the process of adaptation required in order to get satisfactory results is explained.

Some consideration on the data quality application were then reported: the protocol defined specifically for the GMES RDA Lot2 project can be considered a valid instrument for every situation that requires the quality level ranking of a number of geospatial datasets grouped into categories. Although the data quality assessment concluded successfully, some points for protocol modification and improving were found and highlighted, aimed at performing a faster assessment and extend the potential protocol users community. In particular the modification of some indicator thresholds should be considered for exploiting the entire score range in a wider situation variability. Also some scripts in a suitable programming language should be written in order to simplify and speed up the work of the data quality analysts.

On a wider perspective the development of guided procedures in a simplified version of the protocol could allow non-experienced GIS users to take advantage of the protocol, when following a way of protocol publication and dissemination would potentially become an objective. Moreover in such a context the exploitation of open source software should be considered, in the quality analysis that at time being requires proprietary software, for broadening the community of the protocol potential users.

In chapter 6 the focus switched towards the GMES RDA Lot 2 project demonstrators, detailing the technology rather than their content. In particular the GeoNode application, that is the software constituting one of them, was described. However the other demonstrators are based on GeoNode components, therefore technical information on the demonstrators can be derived reading its description.

The nature of GeoNode as Web platform for geospatial data sharing was described and most of the functionalities on the client side were introduced. It combines together the innovation in different fields such as WebGIS, standard communication protocols, free and open source software, protocols and programming languages. Even if it has a high degree of complexity and delicate relationship among its components, it was made to be easy to use and thinking to GIS-illiterate users. Nevertheless it has some fundamental features useful for GIS experts as well as some characteristics typical of a spatial data infrastructure.

The current GeoNode role in emergency management around the world played in the challenge of allowing interoperability among international key agencies and of building resilience against natural disasters in developing countries was disclosed.

To sum up, the GeoNode is a complex application not yet completely stable in some features, but has the potentiality to become the standard application for data sharing in the field of environmental emergencies and natural disasters management; in particular it has a high innovation content, it is easy to use for GIS-illiterate people due to the simple and direct interface and the similarity with social networks. Moreover it allows direct and simple data sharing and has an active developer community.

In chapter 7 the technical strategies for the GeoNode development were introduced and the possible ways of exploiting the data quality assessment results in order to make them available to the end users were taken into consideration.

The first topic was faced by describing an approach for developing a GeoNode extension avoiding issues in future upgrades of the main repository, addressing the potentiality of each component development, suggesting which files should be modified and for what purposes, the path of data and file storage. All the code developed for the purposes of this chapter is contained in the annexes.

On the other side two exploitation possibilities of the quality assessment results within this application were introduced: one is the publication of the results themselves among the rest of the metadata; while the other is the implementation of a tool for dataset choosing among valid candidate datasets belonging to a certain content category. The development work allowed to experience a relative simplicity in the GeoNode development and the smart configuration that allows to generate extensions and maintain the original code, avoiding in such way issues when upgrading the software to newer versions.

GeoNode is a good application to publish data quality results as well as the other metadata, due to its metadata catalogue. It is also a good platform for the implementation of a tool for data choosing due to it being targeted also to non-GIS experts and the direct contact with the end users. The Web is a good channel for the publication of data quality information; in this way it can reach easily the end user and help to make conscious quality-based choices. The tool can help all kind of users, even if experienced user could better make their own choice on the basis of previous experience with datasets and maps. Some developments to the tool are still needed before the implementation in the final version of the GMES RDA Lot 2 project demonstrator, such as the possibility of automatic map generation on the basis of some user inputs in terms of quality needs and content and the possibility for experienced users to define some quality results instead of being only in read-only mode.

In conclusion data and technology are fundamentals for building a reference data access and this is a base component of an environmental global monitoring service. Data quality is a key issue for users to rely on environmental monitoring services and for citizens to really take advantage on the services themselves.

Innovations and easiness of use are key issues for a WebGIS application to be helpful and become popular and widely exploited. Finally the implementation of a reference data access service is a complex activity that requires huge efforts and should follow specific rules and tasks in order to be successful.

Bibliography

- [1] Beare M., Riikka H., Jakobsson A., Marttinen J., Onstein E., Tsoulos L., Williams F., Mäkelä J., De Meulenaer L., Persson I., Kavadas I., *D 8.4 ESDIN Quality Final Report – Part A*, EuroGeographics 2010
- [2] Broglia M., Corbane C., Carrion D., Lemoine G., Pesaresi M. *Validation Protocol for Emergency Response Geo-information Products*, Ispra (VA), Italy, European Commission Joint Research Centre, 2010
- [3] Comoglio G. *Elementi di Cartografia*, Politecnico di Torino, DITAG, 2004
- [4] Dalmasso S., *WebGIS Architectures for Emergency Response* PhD Thesis, Politecnico di Torino, 2011
- [5] Devillers R., Jeansoulin R. *Fundamentals of Spatial Data Quality*, London, ISTE Ltd, 2006
- [6] European Commission, Enterprise and Industry Directorate-General, Space Research and Development *Service Contract Number SI2.547842, Draft Annex I, Tender Specification and Monitoring* Brussels-Turin 2009
- [7] European Environment Agency *Report on in-situ requirements* GMES In-Situ Coordination 2011
- [8] Federal Geographic Data Committee, Subcommittee for Base Cartographic Data *Geospatial Positioning Accuracy Standards - Part 3: National Standard for Spatial Data Accuracy* FGDC-STD-007.3-1998
- [9] GeoNetwork OpenSource *User Manual v2.6.4 - Geographic Information Management for all*, 2012
- [10] Gomarasca M., *Basics of Geomatics* Springer, London 2009
- [11] INSPIRE D2.8.I.4 *Data Specifications on Administrative units – Guidelines*
- [12] INSPIRE *Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119* EC Joint Research Centre, 2007

- [13] ISO/TC 211, 19113 *Geographic Information – Quality Principles*, International Organization for Standardization (ISO), 2002
- [14] ISO/TC 211, 19115 *Geographic Information – Metadata*, International Organization for Standardization (ISO), 2004
- [15] ITHACA, SiTI *CALL FOR TENDERS No ENTR/2009/27 Implementation of an initial GMES Service for Geospatial Reference Data Access SECTION 4 – TECHNICAL PROPOSAL*, Torino 2009
- [16] ITHACA, SiTI *Call For Tenders No ENTR/2009/27 Lot 2 – Implementation of an initial GMES service for geospatial reference data access covering areas outside Europe Deliverable 11.1 Interim Report M9*, Torino 2010
- [17] ITHACA, SiTI *Call For Tenders No ENTR/2009/27 Lot 2 – Implementation of an initial GMES service for geospatial reference data access covering areas outside Europe Deliverables 10.1, 10.2, 10.3, 11.1, 11.2, 12, 13, 14.1, 14.2, 14.3 Interim Report M21*, ITHACA 2011
- [18] Joint Research Centre, European Commission *Guidelines for Best Practice and Quality Checking of Ortho Imagery* Ispra (VA) – Italy 2003
- [19] Nunes de Lima V., Tóth K., *Data Quality and Scale in Context of European Spatial Data Harmonisation* 1th EC-GIS Workshop, Alghero, 29 June – 01 July 2005
- [20] Open Geospatial Consortium Inc. *OpenGIS® Web Map Server Implementation Specification*, 2006
- [21] Open Geospatial Consortium Inc. *OpenGIS Web Feature Service 2.0 Interface Standard*, 2010
- [22] Open Geospatial Consortium Inc. *The OpenGIS Abstract Specification, Topic 6: Schema for coverage geometry and functions*, 2006
- [23] Open Geospatial Consortium Inc. *OGC WCS 2.0 Interface Standard – Core*, 2010
- [24] Patrignani N., *Computer Ethics* Materiale del Corso di III Livello, Politecnico di Torino 2009

Webography

- [1] Cunningham & Cunningham Inc. website, *c2.com : Don't Repeat Yourself* – Available on November 22nd, 2011 at: <http://c2.com/cgi/wiki?DontRepeatYourself>
- [2] Django project website – Available on January 12th, 2012 at: <https://www.djangoproject.com>
- [3] European Commission website, GMES pages – Available on October 25th, 2011 at: <http://ec.europa.eu/enterprise/policies/space/gmes>
- [4] ESRI website, ArcGIS Desktop 9.3 Help page – Available on December 12th, 2011 at: <http://webhelp.esri.com/arcgisdesktop/9.3>
- [5] FAO GeoNetwork website, GAUL metadata page – Available on January 13th, 2011 at: www.fao.org/geonetwork/srv/en/metadata.show?id=12691
- [6] GADM website – Available on January 13th, 2012 at: www.gadm.org
- [7] GeoExt documentation, *GeoExt : Documentation : Core Concepts*, Available on February 3rd, 2012 at : <http://geoext.org/primers/index.html>
- [8] GeoNode website – Available on October 26th, 2012 at: <http://geonode.org>
- [9] GeoServer website, *GeoServer : What is GeoServer*, Available on February 3rd, 2012 at: <http://geoserver.org/display/GEOS/What+is+GeoServer>
- [10] GFDRR website, Labs page – Available on January 29th, 2012 at: <http://www.gfdr.org/gfdr/node/175>
- [11] GFDRR website, OpenDRI page - Available on January 29th, 2012 at: <http://www.gfdr.org/gfdr/pendri>
- [12] GMES website – Available on October 23rd, 2011 at: www.gmes.info
- [13] GMES website, Land Monitoring page – Available on October 23rd, 2011 at: www.gmes.info/pages-principales/services/land-monitoring
- [14] GMES website, Emergency Management page – Available on October 23rd, 2011 at: www.gmes.info/pages-principales/services/emergency-management

- [15] gMosaic website – Available on October 22nd, 2012 at: <http://www.gmes-gmosaic.eu>
- [16] INSPIRE website, Data Specifications page – Available on December 21st, 2012 at: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/7>
- [17] ITHACA website – Available on January 15th, at: <http://www.ithacaweb.org>
- [18] Landscan website – Available on December 17th, at: www.ornl.gov/sci/landscan/index.shtml
- [19] MACC website – Available on November 24th, at: <http://www.gmes-atmosphere.eu>
- [20] My Ocean website – Available on October 27th, at: <http://www.myocean.eu.org>
- [21] OGC website, *OGC : FAQs - OGC's Purpose and Structure* – Available on January 30th, 2012 at: <http://www.opengeospatial.org/ogc/faq>
- [22] OpenGeo website, *OpenGeo blog : GeoExplorer preview*, Available on January 30th, 2012 at: <http://blog.opengeo.org/2009/06/17/geoexplorer-preview/>
- [23] OpenLayers website, *OpenLayers : Home*, Available on January 30th, 2012 at: <http://openlayers.org/>
- [24] PostGIS website, *PostGIS : Home*, Available on February 3rd, 2012 at: <http://www.postgis.org/>
- [25] SAFER website – Available on October 23rd, 2011 at: <http://www.emergencyresponse.eu/gmes/en/ref/home.html>
- [26] SAFER website, Landslides page – Available on October 23rd, 2011 at: http://safer.emergencyresponse.eu/site/FO/scripts/myFO_contenu.php?noeu_id=50&lang=EN
- [27] SAFER website, Cambodia floods - Available on October 23rd, 2011 at: http://www.emergencyresponse.eu/gmes/en/event/Floods-in-Cambodia_115.html
- [28] SEDAC website, information page – Available on November 24th, 2011 at: <http://sedac.custhelp.com>
- [29] Sencha website, *Sencha : Products : ExtJS4 : Overview*, Available on February 3rd, 2012 at: <http://www.sencha.com/products/extjs/>
- [30] SiTI website – Available on January 15th, 2012 at: <http://www.siti.polito.it>
- [31] UNESCO website, WWDR page – Available on December 14th, 2011 at: www.unesco.org/new/en/natural-sciences/environment/water/wwap/wwdr/
- [32] WFP website – Available on December 1st, 2011: www.wfp.org
- [33] Wikipedia, Framework Programmes for Research and Technological Development – Available on November 5th, 2011 at: http://en.wikipedia.org/wiki/Framework_programme
- [34] Wikipedia, Model-View-Controller – Available on November 18th, 2011 at: http://en.wikipedia.org/wiki/Model_view_controller
- [35] Wikipedia website, *Wikipedia : PostgreSQL*, Available on February 3rd, 2012 at: <http://en.wikipedia.org/wiki/Postgresql>

- [36] WWDRII website, download page, Available on December 17th, 2011 - <http://wwdrii.sr.unh.edu/download.html>

Annex A – GeoNode Quality Application Code

quality/README.rst

Data Quality Application for GeoNode

=====

This is a GeoNode application aimed to publishing the data quality assessment results for the GMES Reference Data Access Lot 2 project. It defines a data model for quality results and some templates in order to retrieve and publish them. Furthermore it implements a tool for data choosing on a quality basis.

Installation

#. Install the application:
source bin/activate

```
pip install -e git+git://github.com/sbsimo/quality.git
```

#. Add `quality` to the variable `INSTALLED_APPS` in the Django `settings.py` file.

#. Add the following line to the variable `TEMPLATE_DIRS` in the Django `settings.py` file:

```
os.path.join(PROJECT_ROOT, "..", "..", 'quality', 'quality',  
"templates"),
```

#. Append the following line to the Django `urls.py` module, in the variable `urlpatterns`:

```
(r'^quality/', include('quality.urls')),
```

#. Modify the following line in the `urls.py` module:

```
(r'^data/(?P<layername>[^/]*)$',  
'geonode.maps.views.layerController'),
```

and substitute with the following:


```
(r'^data/(?P<layername>[^\/]*)$', 'quality.views.layerController'),  
  
#. Run ``syncdb`` command and reload the web server in order to get to see  
the application working
```

quality/setup.py

```
import os  
from setuptools import setup, find_packages  
  
def read(*rnames):  
    return open(os.path.join(os.path.dirname(__file__), *rnames)).read()  
  
setup(  
    name="quality",  
    version="0.1",  
    author="Simone Dalmaso, Simone Balbo",  
    author_email="simone.blb@gmail.com",  
    description="Data quality extension for GeoNode",  
    long_description=(read('README.rst')),  
    classifiers=[  
        'Development Status :: 2 - Pre-alpha',  
        'Framework :: GeoNode',  
        'License :: OSI Approved :: BSD',  
    ],  
    license="BSD",  
    keywords="geonode django",  
    url='https://github.com/sbsimo/quality',  
    scripts = [  
    ],  
    packages=find_packages('.'),  
    include_package_data=True,  
    zip_safe=False,  
)
```

quality/quality/models.py

```
from django.db import models  
from geonode.maps.models import Map, Layer  
from django.db.models import signals  
import os  
  
class QualityMatrix(models.Model):  
    layer = models.OneToOneField(Layer)  
    geographicExtent = models.IntegerField('Geographic Extent')  
    licensingConstraint = models.IntegerField('Licensing and Constraint')  
    scaleDenominator = models.IntegerField('Scale Denominator')  
    update = models.IntegerField('Update')
```

```
temporalExtent = models.IntegerField('Temporal Extent')
fitness4Use = models.IntegerField('Fitness for Use')
thematicRichness = models.IntegerField('Thematic Richness')
integration = models.IntegerField('Integration')
dataIntegrity = models.IntegerField('Data Integrity')
positionalAccuracy = models.IntegerField('Positional Accuracy')
thematicAccuracy = models.IntegerField('Thematic Accuracy')
completeness = models.IntegerField('Completeness')

class Subtopic(models.Model):
    title = models.CharField(max_length=100)

class LayerSubtopic(models.Model):
    layer = models.OneToOneField(Layer)
    subtopic = models.ForeignKey(Subtopic, verbose_name="the related
subtopic")
```

quality/quality/urls.py

```
from django.conf.urls.defaults import patterns, url

urlpatterns = patterns('',
    url(r'^subtopics/', 'quality.views.listSubtopics'),
    url(r'^ask4weights/', 'quality.views.ask4weights'),
    url(r'^weights/', 'quality.views.calculateBest')
)
```

quality/quality/views.py

```
from geonode import settings
from geonode.maps.views import _perms_info, MAP_LEV_NAMES,
    _perms_info_json, \
    LAYER_LEV_NAMES, _describe_layer
from django.shortcuts import render_to_response, get_object_or_404
from django.http import HttpResponse
from geonode.maps.models import Map, MapLayer, Layer
import json
from django.template import RequestContext, loader, Context
from django.utils.translation import ugettext as _
#from cartography.models import Document
from django.contrib.auth.decorators import login_required
from geonode.maps.views import default_map_config
from django.views.decorators.csrf import csrf_exempt
from quality.models import Subtopic, LayerSubtopic, QualityMatrix

@csrf_exempt
def layerController(request, layername):
    DEFAULT_MAP_CONFIG, DEFAULT_BASE_LAYERS = default_map_config()
```

```

layer = get_object_or_404(Layer, typename=layername)
if (request.META['QUERY_STRING'] == "describe"):
    return _describe_layer(request, layer)
if (request.META['QUERY_STRING'] == "remove"):
    return _removeLayer(request, layer)
if (request.META['QUERY_STRING'] == "update"):
    return _updateLayer(request, layer)
if (request.META['QUERY_STRING'] == "style"):
    return _changeLayerDefaultStyle(request, layer)
else:
    if not request.user.has_perm('maps.view_layer', obj=layer):
        return HttpResponse(loader.render_to_string('401.html',
            RequestContext(request, {'error_message':
                _("You are not permitted to view this layer")})),
status=401)

    metadata = layer.metadata_csw()

    maplayer = MapLayer(name = layer.typename, ows_url =
settings.GEOSERVER_BASE_URL + "wms")

    # center/zoom don't matter; the viewer will center on the layer
bounds
    map = Map(projection="EPSG:900913")

    qualityRecord = layer.qualitymatrix

    return render_to_response('quality/layer.html',
RequestContext(request, {
        "layer": layer,
        "metadata": metadata,
        "viewer": json.dumps(map.viewer_json(* (DEFAULT_BASE_LAYERS +
[maplayer]))),
        "permissions_json": _perms_info_json(layer, LAYER_LEV_NAMES),
        "GEOSERVER_BASE_URL": settings.GEOSERVER_BASE_URL,
        "qualityRecord": qualityRecord
    }))

GENERIC_UPLOAD_ERROR = _("There was an error while attempting to upload
your data. \
Please try again, or contact and administrator if the problem continues.")

def listSubtopics(request):
#     access to the table that contains the list of subtopics
    allSubtopics = Subtopic.objects.all()
    return render_to_response('quality/subtopics.html',
RequestContext(request, {
    'allSubs' : allSubtopics,
    }))

def ask4weights(request):
    if request.method == 'GET':
        subtopic_pk = request.GET.__getitem__("subtopic")[0]
        subtopic = Subtopic.objects.get(pk=subtopic_pk)

```

```

        return render_to_response('quality/ask4weights.html',
RequestContext(request, {
            'subtopic': subtopic,
            'subtopic_pk': subtopic_pk,
        }))
    else:
        return HttpResponse(loader.render_to_string('401.html',
RequestContext(request, {'error_message':
            _("You are not permitted to view this layer")})),
status=401)

def calculateBest(request):
    if request.method == 'GET':
#         get the weights input by the client
        weightVector = [request.GET.__getitem__("geographicExtent")]

        weightVector.append(request.GET.__getitem__("licensingConstraint"))

        weightVector.append(request.GET.__getitem__("scaleDenominator"))
        weightVector.append(request.GET.__getitem__("update"))
        weightVector.append(request.GET.__getitem__("temporalExtent"))
        weightVector.append(request.GET.__getitem__("fitness4use"))

        weightVector.append(request.GET.__getitem__("thematicRichness"))
        weightVector.append(request.GET.__getitem__("integration"))
        weightVector.append(request.GET.__getitem__("dataIntegrity"))

        weightVector.append(request.GET.__getitem__("positionalAccuracy"))

        weightVector.append(request.GET.__getitem__("thematicAccuracy"))
        weightVector.append(request.GET.__getitem__("completeness"))
#         get the subtopic and the set of related layersubtopics
        subtopic_id = request.GET.__getitem__("subtopic_pk")
        subtopic = Subtopic.objects.get(pk=subtopic_id)
        layersubtopics = subtopic.layersubtopic_set.all()
#         define initial values
        winner_layer_id = 0
        bestScore = 0
#         loop on layersubtopics in order to calculate the total score
for each one
        for layersubtopic in layersubtopics:
            currentLayer = layersubtopic.layer
            qualityVector =
QualityMatrix.objects.get(layer=currentLayer)
#             calculate the quality total score of the layer
            comparisonScore =
qualityVector.geographicExtent*weightVector[0]+\
            qualityVector.licensingConstraint*weightVector[1]+\
            qualityVector.scaleDenominator*weightVector[2]+\
            qualityVector.update*weightVector[3]+\
            qualityVector.temporalExtent*weightVector[4]+\
            qualityVector.fitness4Use*weightVector[5]+\
            qualityVector.thematicRichness*weightVector[6]+\
            qualityVector.integration*weightVector[7]+\

```

```

        qualityVector.dataIntegrity*weightVector[8]+\
        qualityVector.positionalAccuracy*weightVector[9]+\
        qualityVector.thematicAccuracy*weightVector[10]+\
        qualityVector.completeness*weightVector[11]
        if comparisonScore > bestScore:
            bestScore = comparisonScore
            winner_layer_id = layersubtopic.layer.id

    winnerLayer = Layer.objects.get(id=winner_layer_id)
    layername = winnerLayer.typename
    return layerController(request, layername)
#     return render_to_response('quality/temp.html',
RequestContext(request, {
#         'weightVector': weightVector,
#         'layername': layername,
#     }))
    else:
        return HttpResponse(loader.render_to_string('401.html',
            RequestContext(request, {'error_message':
                _("You are not permitted to view this layer")})),
status=401)

```

quality/quality/page_layout.html

```

{% extends "../../../../../GeoNodePy/geonode/templates/page_layout.html" %}
{% load i18n %}

{% block nav %}
    <ul id="sitenav" class="selfclear">
        <li class="home"><a href="{% url geonode.views.index %}">{% trans
"Home" %}</a></li>
        <li class=""><a class="" id="dataLink" href="{% url
geonode.maps.views.browse_data %}">{% trans "Data" %}</a></li>
        <li class=""><a class="" id="mapsLink" href="{% url
geonode.maps.views.maps %}">{% trans "Maps" %}</a></li>
        <li class=""><a class="" id="documentLink" href="{% url
cartography.views.upload_document %}">{% trans "Documents" %}</a></li>
        <li class=""><a href="{% url quality.views.listSubtopics
%}">Quality</a></li>
        {% if user.is_authenticated %}
            <li class=""><a class="" id="profileLink" href="{% url
profiles_profile_detail user.username %}">{% trans "Profile" %}</a></li>
        {% if user.is_staff %}
            <li class=""><a class="" id="adminLink" href="/admin/">{% trans
"Admin" %}</a></li>
        {% endif %}
    {% endif %}
    </ul>
{% endblock %}

```

quality/quality/templates/quality/ask4weights.html

```

{% extends "page_layout.html" %}
{% load i18n %}

{% block title %} {% trans "Browse Data" %} - {{ block.super }} {% endblock %}

{% block head %}
{% include "geonode/ext_header.html" %}
{% include "geonode/app_header.html" %}
{{ block.super }}
<script type="text/javascript">
Ext.onReady(function(){
{% autoescape off %}
    var submitForm = function() {
        Ext.query('#search_form form')[0].submit();
    };
    var searchButton = new Ext.Button({renderTo: 'submit_search', text: '{%
trans "Search" %}',
                                     handler: submitForm});
    var searchInput = new Ext.form.TextField({applyTo: 'search_query', width:
350});

    var goButton = new Ext.Button({
        renderTo: 'ask_weights',
        text: 'Go!',
        handler: function(){window.location = "{%url
geonode.maps.views.maps_search_page%}";
    });
{% endautoescape %}
});
</script>
{% endblock %}

{% block main %}
<div class="twocol">
    <h2>{{subtopic.title}}</h2>
    <p>Please define your own weights between 0 and 10!</p>
    <p>When you're done, press the "Calculate" button in order to get the
dataset with best quality</p>
    <form action="/quality/weights" method="GET">
        <fieldset>
            <legend>Quality Indicators:</legend>
            <table>
                <tr>
                    <td>Geographic Extent</td>
                    <td><input type="text" size=3
name="geographicExtent"></td>
                </tr>
                <tr>

```

```

        <td>Licensing and Constraints </td>
name="licensingConstraint"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Scale Denominator</td>
name="scaleDenominator"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Update</td>
name="update"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Temporal Extent</td>
name="temporalExtent"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Fitness for Use</td>
name="fitness4use"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Thematic Richness</td>
name="thematicRichness"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Integration</td>
name="integration"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Data Integrity</td>
name="dataIntegrity"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Positional Accuracy</td>
name="positionalAccuracy"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Thematic Accuracy</td>
name="thematicAccuracy"></td>
        <td><input type="text" size=3
</tr>
        <tr>
        <td>Completeness</td>
name="completeness"></td>
        <td><input type="text" size=3
</tr>

```

```

        <tr>
            <td>The subtopic code is</td>
            <td><input type="text" size=3
name="subtopic_pk" value={{subtopic.pk}} readonly="readonly"></td>
        </tr>
    </table>
    <br>
    <input type="submit" value="Calculate!">
</fieldset>
</form>

```

quality/quality/templates/quality/layer.html

```

{% extends "page_layout.html" %}
{% load geonode_auth %}
{% load i18n %}

{% block title %} {{ layer.resource.title|default:layer.typename }} - {{
block.super }} {% endblock %}

{% block head %}
{% include "geonode/ext_header.html" %}
{% include "geonode/app_header.html" %}
{% include "geonode/geo_header.html" %}

<link rel="stylesheet" type="text/css" href="{{ GEONODE_CLIENT_LOCATION
}}theme/ux/colorpicker/color-picker.ux.css" />

{{ block.super }}

    <script type="text/javascript">
{% autoescape off %}
    var styleEditor, modified = false;
    Ext.onReady(function() {
        var config = {
            tools: [{
                ptype: "gxp_wmsgetfeatureinfo",
                // uncomment the line below if you want feature info in
a grid
                //format: "grid",
                actionTarget: "main.tbar",
                outputConfig: {width: 400, height: 200, panIn: false}
            }],
            proxy: "/proxy/?url=",
            localGeoServerBaseUrl: "{{GEOSERVER_BASE_URL}}",
            authorizedRoles: "{{
user.is_authenticated|yesno:"ROLE_ADMINISTRATOR,ROLE_ANONYMOUS" }}" ,

            /* The URL to a REST map configuration service. This
service
            * provides listing and, with an authenticated user, saving

```



```

of
    * maps on the server for sharing and editing.
    */
    rest: "/maps/",

    portalConfig: {
        renderTo: "preview_map",
        height: 350
    },

    createTools: function() {
        return [new Ext.Button({
            tooltip:
GeoExplorer.prototype.backgroundContainerText,
            iconCls: 'icon-layer-switcher',
            menu: new gxp.menu.LayerMenu({
                layers: this.mapPanel.layers
            })
        })]
    },
    listeners: {
        "ready": function() {
            var map = app.mapPanel.map;
            map.zoomToExtent(map.layers.slice(-
1)[0].maxExtent);
        },
        "beforeunload": function() {
            if (modified) {
                styleEditor.show();
                return false;
            }
        }
    }
};

config = Ext.apply(config, {{ viewer|safe }});

app = new GeoExplorer.Viewer(config);

Ext.get(Ext.DomQuery.select("input[@name='style']")).on("click",
function(evt, elem) {
    app.selectedLayer.getLayer().mergeNewParams({
        "STYLES": elem.value,
        "_dc": Math.random()
    });
});

Ext.get(Ext.DomQuery.select("select[@name='default-
style']")).on("change", function(evt, elem) {
    Ext.Ajax.request({
        method: "post",
        url: "?style",
        params: {"defaultStyle" : elem.value}
    });
});

```

```

    });
  });
  {% has_obj_perm user layer "maps.change_layer" as can_change %}
  {% if can_change %}
    Ext.get(Ext.DomQuery.select(".style-title")).on("click",
function(evt, elem) {

showStyle(Ext.get(elem).prev("input").getAttribute("value"));
  });
  {% endif %}

  {% has_obj_perm user layer "maps.change_layer_permissions" as
can_change_permissions %}
  {% if can_change_permissions %}
    new GeoNode.PermissionsEditor({
      renderTo: "permissions_form",
      userLookup: "{% url geonode.views.ajax_lookup %}",
      permissions: {{ permissions_json }},
      levels: {
        'admin': 'layer_admin',
        'readwrite': 'layer_readwrite',
        'readonly': 'layer_readonly',
        '_none': '_none'
      },
      listeners: {
        updated: function(perms) {
          var submitTo = "{% url
geonode.maps.views.ajax_layer_permissions layer.typename %}";
          Ext.Ajax.request({ url: submitTo, jsonData:
perms.writePermissions() });
        }
      }
    });
  {% endif %}
  });

  function showStyle(styleName) {

    if (!styleName) {
      Ext.ComponentMgr.all.on("add", function(index, object) {
        if (object instanceof gxp.WMSStylesDialog) {
          Ext.ComponentMgr.all.un("add", arguments.callee);
          object.on("ready", function() {
            object.addStyle();
          }, this, {single: true});
        }
      });
    }
    app.tools["styler"].actions[0].execute();
  }
  {% endautoescape %}
</script>
{% endblock %}

```

```

{% block main %}
<div class="twocol">
  <div id="description"> <h3> {{
layer.resource.title|default:layer.typename }} </h3> </div>
  <p> <strong>{% trans "Abstract" %}</strong> {{
layer.abstract|default:_("No abstract for this layer.") }} </p>

  <div id="preview_map"></div>

  <p><strong>GMES Quality Assessment Results</strong></p>

  <table border="5" frame="vsides">
    <tr>
      <th>Indicator</th>
      <th>Score</th>
    </tr>
    <tr>
      <td>Geographic Extent</td>
      <td align='center'>{{qualityRecord.geographicExtent}}</td>
    </tr>
    <tr>
      <td>Licensing and Constraints</td>
      <td align='center'>{{qualityRecord.licensingConstraint}}</td>
    </tr>
    <tr>
      <td>Scale Denominator</td>
      <td align='center'>{{qualityRecord.scaleDenominator}}</td>
    </tr>
    <tr>
      <td>Update</td>
      <td align='center'>{{qualityRecord.update}}</td>
    </tr>
    <tr>
      <td>Temporal Extent</td>
      <td align='center'>{{qualityRecord.temporalExtent}}</td>
    </tr>
    <tr>
      <td>Fitness for Use</td>
      <td align='center'>{{qualityRecord.fitness4Use}}</td>
    </tr>
    <tr>
      <td>Thematic Richness</td>
      <td align='center'>{{qualityRecord.thematicRichness}}</td>
    </tr>
    <tr>
      <td>Integration</td>
      <td align='center'>{{qualityRecord.integration}}</td>
    </tr>
    <tr>
      <td>Data Integrity</td>
      <td align='center'>{{qualityRecord.dataIntegrity}}</td>
    </tr>
  </table>

```

```

        <td>Positional Accuracy</td>
        <td
align='center'>{{qualityRecord.positionalAccuracy}}</td>
    </tr>
    <tr>
        <td>Thematic Accuracy</td>
        <td align='center'>{{qualityRecord.thematicAccuracy}}</td>
    </tr>
    <tr>
        <td>Completeness</td>
        <td align='center'>{{qualityRecord.completeness}}</td>
    </tr>
</table>

{% if layer.display_type %}<p> <strong>{% trans "Type" %}</strong> {{
layer.display_type }} </p>{% endif %}
{% if layer.keywords %}<p> <strong>{% trans "Keywords" %}</strong> {{
layer.keywords }}</p>{% endif %}
{% if layer.edition %}<p> <strong>{% trans "Edition" %}</strong> {{
layer.edition }}</p>{% endif %}
{% if layer.topic_category %}<p> <strong>{% trans "Topic Category"
%}</strong> {{ layer.topic_category }}</p>{% endif %}

{% if layer.constraints_use or layer.constraints_other %}
<p> <strong>{% trans "Citation" %}</strong>
    {{ layer.constraints_use }} {{ layer.constraints_other }}
</p>
{% endif %}

{% if layer.owner %}
<p> <strong>{% trans "Owner" %}</strong>
    <a href="{{ layer.owner.get_profile.get_absolute_url }}"> {{
layer.owner.username }} </a>
</p>
{% endif %}

{% if layer.poc %}
<p> <strong>{% trans "Point of Contact" %}</strong>
    {% if layer.poc.user %}
        <a href="{{ layer.poc.user.get_profile.get_absolute_url }}"> {{
layer.poc.user.username }} </a>
    {% else %}
        <ul>
            {% if layer.poc.name %}<li><strong>{% trans "Name" %}</strong> {{
layer.poc.name}}</li>{% endif %}
            {% if layer.poc.email %}<li><strong>{% trans "Email" %}</strong>
{{ layer.poc.email }}</li>{% endif %}
            {% if layer.poc.organization %}<li><strong>{% trans "Organization"
%}</strong> {{ layer.poc.organization}}</li>{% endif %}
            {% if layer.poc.position %}<li><strong>{% trans "Position"
%}</strong>: {{ layer.poc.position}}</li>{% endif %}
            {% if layer.poc.voice %}<li><strong>{% trans "Voice" %}</strong>
{{ layer.poc.voice}}</li>{% endif %}
            {% if layer.poc.fax %}<li><strong>{% trans "Fax" %}</strong> {{

```

```

layer.poc.fax}}</li>{% endif %}
    {% if layer.poc.delivery %}<li><strong>{% trans "Delivery"
%}</strong> {{ layer.poc.delivery}}</li>{% endif %}
    {% if layer.poc.area %}<li><strong>{% trans "Area" %}</strong> {{
layer.poc.area}}</li>{% endif %}
    {% if layer.poc.zipcode %}<li><strong>{% trans "Zip Code"
%}</strong> {{ layer.poc.zipcode}}</li>{% endif %}
    {% if layer.poc.country %}<li><strong>{% trans "Country"
%}</strong> {{ layer.poc.country}}</li>{% endif %}
</ul>
    {% endif %}
</p>
{% endif %}

{% if layer.attribute_names %}
<p> <strong>{% trans "Attributes" %}</strong>
    {{ layer.attribute_names|join:", " }}
</p>
{% endif %}

{% if layer.supplemental_information %}
<p> <strong>{% trans "Supplemental Information" %}</strong>
    {{ layer.supplemental_information }}
</p>
{% endif %}

{% if layer.data_quality_statement %}
<p> <strong>{% trans "Data Quality Statement" %}</strong>
    {{ layer.data_quality_statement }}
</p>
{% endif %}

{% if layer.geographic_bounding_box %}<p> <strong>{% trans "Bounding Box"
%}</strong> {{ layer.geographic_bounding_box }} </p>{% endif %}
{% if layer.metadata.crsOptions %}<p> <strong>{% trans "Native SRS"
%}</strong> {{ layer.metadata.crsOptions.0|default:_("No SRS specified.")
}} </p>{% endif %}

</div>
{% endblock %}

{% block sidebar %}
<div id="sidebar" class="threecol">
    <h3> {% trans "Download" %} </h3>
    <p> <strong>{% trans "Data" %}: </strong>
        {% for extension, type, link in layer.download_links %}
            <a href="{{link}}">{{type}}</a>
        {% endfor %}
    </p>
    <p> <strong>{% trans "Metadata" %}: </strong>
        {% for mimetype, mdtype, link in layer.metadata_links %}
            <a href="{{link}}">{{mdtype}}</a>
        {% empty %}
            <em> {% trans "No metadata links provided for this data." noop

```

```

%} </em>
    {% endfor %}
</p>
<h3>{% trans "Maps" %} </h3>
{% if layer.maps %}
<p>{% trans "The following maps use this data set" %}:
    <ul>
        {% for map in layer.maps %}
            <li> <a href="{map.get_absolute_url}">{map.title}</a> </li>
        {% endfor %}
    </ul>
{% else %}
    <p> {% trans "This layer is not currently used in any maps." %}
</p>
{% endif %}
<p><a href="{% url geonode.maps.views.newmap
%}?layer={layer.typename}">Create new map</a></p>

    {% has_obj_perm user layer "maps.change_layer" as can_change %}
    {% has_obj_perm user layer "maps.delete_layer" as can_delete %}

    <h3> {% trans "Styles" %} </h3>
    <p> {% trans "The following styles are associated with this data set.
Choose a style to view it in the preview to the left. Click on a style
name to view or edit the style." %}
    <br/>
    <span class="styles-list">
        <input type="radio" name="style"
id="{layer.publishing.default_style.name}"
value="{layer.publishing.default_style.name}" checked="checked"/>
        <label for="{layer.publishing.default_style.name}" class="style-
title"> {{ layer.default_style.name|title }} </label>
        <a href="{ layer.default_style.body_href }">SLD</a><br/>
        {% for style in layer.styles %}
            <input type="radio" name="style" id="{style.name}"
value="{style.name}"/>
            <label for="{style.name}" class="style-title"> {{
style.name|title }} </label>
            <a href="{ style.body_href }">SLD</a><br/>
        {% endfor %}
    </span>
    {% if user.is_authenticated and can_change %}
    {% trans "Default style:" %}
    <select name="default-style">
        <option value="{layer.publishing.default_style.name}"
selected="selected">
            {{layer.publishing.default_style.name|title}}
        </option>
        {% for style in layer.styles %}
            <option value="{style.name}"> {{ style.name|title }} </option>
        {% endfor %}
    </select><br/>
    <a href="javascript:showStyle()">Create new style</a>
    {% endif %}

```

</p>

```

{% if user.is_authenticated %}
{% if can_change or can_delete or can_change_permissions %}

    <h3> {% trans "Manage" %} </h3>
    <ul>
        {% if user.is_authenticated %}
        {% if can_change %}
            <li><a href="?describe">{% trans "Update the description of
this data" %}</a></li>
            <li><a href="?update">{% trans "Upload a new version of
this data" %}</a></li>
        {% endif %}
        {% if can_change_permissions %}
            <li><a href="{% url edit_layer_permissions layer.typename %}">{%
trans "Edit Permissions" %}</a></li>
        {% endif %}
        {% if can_delete %}
            <li><a href="?remove">{% trans "Remove" %}</a></li>
        {% endif %}
        {% else %}
            <li>Please authenticate to update or remove
layers</li>
        {% endif %}
    </ul>
    {% endif %}
    {% has_obj_perm user layer "maps.change_layer_permissions" as
can_change_permissions %}
    {% if can_change_permissions %}
    <h3>{% trans "Permissions" %}</h3>
    <div id="permissions_form"></div>
    {% endif %}
    {% endif %}
</div>
{% endblock %}

```

quality/quality/templates/quality/subtopics.html

```

{% extends "page_layout.html" %}
{% load i18n %}

{% block title %} {% trans "Browse Data" %} - {{ block.super }} {% endblock
%}

{% block head %}
{% include "geonode/ext_header.html" %}
{% include "geonode/app_header.html" %}

```

```

{{ block.super }}
<script type="text/javascript">
Ext.onReady(function(){
{% autoescape off %}
    var submitForm = function() {
        Ext.query('#search_form form')[0].submit();
    };
    var searchButton = new Ext.Button({renderTo: 'submit_search', text: '{%
trans "Search" %}',
                                     handler: submitForm});
    var searchInput = new Ext.form.TextField({applyTo: 'search_query', width:
350});

    var goButton = new Ext.Button({
        renderTo: 'ask_weights',
        text: 'Go!',
        handler: function(){window.location = "{%url
geonode.maps.views.maps_search_page%}";
    });
{% endautoescape %}
});
</script>
{% endblock %}

{% block main %}
<div class="twocol">

    <h2>This is the list of the data sub-topics available in the
server</h2>
    <p>Please select your subtopic of interest in order to get the dataset
of best quality!</p>

    <form action="/quality/ask4weights" method="GET">
        <fieldset>
            <legend>Available sub-topics</legend>
            {% for oneSub in allSubs %}
                {% if oneSub.layersubtopic_set.count > 0 %}
                    <input type="radio" name="subtopic"
value={{oneSub.pk}}/>{{oneSub.title}}
                    <br>
                {% endif %}
            {% endfor %}
            <input type="submit" value="Next">
        </fieldset>
    </form>

</div>
{% endblock %}

{% block sidebar %}
    {% block sidebar_info %}
    {% endblock %}

{% endblock %}

```


Annex B - Malawi Disaster And Natural Resources Spatial Data Sharing Platform: Technical Proposal

Technical Approach and Methodology

GeoNode and Roadmap

In recent years several studies have been conducted in African countries in order to investigate the vulnerability to natural hazards also in relation to the effects of climate change. Geospatial data containing information on the vulnerability, impacts, climate change, disaster risk reduction and preparedness is usually part of the output of such studies. This geospatial data is much useful to governments and other important stakeholders involved in building resilience against natural disasters; nevertheless this data often is not widely distributed and readily accessible to the same stakeholders and often resides in remote storage solutions hardly reachable.

Spatial Data Infrastructures are technical solutions capable to solve this issue, by storing geospatial data and making them widely available through the internet. Among these solutions, GeoNode, an online platform for geospatial data sharing, that has been developed in recent years appears to be an excellent one.

GeoNode is a platform for the management and publication of geospatial data. It brings together mature and stable open-source software projects under a consistent and easy-to-use interface allowing users, with little training, to quickly and easily share data and create interactive maps.

GeoNode data management tools allow for integrated creation of data, metadata, and map visualizations. Each dataset in the system can be shared publicly or restricted to allow access to only specific users. Planned social features like user profiles and commenting and rating systems will allow for the development of communities around each platform to facilitate the use, management, and quality control of the data the GeoNode instance contains [geonode.org].

The GeoNode community is very active in proposing new improvements and functionalities. The current development roadmap is based on Improvement Proposals (GNIPs), where the core

committers officially track and explain the desired features. The major GNIPS, in terms of new features, now proposed are:

- multiple metadata format management and export
- add group management and group permissions
- feedbacks and user rating on the content with social features
- resources collection on particular topics
- improved search
- GEORSS visualization

Besides the official improvement proposals there are also a relevant number of GeoNode forks with branches created by people working on GeoNode. Several functionalities such as editing, document managements, social features, improved metadata are already implemented as extensions or upgrades to GeoNode, even if they are not included in the main release.

GeoNode state-of-the-art

On February 6th, 2012 the GeoNode community released version 1.1 of the platform. The community focused their work on system stability and bug fixing since the previous release and provided a reliable version of the platform.

The GeoNode is the reference software of the GFDRR Open Data for Resilience Initiative, that aims to reduce the impact of disasters by empowering decisions-makers with better information and the tools to support their decisions. Several GeoNode instances are running in countries such as Haiti, Grenada, Saint Vincent and the Grenadines, Saint Lucia in the Caribbean region, in Kenya and in the Horn of Africa Region. This tool enables individuals and organizations to download as well as contribute with their own data, thereby promoting collaboration around the management of the information needed to rebuild Haiti [<http://www.gfdr.org/gfdr/opensdr>]

Country focus: Malawi

Landlocked Malawi is among the world's least developed and most densely populated countries. With a poverty headcount of 39% , Malawi faces severe and widespread poverty conditions that, combined with the high incidence of HIV/AIDS (12%), the high percentage (85%) of the population living in rural areas with poor access to basic health and education services, and the critical land use and its derived increased erosion, lead to a grave vulnerability of the poor to the impacts of natural disasters.

As a Highly Indebted Poor Country, Malawi therefore represents a priority in the GFDRR Country Programming Framework and specific support and operations have been undertaken towards disaster risk reduction.

In particular, in the last years GFDRR has been cooperating with the Government of Malawi to strengthen disaster management strategies by providing technical assistance for the elaboration of national development policies. Quite a few studies and analyses have been already conducted in order to contribute to flood risk reduction in the Shire River basin and the large amount of spatial

data collected suggests that a proper data sharing platform be developed and implemented for an optimal exploitation of the retrieved information.

In the framework of the World Bank Open Data for Resilience Initiative (OpenDRI), which fosters data sharing for disaster vulnerability reduction, data accessibility is considered a powerful tool to support governmental decisions. In this context, a data sharing tool based on GeoNode is to be implemented and delivered to the members of the Shire River Basin Management Program Technical Taskforce.

Workplan

The main activities that will be performed under this assignment will be aimed at fulfilling the requirements set out in the Terms of Reference.

Here below follows a detailed description of each single proposed activity.

Activity I - Installation and testing

In accordance with Task 1 of the TORs the firm will perform the installation of a GeoNode instance on a server managed by the Malawian National Spatial Data Center. The installation activities will follow a discussion with the personnel of the mentioned Center, in order to get information about the IT management and hardware availability, and a subsequent support in the production configuration. Particular attention will be devoted to the choice between a cloud computing instance or a devoted server; in both cases support will be given in cloud instance solution or hardware choice.

All necessary tasks and operations of network configuration will be performed in order to have the server reachable in the local area network and on the internet.

The installation will be performed on a Linux based operating system, preferably on a fresh Ubuntu OS. The GeoNode installation requires the availability of different software on the server, such as a Python interpreter, a Java Runtime Environment, an HTTP server and a Java server applet; therefore the installation procedure will be tailored on the basis of the presence of the aforementioned software requirements on the hosting server. Accordingly either the Personal Package Archive prepared by the GeoNode community or a check-out of the GeoNode official code repository will be installed. All the software installed on the server will be Free and Open Source.

After the installation, the necessary steps for the server configuration will be performed, such as the definition of the correct communication ports according to potential local firewalls and IP tables, the definition of the correct paths/folders for GeoServer data storage and the definition of the necessary Django settings.

Finally a test session will follow, based on a double approach: on one side the Unit Test Suite developed by the GeoNode community will be run in order to detect and solve potential errors and failures; on the other hand the main GeoNode functionalities, such as data uploading/downloading, map generation, user/group creation and data permission respect, will be tested in order to verify the correct behavior. Potential issues detected during the test phase

will be fixed by the firm and/or, if interesting for the entire GeoNode community, reported to the GeoNode developers through the official channels.

Activity 2 - Database population

The second activity that will be performed right after having installed GeoNode is the data import. Different kinds of data will be uploaded onto the database, namely Local Data, Open Source Data and Value Added Data.

Local Data

During the Malawi deployment ITHACA will work together with the local National Spatial Data Center staff to populate the GeoNode platform with available datasets. The activities can be described as follows:

- Uploading identified spatial layers. A preliminary data assessment will be performed. If it won't be possible to complete the task due to the amount of data, a shared working plan will be established in order to cover all the possible data formats issues. It is envisioned that during a 5 day work ITHACA will train local staff in order to make them independent users and possibly complete this task by their own. In particular, it will be taught to manage all the different type of data formats in GeoServer. Ensuring metadata is entered into the platform, in compliance with World Bank GeoSpatial standards (based on ISO 19115: 2003). A specific training will be performed on GeoNetwork metadata management in combination with GeoNode.
- Styling of uploaded datasets using the SLD format. The applied training will focus also on OGC filtering and SLD Styling (possibly also using the CSS styling plug in) and GeoServer extensions.
- Creation of a suite of basic mashups or maps with data available on the platform in collaboration with local staff.

OS Data

The free and open access to spatial data can be a key factor in many countries developing process. Licensing restrictions are barriers that can hinder or prevent the local economic growth and the knowledge of the territory.

ITHACA performed an analysis of non-European reference data availability (data inventory, data modeling and published on an OGC compliant catalogue), quality and consistency. The project, titled "Implementation of an initial GMES (Global Monitoring for Environment and Security) Service for Geospatial Data Access covering areas outside Europe", was funded by the European Commission.

Reference data are vital to end-user applications in order to provide the basic geographic framework on top of which additional (thematic) spatial information can be produced and disclosed (e.g. land use/land cover maps, asset maps and damage assessment maps in response to crisis). Reference data are typically provided by data providers at regional or national level. In

order to take full benefit from the data available a central service is needed to access these reference data providers in a unified manner.

ITHACA will provide a subset (Malawi bound) of the best evaluated public reference and thematic data covering the following topics:

ITHACA will provide a subset (Malawi bound) of the best evaluated public reference and thematic data covering the following topics:

- Boundaries: costal, administrative and areas of special interest
 - Country Political Level 0 Boundaries
 - Sub-national Boundaries (from 1st Level on)
 - Areas of conflict and Disputed Boundaries, if any
- Parks, conservancies, and Protected areas
- Surface hydrology: surface water bodies, water points, drainage and watersheds
 - Drainage, Rivers and Flow Routing
 - Lake Malawi, Lake Malombe, Lake Chilwa and other minor Surface Water Bodies
 - Watersheds and River Basins
 - Water Points
- Human population: population centres and distributions
 - Population Centers;
 - Urban and Rural Population Density and Distribution
- Transportation: roads, railways, airports, harbors and nautical routes
 - Roads;
 - Railway lines, Stations, and Marshalling Yards;
 - Airport;
- Bathymetry and terrestrial elevation
 - Bathymetry;
 - Terrestrial Elevation

The open data will be stored in a PostgreSQL / PostGIS, loaded as layers in GeoServer and stylized according to scale and feature classes and then will be made available in GeoNode. All the datasets will be uploaded with detailed metadata according to CSW standards. Metadata will be available both in GeoNode (core metadata) and in GeoNetwork (ISO metadata).

Three Malawi reference maps (based on road network, physical features and satellite images) will be created in GeoNode as a starting point for further maps.

Value added data

In addition to the former types of data that will be imported onto the system, Ithaca will make available to GeoNode value added data coming from different sources: the Extreme Rainfall Detection System and GIO-EMS mapping in rush mode.

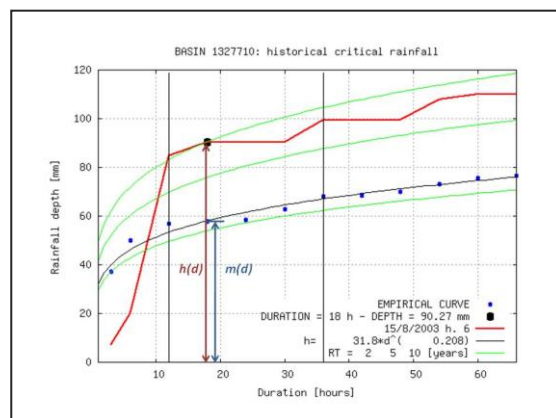
EXTREME RAINFALL DATA SYSTEM

Ithaca has been active in the field of flood risk, especially in the preparedness phase of the emergency cycle, for a few years now, due to the development of the so called Extreme Rainfall Detection System; this is a simplified and efficient Early Warning System for flood events.

In particular, the ERDS system is able to give an alert in advance about the occurrence of floods around the world monitoring the heavy rainfalls in near real time and it is currently used by WFP (World Food Programme) and other humanitarian assistance organizations to assess event magnitude and to identify potentially floodable areas in places where their assistance is needed. Alerts detected by this system are distributed through the Humanitarian Early Warning Service (<http://www.hewsweb.org/floods/>).

In details the method is based on precipitation analysis and it uses rainfall data from satellite at worldwide extent. This system works on river basins scale with global coverage by using satellite rainfall data from Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA).

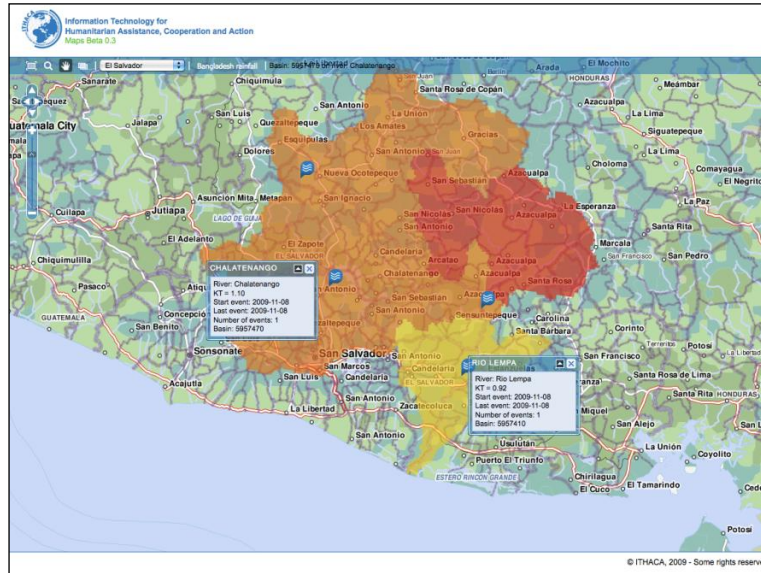
On one hand The 3-hourly data from 1998 are used to detect all historical flood events in the past ten years using a hydrological method based on Depth Duration Frequency (DDF) curves. The final output of this procedure is the creation of a complete historical database of heavy rainfalls with global coverage and single river basin resolution.



Depth Duration Frequency (DDF) curves with different return time are represented in green, cumulated rainfalls are represented in red and critical rainfall, detected through the comparison with historical values, are shown by the black dot.

On the other hand the same analysis is performed using near real-time data to monitor current rainfall conditions and to make alerts in near real time: it can be considered the proper “Early Warning System”. The 3 hourly real time rainfall data (product) are compared with the DDF curves, previously determined using historical data.

In conclusion the outputs consist of thematic maps that show the areas in which heavy rainfalls are occurring. River basins are shown in different colors indicating the level of criticality, according to the severity of rainfalls. The outputs of the ERDS systems, clipped on the extent of the country of Malawi, in case the Malawian catchments will be alerted by the system will be uploaded into the GeoNode platform object of the contract, during the period of the contract assignment.



Example of the output map: the alerted basins are shown with 3 different possible colours, according to the severity of rains.

GMES INITIAL OPERATIONS EMERGENCY MANAGEMENT SYSTEM – MAPPING IN RUSH MODE

As previously mentioned, Ithaca has been awarded two European Commission contract for the services “Mapping in rush mode” and “Mapping in non-rush mode”. The services concern the on-demand fast provision of geo-spatial information in support to emergency management activities both immediately following an emergency event and during the phases of the emergency management cycle which are not related to the immediate response, i.e. prevention, preparedness and reconstruction activities.

In the rush mode service, the activations will be launched for relevant disastrous events hitting the European continent but also the rest of the globe. The spatial data regarding the country of Malawi which will be potentially acquired and derived in this context will be shared on GeoNode thus providing a high added value to the system. Similarly, in case of non-rush activations, Ithaca will share the results of the spatial analyses when activated.

Activity 3 - GeoNode customization

Layout and Styling

The customization is necessary to tailor the instance to fit the user needs.

The interface customization is achieved by modifying the GeoNode HTML templates. The changes should involve:

- header with custom colors and logo
- footer with references
- page layout with desired information displayed as well as application shortcuts.

The template customization can involve each GeoNode page and not just the home page. The application style will be managed through the use of Cascading Style Sheets.

The background customization is achieved by modifying the client side logic and the server side logic.

The client side is based on the Javascript language and is composed by different tools such as Geoexplorer, Gxp and GeoNode specific customization. The client side logic is used for data/maps search and for the map composer/viewer.

The server side logic is Python based and is the core of GeoNode, customization on this side can be done as extensions that add features without modifying the core code or as custom implementations of the source code.

Videos, pictures and documents

Ithaca is the developer of a GeoNode extension called GeoNode-Documents, <https://github.com/simod/geonode-documents> - readme.

This extension enables document management in GeoNode - where “document” is intended any kind of non spatial information - and provides a form that allows the user to upload a file and link it to an existing GeoNode map (or creating one on the fly if not available). The map info interface shows the list of the related documents ordered by extension which can be accessed or downloaded. The documents have their specific permission layer.

As part of the customization required in task 1, the GeoNode-Documents extension will be used as a base tool for enabling videos and picture management. The videos and pictures, once uploaded, will be linked to one or more maps and will be shown in the map info page. They can be downloaded or viewed through links and it will also be possible to view each map related to the document.

Activity 4 - Training and documentation

In accordance with Task 3 of the TOR, two separate training sessions in Lilongwe are envisioned. The first training session regards basic topics both on the user and on the administrator side.

- The basic training on the End User side addresses the following topics:
 - Overview of the software
 - Map and data viewing
 - Data uploading (with focus on data types, Coordinate Systems, ISO standard for metadata)
 - Map creating (focus on Styling, printing and web publication)
 - Data downloading (focus on various available formats)
 - Assigning permissions to maps and data
- The basic training on back end and system administration side addresses the following topics:
 - Overview of the main components installed with GeoNode (Tomcat server applet, Apache webserver)
 - Overview of GeoNetwork (focus on interactions with GeoNode)
 - GeoServer main functionalities and objects
 - Django Admin interface (focus on users and groups management and DBMS interaction)
 - Overview of PostgreSQL and PostGIS and data storage

The second training session regards advanced topics both on the user and on the administrator side:

- The advanced training on the End User side addresses the following topics:
 - Interaction between the different GeoNode components
 - SLD – Style Layer Descriptor in xml format and examples
 - SLD generation through offline tools such as Atlas Styler, ArcMaptoSLD
 - WMS, WCS, WFS web services for data publication
 - map generation by means of third part web services
 - use of GeoNode Documents application
- The advanced training on the back end and system administrator side addresses the following topics:
 - Django web framework overview and MVC pattern
 - GeoNode command line (focus on batch data upload and system synchronization)
 - GeoNode look and feel customization
 - GeoNode troubleshooting: reading and understanding log files
 - GeoNode main code modules and development howto

The trainings are held with hands-on and visual support: all the documentation produced with the objective of the trainings will be distributed to the attendant in open formats, such as pdf. The production of the trainings material will start from the GeoNode online documentation, that will be integrated and extended in order to match the TOR specifications. All the materials produced in the framework of the present project proposal will be made available to the GeoNode community in order to facilitate future work and trainings.

Activity 5 – Maintenance and support

As required in the TOR, Ithaca will provide 10 hours per month for 5 months starting from July 1, 2012 of specific remote technical assistance, bug-fixing and fine-tuning of the system.