UN WFP SPATIAL DATA INFRASTRUCTURE

A. Ajmar [a], P. Boccardo [b], O. Terzo [c]

[a] ITHACA, via P.C. Boggio 61, 10138 Torino, Italy - andrea.ajmar@ithaca.polito.it
[b] DITAG - Politecnico di Torino, Corso Duca degli Abruzzi 24, 10135 Torino - piero.boccardo@polito.it
[c] Istituto Superiore Mario Boella, via P.C. Boggio 61, 10138 Torino, Italy - terzo@ismb.it

Inter Commission WG IV/VIII

KEY WORDS: Cartography, GIS, Technology, Management, Interoperability, Global-Environmental-Databases, Reference Data, Automation

ABSTRACT:

UN World Food programme (WFP) is implementing its own Spatial Data Infrastructure (SDI), in the framework and following the guidelines developed by United Nations Geographic Information Working Group (UNGIWG) with the UNSDI initiative. This paper will present an updated situation report of WFP SDI implementation, project realized with the technical support of ITHACA association. Geodatabase content has been structured in a specific data model developed in UML. The geodatabase content includes reference and transportation datasets; this second component is specifically conceived to efficiently store all objects and attributes needed for logistic assessment, one of the WFP mandates. Especially for the use during emergencies, up-to-date information, system availability, redundancy and performances are key factors; for that reason WFP SDI architecture is based on a federated geodatabase structure, with replicas distributed at regional and local offices. Furthermore, streaming features are used to create and maintain aligned a set of read-only geodatabase instances; streaming environment is created and set to maintain geodatabase integrity, without having to establish strict constraints on the configuration of the different nodes. The objective is to realise a system made of several distributed instances, possibly on different DBMS platforms, automatically aligned with the master site, granting system redundancy, network load reduction and increasing performances through the implementation of load balancing procedures. The proposed solution would be able to grant the maximum level of availability, especially during emergencies and for users operating in difficult environmental condition and with different hardware/software configurations. Synchronization procedures provide the capability of keeping all nodes aligned and up-to-date, simplifying data management processes.

1. INTRODUCTION

In the frame of humanitarian operations, the United Nations *World Food Programme* (WFP) is in the front line. The WFP is the biggest UN Agency and responds to more than 120 emergencies per year worldwide. WFP operations reached just over 102 million people with food assistance in 2008. The beneficiaries were in 78 countries. As the United Nations frontline agency in the fight against hunger, WFP is continually responding to emergencies; WFP also works to help prevent hunger in the future through programmes that use food as a means to build assets, spread knowledge and nurture stronger, more dynamic communities. According to the UN reform, WFP is also the leader of logistics for UN and international bodies during emergency response operations. Logistic coordination is in charge of the UN Logistic Cluster, a group of humanitarian organizations and other stakeholders committed to commonly address logistics needs during humanitarian crises on a broad partnership basis. WFP is the lead agency for Logistics Cluster.

In 2008, WFP total expenditure was 3.72 billion US$, with 93% of its funding gone either to cover food costs or to pay for its transport.

WFP is a worldwide distributed organization, having operating offices dislocated in 79 different countries, with a hierarchical subdivision in Regional Bureaux and Country Offices; WFP employed 10,200 people in 2008, with 91% of staff serving in the field (Figure 1)
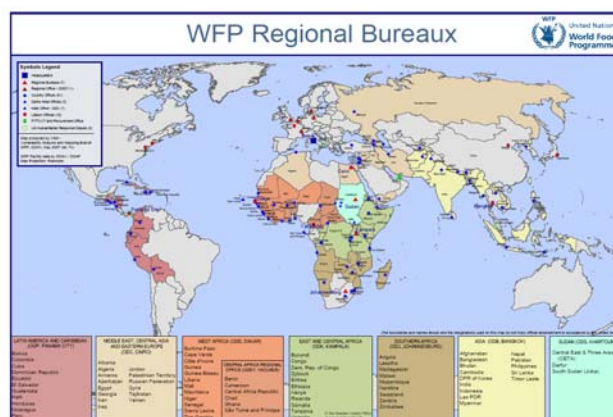

Figure 1 - WFP Regional Bureaux (red triangles) and Country Offices (blue dots) dislocation.

The production and use of geospatial information within the United Nations has been accomplished historically by its component organizations in accordance with their individual needs and expertise. This has resulted in multiple efforts, reduced opportunities for sharing and reuse of data, and an unnecessary cost burden for the United Nations as a whole. In that framework, UNGIWG (UN Geographic Information Working Group) has developed recommendations for implementing Spatial Data Infrastructures (SDI) at UN level (UNGIWG, 2007).

The term SDI is often used to denote the relevant base collection of technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data (Executive Office Of The President - Office of Management and Budget, 2002). A spatial data infrastructure provides a basis for spatial data discovery, evaluation, download and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and the general public. Spatial data infrastructures facilitate access to geographically-related information using a minimum set of standard practices, protocols, and specifications. Spatial data infrastructures are commonly delivered electronically via internet (Groot and McLaughlin; 2000, Masser, 2005; Nebert, 2004).

The ITHACA (*Information Technology for Humanitarian Assistance Cooperation and Action*) association is supporting UN WFP in developing and implementing an SDI as solution for several issues, related to distributed management and exploitation of spatial data, among them:
- inconsistent data in terms of content and format;
- existence of "invisible" data, not computerized or hidden in local computers;
- confidentiality and sensitivity of certain data and information;
- difficulties in implementing data/systems integration;
- poor application of standards;
- lack of extensive and reliable metadata catalogues;
- lack of streamlining of spatial analysis in decision making;
- unproductive competitive practices.

2. THE STAKEHOLDER NETWORK

The stakeholders involved in the project are (Figure 2):
- **International organizations**, that have a strong connection with *master node design and implementation team* and with *HW/SW developers and vendors*:
  o Open Geospatial Consortium (OGC), in charge of the definition of standards for formats and services related with geographic informations;
  o United Nations Geographic Information Working Group (UNGIWG), responsible for defining guidelines for SDI implementation at UN level.
- **HW/SW vendors**, that develop and distribute infrastructure components used at central level (*master node* and *replica node*)and at *local level*, according to *international standards*;
- **Geodatabase and services provider (master node ITHACA)** is the node containing the master version of the geodatabase and that provides *web based services* (visualization, analysis, editing, etc.) for all other network nodes (*replica* and *locals*). Infrastructure (*hardware and software*) is designed taking into account *international standards*;
- **Geodatabase and services provider (replica node WFP)** is the exact replica of the master node, for granting better overall system performances and reliability;
- **Internet service providers** are the entities controlling the networks used for sharing information and are the base infrastructure allowing communications between central nodes (*master* and *replica*) and the *local nodes*. Their infrastructure is more and more the preferred way to distribute analysis results to the *citizens*, for managing emergency operations. They have normally strong

relationships with national *governments*, as they control strategic services;
- **Governments and Governmental Authorities** have strong link with international aid agencies (*UN agencies*, *NGO's*, etc.) as any operations be requested and receive the approval of local authorities. Governments are also the major donors for *WFP*, and in general of UN agencies, controlling in such way their activity. Governments are the reference for *citizens*, especially during humanitarian emergencies. Governments may use their control power over *Internet service providers* in order to control information flows. Governmental Authorities are frequently the geographical information providers and they establish rules in accessing data (as instance rainfall information).
- **Local users** are WFP and NGO's officers in charge of managing emergency preparedness and response. They need to access, through the *web*, to the geodatabase and related services provided by central nodes (*master* and *replica*) in order to take appropriate and timely decisions based on geographic base data and analysis made on current event. They act in support and strict collaboration with local *governments*. They interact directly with affected *citizens*, that needs help and recovery actions.
- **Citizens** are keep informed and their condition monitored, possibly through *internet* based services, by *local users* and *government* authorities.
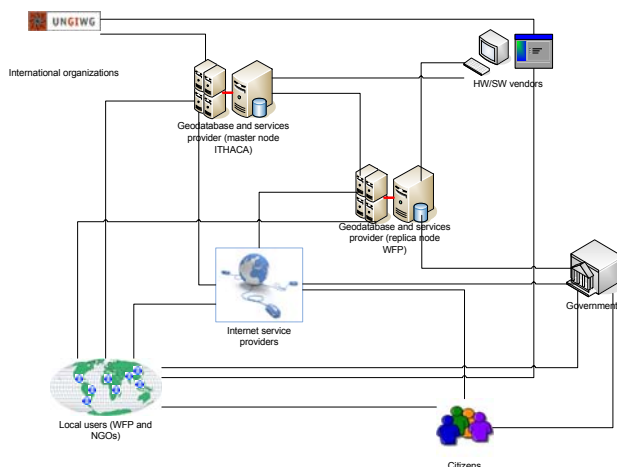

Figure 2 - The representation of the WFP SDI stakeholder network.

3. WFP SDI GEODATABASE

The geodatabase schema is composed by 2 main entities:
- **Reference data**: according with the global extension of the areas of interest for WFP activities, globally consistent data have been selected to be included. Moreover, particular attention has been paid to identify and acquire geographic data classified as public domain; these data cannot be copyrighted, restricted or licensed once they are released, allowing their right use and re-use, at least for no-profit purposes;
- **Transportation**, with particular attention to sea and airlift capacity and availability, transport procedures and schedules, infrastructure assessments and updates, customs issues and logistics bottlenecks affecting the humanitarian community. Transportation dataset data collection is a direct responsibility of UN Logistic Cluster and derives from very different sources, such as pre-existing

geodatasets, digitalization of paper sources, field missions with GPS, reports, interviews, etc.

A logical relational database model has been implemented, using UML as modelling language. UML model has been then integrated with specific geographic behaviour (i.e. topology, network, raster catalogues) that cannot be universally expressed in a UML design. A component of the UML model is shown in Figure 3.
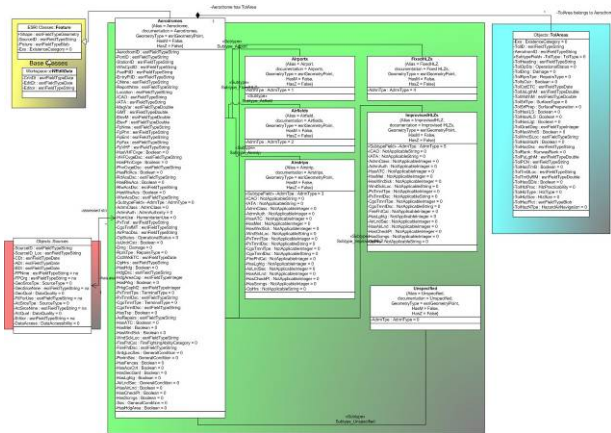

Figure 3 - UML component for aerodromes (Transportation).

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 (WFP units);
- feedback collection and model amendment.

In the designing process structural aspects, dynamic aspects and quantitative aspects are studied. The structural aspects are related to the structure of the concrete knowledge and the abstract knowledge. The dynamic aspects concern the procedural knowledge. Finally, the quantitative aspects are related to the information that describe quantitatively the represented facts and the way they evolve over time. The collection of this information is important in the logic and physical design phases to make the best use of the DBMS and to ensure that applications can be made with the desired performance. Table 1shows a summary of the aspects on which the different stages are focused.

	Needs assessment	Conceptual design	Logical design	Physical design
Structural aspects	Yes	Yes	Yes	Yes
Dynamic aspects	Yes	Yes	Yes	Yes
Quantitative aspects	Yes	No	Yes	Yes
DBMS data model	No	No	Yes	Yes
DBMS features	No	No	No	Yes

Table 1 - Different aspects considered during design steps. (Albano & al., 2007)

4. WFP SDI ARCHITECTURE

The solution for the final architecture has been defined in order to efficiently execute complex functionalities for spatial data management, to maintain big volumes of data and to grant good performances in a distributed environment. Trying to provide a suitable solution for all the above mentioned elements, two parallel environments, the production area and the publication area, has been conceived and implemented. This two-levels architecture has the following major advantages:
- to allow to use built-in and ready available solution from existing applications, that are platform already well known and widely used inside WFP;
- to increase performances by splitting the production and publication environments and by defining specific configurations and tuning options;
- to study new features in order to implement a progressive porting of the geodatabase from a commercial to a non commercial DBMS.

The production layer is implemented mainly using commercial software, where the publication layer developed completely in an Open Source environment.

4.1 Solutions for High-Availability

In order to provide system redundancy, network load reduction and increased performances through the implementation of load balancing procedures, maintaining several distributed instances automatically aligned with the master site, has been considered as a solution. The objective is to implement an active-passive clustering, where the secondary nodes are in hot standby mode, possibly activated in full transaction database mode in case of failure of the master node. Failover procedures automatically switches requests to the standby node, in most cases transparently to the user. The standby node is also used for read-only processing, incrementing overall system performances.

The approach based on the implementation of Oracle Stream feature has been tested in a pre-production environment of WFP SDI architecture (Figure 4). During the first phase of a test, a clone of the WFP SDI master server has been configured and a workstation has been used as downstream node; both machines were resident in the same network.
Several test where executed in order to find the best streaming configuration. It was defined witch schemas must be managed and that both Data Manipulation Language (DML) and Data Definition Language (DDL) should be maintained.
After starting the streaming process, several tests were made in order to verify if propagation and management of data was managed correctly and the efficiency of the solution. We verified that both DML and DDL made on tables associated with feature datasets, feature classes and stand-alone tables were correctly managed.
A still open issue is the one related with the management of big burst of data, occurring when some user executes a massive data loading. In that case the stream process, with the available configuration, is not capable to manage the flow of data. Testing are now in execution in order to understand if with more powerful hardware configuration that problem may be solved.
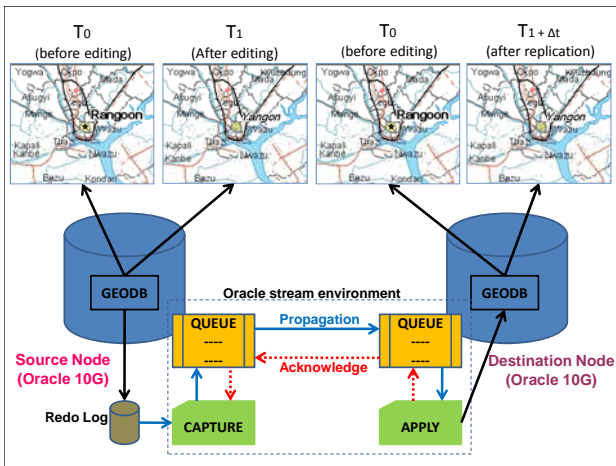
Figure 4 – WFP SDI streaming environment and its major component invoked during an update process (i.e the downgrading and renaming of Rangoon/Yangoon city, Burma).

4.2 Remote nodes

WFP acts in support of local governments, in function of specific help requests; evidently, those requests come essentially from less developed countries, needing for international support for facing major calamities. Less developed countries are also the ones with less effective Internet infrastructures and low penetration rates; mean Internet penetration rates in countries where WFP operates it's 15% of the population, against a world mean of 23.5% (source: internetworldstats.com, 2009).

Internet availability, performances and service reliability are then a major issue while designing an infrastructure that wants to deliver services through the web. It's important also to notice that population living in those countries reaches 80% of the total world population.

WFP Regional Bureaux are connected to Internet through local Internet providers, offering different service levels (Table 2).

WFP Regional Bureaux	City	Internet speed
ODB	Bangkok	512 Kbps
ODC	Cairo	1 Mbps
ODD	Dakar	1 Mbps
ODDY	N'Djamena	None
ODJ	Johannesburg	384 Kbps
ODK	Kampala	256 Kbps
ODP	Panama City	2 Mbps
ODS	Khartoum	1 Mbps

Table 2 - Internet speed for WFP Regional Bureaux

According to above mentioned limitations, an effective infrastructure must include operational solutions for disconnected environments, providing tools for replicating part of the geodatabase content and for keeping synchronized master and replicas. The final architecture becomes a federated geodatabase, with local geodatabase instances hosting subset of the global database covering the territory of pertinence of each office (Figure 5). Single offices will also be identified as responsible for maintaining and updating data content for their territory.
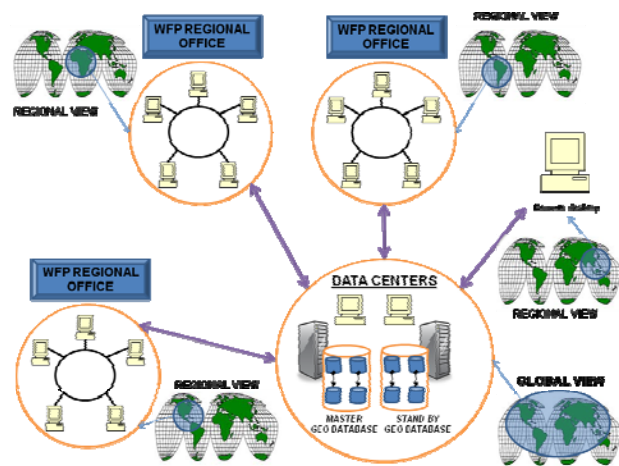

Figure 5 - Schematic representation of a federated geodatabase structure.

5. METADATA CATALOGUE

GeoNetwork is the opens source solution adopted for maintaining the WFP SDI metadata catalogue. An ITHACA GeoNetwork node has been installed and configured. WFP SDI layers are registered in the metadata repository, allowing user to search for available content. The WFP SDI GeoNetwork node is directly accessible through the ITHACA web site (http://www.ithacaweb.org/geonetwork/srv/en/main.home).

6. DEVELOPMENT OF MAP TEMPLATES

While responding to an emergency, outputs of an early-impact analysis are normally in shape of cartographic products. As for the extraction of value added information, the availability of precise and accurate reference geographic datasets is essential to effectively communicate the results of the analysis. Data should be represented following cartographic rules, both in relation with output map scale and with representation rules.

The implementation of an SDI has the general objective to provide to the user tools for spatial data discovery, evaluation, download and application, together with pre-defined representation rules established and shared at international level. Ancillary map elements, such as title, legend, scale and north bar, graticules, overviews, disclaimers etc., can be automatically adapted and updated in function of map scale, content and represented area. The result of that automatization process is the preparation of a map template, ready to incorporate the output of the analysis process and minimizing the effort for managing all ancillary but essential components of the cartographic output. ITHACA has developed a personalization of the mapping environment that allows to significantly reduce time for map preparation and increase overall quality of the final products (Figure 6).
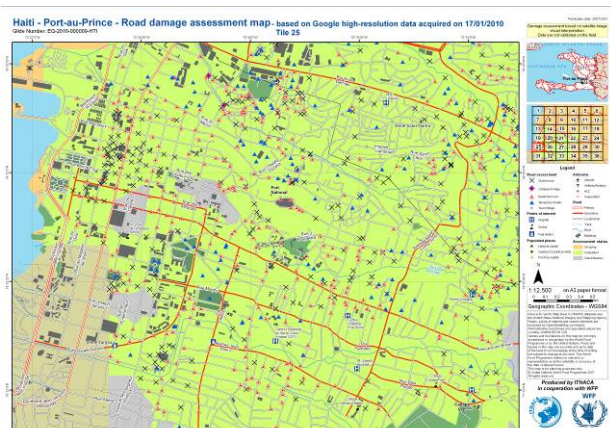
Figure 6 - Example of an Early-Impact analysis represented using ITHACA map template: road damage assessment after the Haiti earthquake, January 2010.

7. WEB SERVICES AND APPLICATIONS

The physical implementation of the WFP SDI allows to start developing Web services and applications, general purposes or specifically conceived to answer to specific needs and execute particular tasks.

Upon the storage level, a Server layer will perform the delivery of OGC web services. The GeoServer open source will be used to implement the OGC server layer. Designed for interoperability, it publishes data from any major spatial data source using open standards. GeoServer is the reference implementation of the Open Geospatial Consortium (OGC) Web Feature Service (WFS) and Web Coverage Service (WCS) standards, as well as a high performance certified compliant Web Map Service (WMS). GeoServer forms a core component of the Geospatial Web. Apache (web server) and Tomcat (servlet container) will be used to expose business logic and business processes for use by GeoServer. An optimization layer for the WMS protocol will be used to improve the map serving across the web. This level is a caching system which stores the map raster tiles. To increase the security level and the maintenance, the applications will be programmed and managed in the Python language under the Django framework.

On the client side, the well known OpenLayers library with the new GeoExt will be used to produce elegant, easy to use and well organized end user interface (Figure 7).
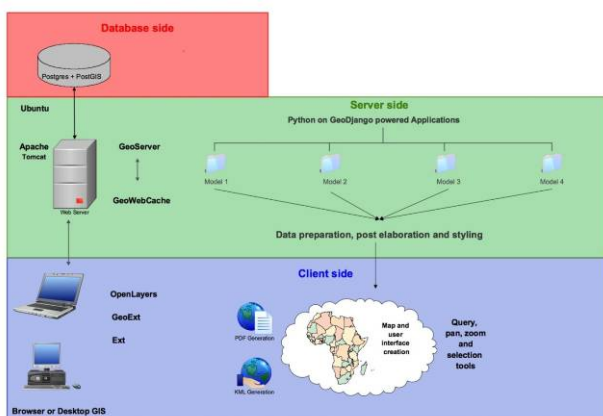

Figure 7 - ITHACA WebGIS architecture.

8. AUTHENTICATION AND AUTHORIZATION SERVICES

WFP operations includes the necessity to produce, store and retrieve sensible data in a secure way. In particular, data security is granted by the applicability of several different approaches, such as:

- *Authentication*, one-factor or two-factor;
- *Authorization;*
- *Privileges*;
- *Data encryption*;
- *Data Integrity algorithms*;
- *Auditing*;
- *Virtual Private Database Column Masking*, allowing only to authorized users to see the content of certain table fields (Figure 8);
- *Label Security Authorizations* (Figure 9).

Figure 8 - Virtual Private Database Column Masking

Figure 9 - Label Security Authorizations

9. CONCLUSIONS

WFP SDI solution, conceived and implemented by ITHACA, is operational since June 2008 and offer an effective solution to main critical issues, related to distributed management and exploitation of spatial data.

Systematic training activity for WFP users is on-going. A group of specialists from WFP headquarters and ITHACA has already met WFP GIS officers of South-Eastern Asia, Western Africa and North-Eastern Africa regions for presenting the solution and providing knowledge in order to effectively implement the federated architecture. Other WFP areas are expected to be visited in the first half of 2010.

The recent Haiti crisis has confirmed the overall system reliability and efficiency for distributing and disseminating reference and thematic data.

The flexibility of the web architecture allows to easily deploy web applications in short times and with specific tools, an important element for better supporting emergency management. Web applications may also become the way to optimize and coordinate the work of the growing community that, in case of major disaster, volunteered for systematic operations of data acquisition.

REFERENCES

References from Books:
Albano A., Ghelli G., Orsini R., 2007. *Basi di dati relazionali e a oggetti*. Zanichelli, Bologna, Italy.

Groot R. and McLaughlin J., 2000. *Geospatial Data Infrastructure: Concepts, Cases and Good Practice*. New York: Oxford University Press.

Masser I., 2005. *GIS Worlds - Creating Spatial Data Infrastructures*, ESRI Press.

References from websites:
Executive Office Of The President - Office of Management and Budget, 2002. Circular No. A-16, United States of America. http://www.whitehouse.gov/omb/circulars/a016/print/a016_rev. html (accessed: 28 January 2010).

Nebert D., 2004. *Developing Spatial Data Infrastructures: The SDI Cookbook*, Version 2.0, Global Spatial Data Infrastructure Association. http://www.gsdi.org/gsdicookbookindex.asp (accessed: 28 January 2010).

UNGIWG, 2007. *STRATEGY for developing and implementing a United Nations Spatial Data Infrastructure in support of Humanitarian Response, Economic Development, Environmental Protection, Peace and Safety*. http://www.ungiwg.org/docs/unsdi/UNSDI_Strategy_Implemen tation_Paper.pdf (accessed: 28 January 2010).

UNGIWG, 2007. *UNSDI COMPENDIUM. A UNSDI Vision, Implementation Strategy and Reference Architecture*. http://www.ungiwg.org/docs/unsdi/UNSDI_Compendium_13_0 2_2007.pdf (accessed: 28 January 2010).