POLITECNICO DI TORINO
Repository ISTITUZIONALE

High Availibilitysystems for Geographics Information

Original
High Availibilitysystems for Geographics Information / Ajmar, A., Terzo, O.. - (2010). (Gi4DM conference Geomatics for Crisis Management Turin Italy February 2-4, 2010).

Availability:
This version is available at: 11583/2382353 since:

Publisher:

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

05 September 2026

HIGH AVAILABILITY SYSTEMS FOR GEOGRAPHIC INFORMATION

A. Ajmar [a], O. Terzo [b]

[a] ITHACA, via P.C. Boggio 61, 10138 Torino, Italy - andrea.ajmar@ithaca.polito.it
[b] Istituto Superiore Mario Boella, via P.C. Boggio 61, 10138 Torino, Italy - terzo@ismb.it

KEY WORDS: Databases, Developing Countries, Hazards, Internet/Web, Decision Support, Spatial Infrastructures, Interoperability

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) in the UNSDI initiative. WFP SDI is built to fulfil organization requirements in terms of availability of geographic information for the achievement of the objectives of saving lives and protect livelihoods in emergencies, prepare for emergencies, restore and rebuild lives after emergencies, reduce chronic hunger and undernutrition everywhere, strengthen the capacity of countries to reduce hunger. Geodatabase component of the WFP SDI is an ESRI ArcSDE relational database that uses ORACLE as Data Base Management System (DBMS).
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. Those replicas contain an extraction from the master geodatabase of the area of interest of the office; local personnel, other than being an authorized user, is also responsible for updating geodatabase replica content and for transferring those updates to master site (synchronization). The capability of working on local replicas is fundamental considering the remote locations, network connectivity issues and the complex environmental condition that are typical of most of WFP activation. Local replication includes the capability of operate on mobile devices, granting the capability of acquiring new and updated information directly on the field.
Furthermore, Oracle streaming features are used to create and maintain aligned a set of read-only geodatabase instances; streaming environment is created and set to maintain ESRI ArcSDE geodatabase integrity. The objective is to have 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.
This architecture is effectively 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 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.

RÉSUMÉ:
Le Programme Alimentaire Mondial (PAM) est en train de mettre en œuvre sa propre Infrastructure de Données Spatiales (IDS), dans le cadre des directives élaborées par l'Organisation des Nations Unies Geographic Information Working Group (UNGIWG), mais aussi sur la base de l'initiative UNSDI. PAM IDS a été conçue pour satisfaire aux exigences de l'organisation en termes de disponibilité de l'information géographique pour la réalisation des objectifs de sauver des vies et protéger les moyens de subsistance pendant les situations d'urgence, de préparer aux situations d'urgence, restaurer et reconstruire une societé après les urgences, réduire la faim chronique et la malnutrition partout dans le monde, renforcer la capacité des pays à réduire la faim. La composante des données de PAM IDS est un ESRI ArcSDE geodatabase qui utilise la base de données relationnelle Oracle comme Data Base Management System (DBMS).
Particulièrement pour l'utilisation en cas d'urgence, la mise à jour des informations, la disponibilité du système, la redondance et les performances sont des facteurs clés. Pour cette raison, l'architecture du PAM IDS est basé sur une structure de géodatabase fédérées, avec des répliques distribuées dans les bureaux régionaux et locaux. Ces répliques contiennent une extraction à partir de la géodatabase maître de la zone d'intérêt de l'office; le personnel local, autre que d'être un utilisateur autorisé, est également chargé d'actualiser le contenu du géodatabase répliqué et pour le transfert de ces mises à jour de site principal (synchronisation). La capacité de travailler sur des répliques locales est fondamental considérant les régions éloignées, les problèmes de connectivité réseau et les conditions environnementales complexes qui sont typiques de la plupart des activations du PAM. La réplication locale inclut la capacité de fonctionner sur des appareils mobiles, l'octroi de la capacité d'acquérir de nouvelles informations et mises à jour directement sur le terrain.
En outre, les fonctionnalités de streaming de Oracle sont utilisés pour créer et maintenir aligné en lecture seul la base de donnée géographique L'objectif est d'avoir plusieurs instances distribuées, éventuellement sur différentes plateformes DBMS, automatiquement aligné sur le site principal, l'octroi de la redondance du système, la réduction de la charge réseau et les performances augmentés grâce à la mise en œuvre de l'équilibrage de charge des procédures. Cette architecture est un cluster actif-passif, où les nœuds secondaires sont en mode veille, et activés en cas de défaillance du nœud maître. Des procédures de basculement commute automatiquement les demandes vers le nouveau server actif, dans la plupart des cas de manière transparente pour l'utilisateur. Le server de secours est également utilisé pour l'access en lecture seule, augmentant les performances globales du système.

1. INTRODUCTION

Nowadays, the world is facing disasters on an unprecedented scale: millions of people are affected by natural disasters globally each year and, only in the last decade, more than 80% of all disaster-related deaths were caused by natural hazards.
When an event hits regions of the world where local authorities are not structured to deal with their complex effects, they normally ask the intervention of international organizations, such as UN agencies. Those organizations are in charge of the activation of emergency procedures in order to cover affected population's immediate needs and, with a long-term view, to prepare protracted relief and recovery operations.
The different events that may generate the need for an humanitarian intervention may be grouped in three major kinds:

- sudden disasters, which affect food access and/or cause population displacements;
- slow-onset disasters, as droughts and crop failures;
- complex emergencies, that can involve conflict, widespread social and economic disruption and large population displacements.

In the frame of humanitarian operations, the *World Food Programme* (WFP) of the United Nations is in the front line. The WFP is the biggest UN Agency and responds to more than 120 emergencies per year worldwide.

One major requisite for an effective response to an emergency is the establishment of a logistic infrastructure. Humanitarian Logistics: is "the process of planning, implementing and controlling the efficient, cost-effective flow and storage of goods and materials, as well as related information, from the point of origin to the point of consumption for the purpose of alleviating the suffering of vulnerable people. The function encompasses a range of activities, including preparedness, planning, procurement, transport, warehousing, tracking and tracing and customs clearance." (Thomans et al., 2005). Created in the framework of the Cluster Approach UN reform, the primary mechanism for inter agency coordination of humanitarian assistance involving key UN and non-UN humanitarian partners, the Logistic Cluster is 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.

Short-term emergency response capacities, long-term risk reduction, development and environmental protection activities are sector where a Spatial Data Infrastructure (SDI) may strongly improve efficiency. The term Spatial Data Infrastructure (SDI) is often used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. 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. 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. The WFP SDI solution has the objective to eliminate or minimize deficiencies related with actual way of operate, such as:

- 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.

System availability, redundancy and performances are other critical factor for operational systems managing humanitarian emergencies. All those aspect has been taken into account in the definition of the global WFP SDI architecture.

2. WFP SDI USERS

WFP SDI users are mainly composed by WFP effectives dislocated in 8 Regional Bureaux and several Country Offices instituted in most of the developing countries (Figure 1).

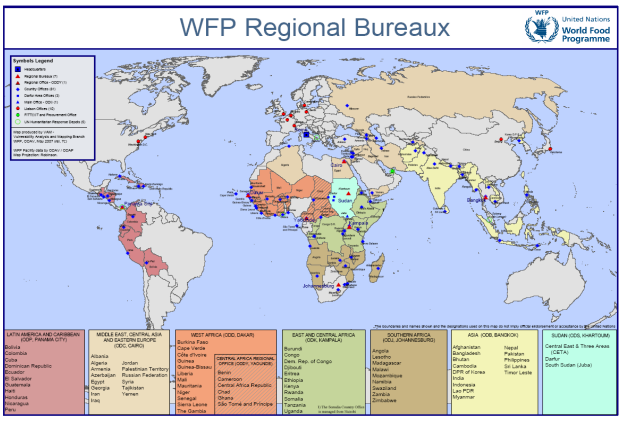

Figure 1 - WFP SDI Users and their dislocation.

Among WFP Regional Bureaux and main offices, Internet connectivity in terms of speed and availability is quite variable (Table 1), but rarely the conditions for a completely Internet based architecture are present. For that reason, the final implementation must include capabilities of transferring and synchronizing several geodatabase instances, creating a federated structure.

Country	City	Internet speed
Kenya	Nairobi	1.5 Mbps
Thailand	Bangkok	512 Kbps
Senegal	Dakar	1 Mbps
Uganda	Kampala	256 Kbps
Egypt	Cairo	1 Mbps
Panama	Panama City	2 Mbps
South Africa	Johannesburg	384 Kbps
Sudan	Khartoum	1 Mbps
Zimbabwe	Harare	No available connectivity

Country	City	Internet speed
Ethiopia	Addis Ababa	256 Kbps
Chad	N'Djamena	No available connectivity

Table 1 - WFP Regional Bureaux and main offices Internet speed.

3. WFP SDI ARCHITECTURE

3.1 Master site

A two-levels architecture has been implemented, to fulfil two major requirements:

- to increase performances by splitting the production and publication environments;
- to study new features in order to implement a progressive porting of the geodatabase from a commercial to a non commercial Database Management System.

Production/Editing environment: back end component accessed by high level users, in charge of database management and of performing complex data analysis procedures. The necessity of having ready-to-use and operative functionalities for ongoing activities and missions, granting high levels of data security and reliability, is the main factor suggesting the implementation of a commercial products based platform.

Publication/Analysis environment: front end applications, mainly dedicated to analysis, processing of project specific geodata and exploratory aspects. This environment is developed on a completely open source platform, for high availability and interoperability of derived applications and services. This environment, in future perspective, may substitute in all functionalities the production/editing commercial based environment, once the development of certain functionalities for data management and security will be considered mature.

Actually implemented architecture includes a production geodatabase based on Oracle 10G and a PostgreSQL/PostGIS database for the publication environment, acting as source for web mapping services. Several scripts have been implemented for exporting data from Oracle to PostgreSQL, by using ora2pg and ogr2ogr tools.

WFP SDI production/publication environments are fully operating since June 2008. The system is composed by:

- an ESRI enterprise geodatabase on ORACLE 10g for data storage for the production environment;
- a PostgreSQL/PostGIS geodatabase for data storage for the publication environment (and relative alignment procedures);
- data access services;
- a GeoNetwork node, for discovery services;
- authentication and authorization processes for users identification and accounting;
- templates for the optimization and the normalization of the map production processes;
- several beta-version web-GIS applications, general purpose and dedicated.

3.2 Secondary sites

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.

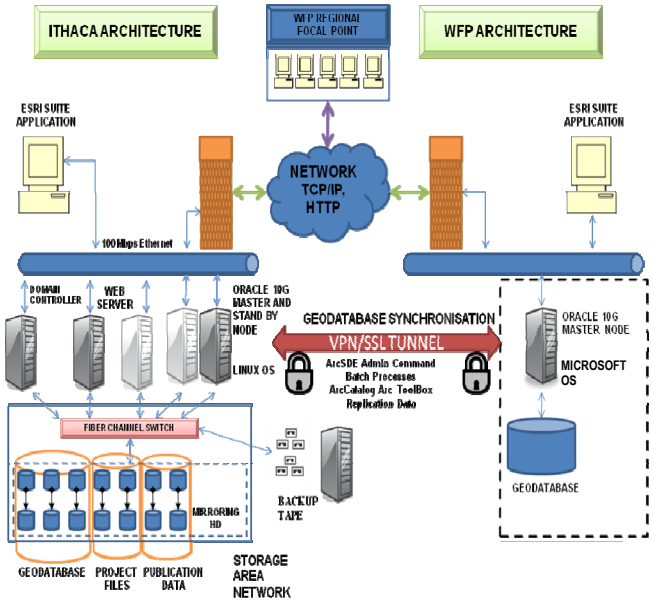

Figure 2 - WFP SDI architecture

The approach based on the implementation of Oracle Stream feature has been tested in a pre-production environment of WFP SDI architecture. 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 (Figure 3).

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. It was defined witch of the ArcSDE system tables should be replicated; that's is a key issue as both server has an operating ArcSDE instance and the incorrect choice of that list may corrupt the downstream instance.

Before starting the streaming, database backup was used to create, on the master and on the downstream node, 2 identical database.

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.
Next step is to extend the text phase, using as downstream node a WFP server. That should help in better understanding if some physical network linked issue may influence streaming functionalities. It will help also in understanding if a different OS platform (ITHACA server uses Linux distribution, WFP servers operate on Windows environment) requires any adaptation of the configuration.

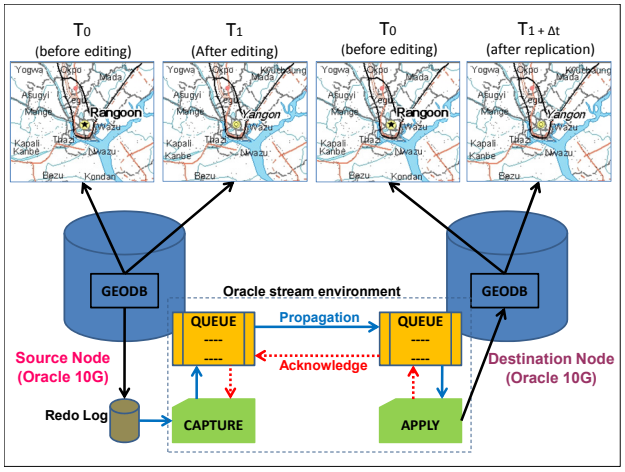

Figure 3 – WFP SDI stream environment

Oracle Streams technology is composed of a set of elements, the capture, the propagation and the apply area:

- The *Capture process:* by using the implicit capture mechanism, the changes made on the geodatabase can be store in a queue events list and replicated to a remote Database Management System. The capture elements extract two type of changes, the DML, typically where we change the value of data, and the DDL, when we change the structure of the objects. Changes are extracted from the redo log element, which log a history of all changes made to the database; each redo log file consists of redo records. After the extraction, the change events are published on the queue area;
- The *propagation processn*: this mechanism provide to sent events captured from one queue to another. The queue which events are propagated is the source. The events received on the destination queue can be consumed by the apply process;
- The *apply process*: this process dequeue messages sent by the propagation feature. The apply handle an user-defined procedure process making that the results of the DDL and DML changes are applied to the objects on the destination database. The default apply engine will detect conflicts where the destination row has been changed and does not contain the expected values. If a conflict is detected, then a resolution routine may be invoked.

3.3 Remote nodes

As mentioned in Chapter 2, dislocation of the users and conditions during operation in emergencies are factor that makes a solution based exclusively on Internet services not applicable. For that reason 2 different solutions has been envisaged and compared, in order to be able to choose the optimal one according with local telecommunication infrastructure constraints.

Centralized (application server) based architecture
Citrix XenApp (formerly Citrix MetaFrame Server and Citrix Presentation Server) is a remote access/application publishing product that allows users to connect to their corporate applications, usually available from central servers. Implementing this solution implies acquisition, configuration and maintenance of an application server where applications for remote users are installed and made available (Figure 4). The advantage of that solution are that all licensing and upgrades activities can be performed centrally, and standard operations and procedures are more easily controlled. Network traffic is also optimized, because only graphical outputs are sent over the network. The main disadvantage of that solution is that, in completely disconnected environments, no services are provided, making the solution completely dependent on the reliability of the telecommunication infrastructure.

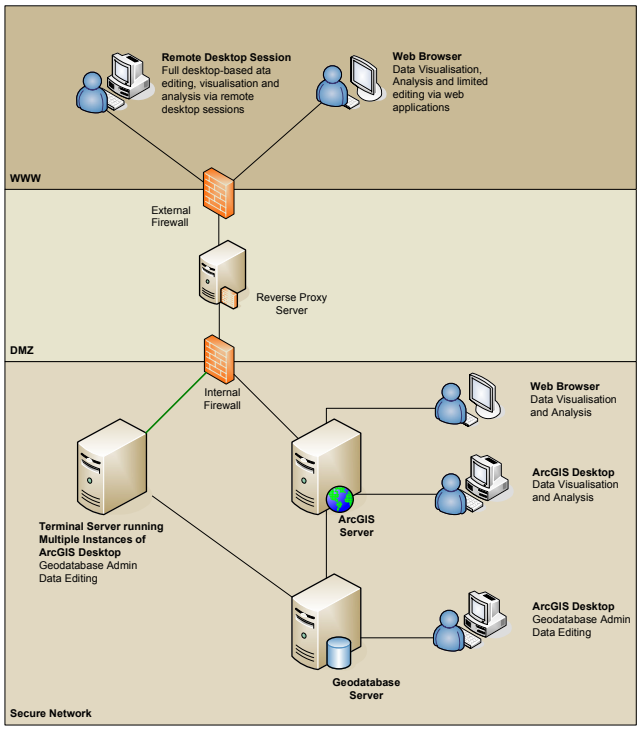

Figure 4 - Centralized architecture schema

Federated architecture
A federated architecture is built on top of the capability to maintain synchronized several geodatabase instances. The objective is to create an infrastructure where Regional Bureaux, main and country offices and local effectives on the field have the capability to extract information on their area of competence or interest, to maintain those data and to synchronize it with their upstream node (Figure 5). The synchronization process should be executed online or offline, granting the operation also in environment with low or no connectivity. The system conceived in that way has the advantage to make local effectives, the ones that are the nearest to the area that must be described by geographic data, responsible for data production and maintenance (Figure 6).

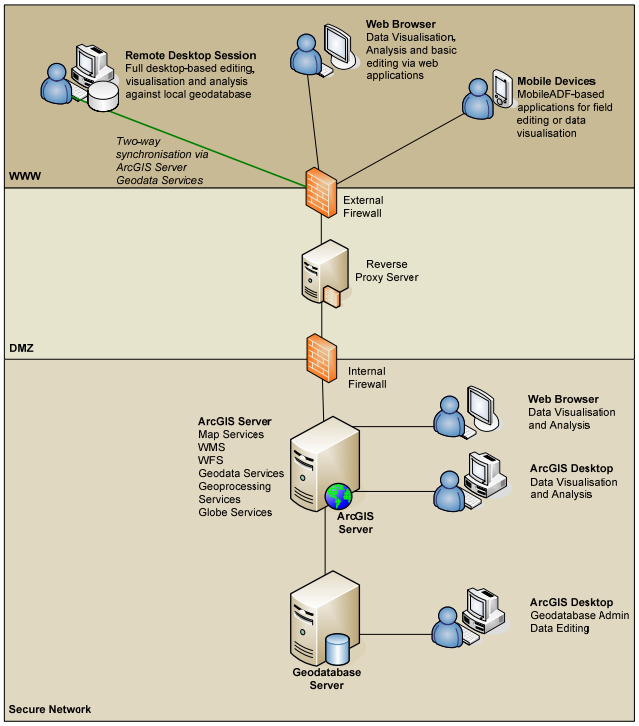

Figure 5 - Federated architecture schema

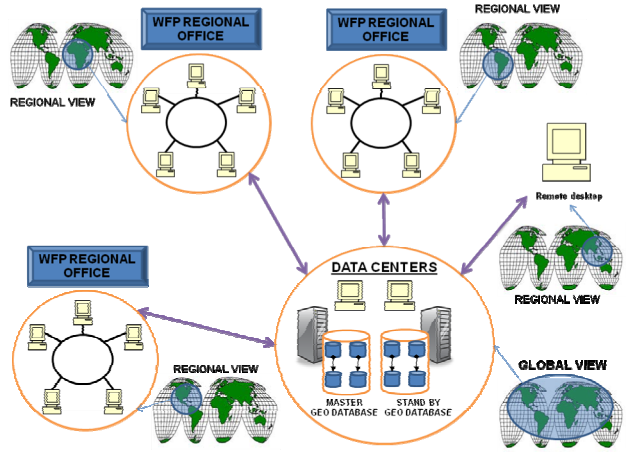

Figure 6 - Data organization in a federated architecture

4. CONCLUSIONS

Key characteristic of Information System architecture for emergency management is to provide fundamental data access, analysis and map production capabilities even if the users have to operate in difficult environmental conditions. In the first hours of a major emergency, in particular, damages to telecommunication infrastructures and related system failures are common.

On top of that, in many of the country where WFP operates the quality and reliability of telecommunication services are not at a level suitable to support complex data management and analysis, such as those requested by Geographic Information Systems.

At the same time, for a broadly distributed organization such as WFP, the capability to provide timely access to the best data available is a major issue, because the risk of the existence of "invisible" data, not computerized or hidden in local computers, is very high.

All above mentioned consideration was taken into account while proposing the solution for WFP SDI.

Federated geodatabases are the way to create commonly agreed and shared data model, data maintaining procedures and services and functionalities for keeping aligned the different instances.

Web services and applications are the mean for increase overall interoperability levels, making the final users independent from specific Operative System platforms and software licensed and installed. Web applications too are conceived to support off-line capacities and automatic synchronization when connectivity is re-established.

High availability may be reached through the implementation of transparent failover procedures, built on top of stream technology. Same technology may also be used to create a network of server, distributed around the globe, in order to increase security and performances.

5. REFERENCES

Crocker C., 2005, *The Oracle Enterprise Geodatabase (EDGB)*, Norcal URISA Chapter Meeting, http://www.fargeo.com (accessed 21 Apr. 2008)

Donsez D., 2003, *Les SGBDs Géographiques*, teaching material

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 21 Apr. 08)

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

Li L., Chen A., Bai Y., Wei Y., 2006, *Implementation of Geospatial Product Virtualization in Grid Environment*, NASA's Earth Science Technology Office (ESTO) Conference – ESTC

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

Mondini M., 2006, *La realizzazione di soluzioni interoperabili in una spatial data infrastructure,* X° Conferenza Nazionale ASITA, Bolzano, Italy

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 21 Apr. 08)

Thomans, A.S and L.R. Kopczak, 2005 - *From Logistics to supply chain management*. The path forward in the humanitarian sector, Fritz Institute.

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_Implementation_Paper.pdf (accessed 21 Apr. 08)

UNGIWG, 2007, UNSDI COMPENDIUM. A UNSDI Vision, Implementation Strategy and Reference Architecture. http://www.ungiwg.org/docs/unsdi/UNSDI_Compendium_13_02_2007.pdf (accessed 21 Apr. 08)

Zwicker C., Pierce B., 2007, *Geodatabase – An Introduction*, ESRI International User Conference