WFP Spatial Data Infrastructure (SDI)

In support to a UN SDI

Preliminary proposal

Version 1.0, 18/05/2007

Summary

ITHACA - Information Technology For Humanitarian Assistance, Cooperation and Action
Sede legale: Via Pier Carlo Boggio 61 • 10138 Torino • ingresso: Corso Castelfidardo 30/A Tel. 011.19751853 • Fax 011.19751122 • www.ithacaweb.org • E-Mail: info@ithaca.polito.it C.F. 97682870015

Soci fondatori:

Istituto Superiore sui Sistemi Territoriali per l'Innovazione

POLITECNICO DI TORINO

1. Introduction

The United Nations Geographical Information Working Group (UNGIWG) is a network of professionals working in the fields of cartography and geographic information science to building the UN Spatial Data Infrastructure needed to achieve sustainable development.

UNGIWG was formed in 2000 to address common geospatial issues - maps, boundaries, data exchange, standards - that affect the work of UN Agencies. UNGIWG also works directly with non-governmental organizations, research institutions and industry to develop and maintain common geographic databases and geospatial technologies to enhance normative and operational capabilities. UNGIWG reports periodically to the UN Chief Executive Board (CEB) on progress made and priority issues.

Specifically UNGIWG aims to:

- improve the efficient use of geographic information for better decision-making;
- promote standards and norms for maps and other geospatial information;
- develop core maps to avoid duplication;
- build mechanisms for sharing, maintaining and assuring the quality of geographic information;
- provide a forum for discussing common issues and emerging technological changes.

During the Sixth UNGIWG Plenary Meeting held in Addis Ababa in October 2005, endorsement for a UN Spatial Data Infrastructure (UNSDI) to support coordinated efforts in the development and management of geo-spatial information was made.

The main issues on Geospatial Information management & expectations of the UNSDI has been identified during the last UNSDI Global Partners Meeting held on 1-2 March 2007 in Frascati (Italy) and are highlighted in the next table.

Table 1 – UNSDI main issues and expectations

	Issues	**Expectations of UNSDI**
Data & Services	• Inconsistent data in terms of content and format • Existence of "invisible" data: not computerized or hidden in local computers • Data access issues (restrictive access) • Confidentiality and sensitivity of certain data and information • Difficulties in implementing data/systems integration	• More efficient search of, and access to data in emergencies: shorter response time, most relevant information is shared and reduction of gaps and duplications
Standards	• Need for standardization • Poor application of standards at country level • Limited use of existing standards for data sharing (and lack thereof)	• Active support to implementation of standards • Certification of spatial data/SDI that adhere to standards • Access to standards and best practices for data collection, analysis and sharing • Standards at international and national levels

Metadata	• Lack of extensive and reliable metadata Catalogues	• Standardized Metadata population and the development of catalogue services • Facilitate metadata creation, discover, retrieval and visualization.
Capacity building	• Gap between national and sub-national data	• Development of national capacity • Repository of common technical knowledge • Strengthening of GIS/Remote Sensing units within respective agencies
Organizational	• Limited human and financial resources for tools development and maintenance • Lack of streamlining of spatial analysis in decision making (outside the responsible units in agencies) • Unproductive competitive practices	• Focus more on governance and sustainability than mere technology • Build partnerships

In that framework it has been considered very important to start immediately with quick win products, to prove the effectiveness of the UNSDI and to keep high the interest of the top decision/policy makers.

2. Needs assessment

A preliminary needs assessment has beencarried out, involving the three structures that, in the actual organization of the WFP, are operating with geographic data stored in GIS:

- 2 units operating directly under UN WFP:
  - ODAP (Emergency Preparedness and Response Branch);
  - ODAV (Vulnerability Analysis and Mapping Branch).
- UNJLC as UN humanitarian common service operating under theWFP umbrella that deals with logistic data.

All significant elements acquired during this assessment has beenresumed in a document titled *Toward a common UNSDI - State of the art* and used as guidelines during the developing and implementing phases.

Common to all units came out the necessity to define, implement, maintain and distribute a minimal level of data for geographic analysis and reporting. The definition of a data model is considered the first step toward the goal of data reliability and integrity, standardization and metadata integration. The resulting geodatabase must have characteristics suitable to be easily interfaced with VAM SIE metadata repository, enabling in that way decentralized structures to accessand exchange geographic relates information.

The geodatabase conceptual model and implementation rules must be platform independent; anyway some users manifested the needto develop an architecture compatible with ESRI ArcGIS suite, available and well known in UN GIS units.

3. Data inventory

The successive phase has been performed with the scope to specify the available geographic data, coming from different sources, which can be used in the support of base mapping and emergency preparedness and response activities.
The sources of data considered in the inventory were identified based on a review of on-line available resources. The only restriction that were applied was that the data should be globally consistent (small map scale data); therefore, country and project specific data were not considered.
It is necessary to point out that many of the data identified are released into the public domain and are available for not commercial use completely free of charge.

Ten topical areas have been initially considered: BOUNDARIES, ELEVATION, HYDROGRAPHY, PHYSIOGRAPHY, POPULATION, TRANSPORTATION, INDUSTRY, UTILITIES, VEGETATION and NAMES. In the following table the specific data layers identified for the purpose of the geodatabase definition are listed and described. The list of the corresponding data sources are shown in table 3.

Table 2 – Thematic areas

BOUNDARIES
BNDPolA - Political Boundaries (areas, polygon) BNDPolL - Political Boundaries (outlines, polyline) BNDOceanSea - Oceans and Seas (polygon) BNDCoastline - Coastlines (polyline)
POPULATION
POPBuiltUpA - Built-Up areas (polygon) POPAnthrFeat - Anthropic features (point) POPPopDensity - Population density (raster)
TRANSPORTATION
according to the databases defined from the UNJLC
INDUSTRY
INDIndA - Industries (extraction/fish, polygon) INDStorpoint - Storage points (point) INDIndP - Industries (extraction/fish, point)
UTILITIES
UTITransmLines -Transmission lines (power, pipelines, etc., polyline) UTITransmNodes - Transmission nodes (plants, pumping, etc., point)
ELEVATION
ELVCntline - Contour lines on land (polyline) ELVDepthline - Depth lines (polyline) ELVElevPoint - Elevation points (point) ELVDtm - Digital Terrain Model (raster)
HYDROGRAPHY
HYDInWaterA - Inland water bodies (polygon) HYDMiscWater - Miscellaneous water elements (polyline) HYDCanal - Inland water canals (polyline) HYDInWaterL - Inland water paths (polyline) HYDMiscWaterP - Miscellaneous water elements (point) HYDDangerPoint - Point of danger on water (point) HYDSwbd - SRTM Water Body Areas (polygon) HYDBasin - Drainage basins (polygon)
PHYSIOGRAPHY
PHYGroundCover - Ground cover broad classes (polygon) PHYLandCover - Land cover (raster) PHYLandsatMos - Orthorectified Landsat TM Mosaics (year 2000, raster) PHYModisLI - Modis Land Imagery (raster) PHYModisNDVI - Modis NDVI data (raster) PHYModisLC - Modis Land Cover data (raster)
NAMES
NMSNamedElements - Geographic elements with name (point)
VEGETATION
No data sources have been currently identified

These datasets have been identified according to the results of the preliminary needs assessment phase (described in section 2). Further researches are required in order to include other necessary data in the geodatabase, according to more specific needs and demands (for instance, human health and socio-economic databases, hydrological observations such as river gauge monitoring and runoff data, soil databases and geological data, climatic observations such as near real time or historical precipitation data).

Table 3 – Data sources

BOUNDARIES
<u>VMAP0 – Vector Map Level 0</u> Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000 <u>SALB – Second Administrative Level Boundaries</u> Data source: UNGIWG Data geometry: vector Native format: shapefile Geographic area: near global Scale: 1:1000000
POPULATION
<u>LandScan – LandScan Global Population</u> Data source: Oak Ridge National Laboratory Data geometry: raster Native format: ESRI Grid Geographic area: world Resolution: 1 km
INDUSTRY
<u>VMAP0 – Vector Map Level 0</u> Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
UTILITIES
<u>VMAP0 – Vector Map Level 0</u> Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
ELEVATION
<u>VMAP0 – Vector Map Level 0</u> Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000 <u>DTM from Shuttle Rada topography Mission (SRTM)</u> Data source: NASA - JPL Data geometry: raster Native format: binary Geographic area: world Resolution: 90 m
HYDROGRAPHY
<u>VMAP0 – Vector Map Level 0</u> Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000 <u>STRM Water Body Data (SWBD)</u> Data source: NASA - JPL Data geometry: vector Native format: shapefile Geographic area: near global <u>Drainage Basins Level 1, 2, 3</u> Data source: USGS – EROS Data geometry: vector Native format: shapefile Geographic area: world Scale: 1:5000000
PHYSIOGRAPHY - VEGETATION
<u>VMAP0 – Vector Map Level 0</u> Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000 <u>Orthorectified Landsat Thematic Mapper Mosaics</u> Data source: Geo Community Data geometry: raster Native format: TIFF/Geotif Geographic area: world Resolution: 30m <u>GLC2000- Global Landcover Classification for the year 2000</u> Data source: JRC- IES Data geometry: raster Native format: ESRI Grid Geographic area: world Resolution: 1 km <u>MODIS Land Imagery</u> Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Resolution: 250 m, 500 m, 1 km <u>MODIS NDVI Data</u> Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Resolution: 250 m, 500 m, 1 km <u>MODIS Land Cover Data</u> Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Resolution: 1 km, 5 km <u>Landsat 7 ETM+ Ortho GeoCover Imagery</u> Data source: GLFC Data geometry: raster Native format: TIFF/Geotif Geographic area: world Resolution: 30m
NAMES
<u>Geonames – Geographic names</u> Data source: NGA Data geometry: table with coordinates Native format: ASCII Geographic area: world Scale: 1:5000000-1:10000
OTHER

4. Geodatabase structure definition

In the process of defining a geodatabase structure it has been decided to split this activity in two successive steps:

- **Step 1**: for any single object, implementation of a data structure matching with the one of identified source data. This structure will not be yet optimized on user needs but has the advantage to preserve any possible information and it must be considered as the starting point for further refining discussion and activities, involving all potential final users.
  A geodatabase structured in that way should in any case provide several immediate advantage to the users:
  - All data sources are integrated, granting the immediate access to the most reliable and updated data for any category;
  - The geodatabase data format provide additional functionalities (domains, subtypes, relations, topology rules, more efficient data indexing, etc.) not available in source data native formats;
  - Source data may be rapidly and easily loaded in the GeoDB, with no or few needs of import procedures development;
  - Users can test the solution looking at familiar data structures and field definition;
  - Data procedures for selection, symbolization, export, etc. eventually already developed inside users structures may be reused with no or little intervention.
- **Step 2**: after the test phase, new needs assessment wil be performed, finalizing the definition of a brand new restructured data model, optimized for the use of UN structures. That phase should include, and eventually complete, the outcomes of the common terminology definition (thesaurus), data collection forms and specifics procedures developed by UNJLC in the framework of their ongoing activity of transportation data model implementation.
  To be evaluated a structure that use Valid Value Tables (VVT) to define feature types.

4.1.Description of Step1 activities

4.1.1.Data sets definition

Geographic data has been grouped in 11 dataset classes, as below described.

Table 4 – Modelled data sets

Datasets		
SeqNum	**Name**	**Acronym**
1	Boundaries	BND
2	Elevation	ELV
3	Hydrography	HYD
4	Physiography	PHY
5	Population	POP
6	Transportation	TRN
7	Industry	IND
8	Utilities	UTI
9	Vegetation	VEG
10	Names	NMS

Datasets		
SeqNum	Name	Acronym
11	DataQuality	DQL

4.1.2. *Data classes definition*

Geographic data has been structured in 37 classes (vector, raster and tables), defining for each class the relative data sources, identified in Primary or Secondary (table 5).

UNJLC is active in the definition and implementation of a geodatabase mainly dedicated to transportation, containing the following datasets:

- Ports;
- Navigable waterways;
- Roads;
- Airports;
- Railways (with stations);
- Fuel supply points;
- Warehouses;
- Border crossings.

The UNJLC geodatabase is structured to be enriched and updated locally in case of emergency activation. In those occasion the interested portion of geodatabase is checked out from the main db, provided to local officer that are requested to check and complete existing information and to acquire brand new data, always following UNJLC best practices definition. The local officers are normally equipped with GPS devices. The new acquired information are inserted in the main geodatabase, after a validation procedure executed by internal GIS officers in UNJLC offices.

A specific activity of model integration must be envisaged, to grand logical compatibility and geodatabase normalization.

Table 5 – Modelled classes with respective data sources

Classes and relative data sources					
Dataset	Class	Description	DataType	DataSource	Primary
Boundaries	BNDCoastline	Coastlines	polyline	VMAP0	Yes
Boundaries	BNDOceanSea	Oceans and Seas	polygon	VMAP0	Yes
Boundaries	BNDPolA	Political Boundaries (areas)	polygon	VMAP0	Yes
Boundaries	BNDPolA	Political Boundaries (areas)	polygon	SALB	No
Boundaries	BNDPolL	Political Boundaries (outlines)	polyline	VMAP0	Yes
Boundaries	BNDPolL	Political Boundaries (outlines)	polyline	SALB	No
Elevation	ELVCntline	Contour lines on land	polyline	VMAP0	Yes
Elevation	ELVDepthline	Depth lines	polyline	VMAP0	Yes
Elevation	ELVDtm	Digital Terrain Model	raster	SRTM	Yes

Classes and relative data sources					
Dataset	**Class**	**Description**	**DataType**	**DataSource**	**Primary**
Elevation	ELVElevPoint	Elevation points	point	VMAP0	Yes
Hydrography	HYDBasin	Drainage basins	polygon	Basin	Yes
Hydrography	HYDCanal	Inland water canals	polyline	VMAP0	Yes
Hydrography	HYDInWaterA	Inland water bodies	polygon	VMAP0	Yes
Hydrography	HYDInWaterL	Inland water paths	polyline	VMAP0	Yes
Hydrography	HYDMiscWaterL	Miscellaneous water elements	polyline	VMAP0	Yes
Hydrography	HYDMiscWaterP	Miscellaneous water elements	point	VMAP0	Yes
Hydrography	HYDSwbd	SRTM Water Body Areas	polygon	STRM SWBD	Yes
Physiography	PHYGroundCover	Ground cover broad classes	polygon	VMAP0	Yes
Physiography	PHYLandCover	Land cover (1km resolution)	raster	GLC2000	Yes
Physiography	PHYLandsatMos	Orthorectified Landsat TM Mosaics (year 2000)	raster	Landsat mosaic	Yes
Physiography	PHYModisLC	Modis Land Cover data	raster	MODIS Land Cover	Yes
Physiography	PHYModisLI	Modis Land Imagery	raster	MODIS Imagery	Yes
Physiography	PHYModisNDVI	Modis NDVI data	raster	MODIS NDVI	Yes
Population	POPAnthrFeat	Anthropic features	point	VMAP0	Yes
Population	POPBuiltUpA	Built-Up areas	polygon	VMAP0	Yes
Population	POPPopDensity	Population density	raster	LandScan	Yes
Industry	INDIndA	Industries (extraction/fish)	polygon	VMAP0	Yes
Industry	INDIndP	Industries (extraction/fish)	point	VMAP0	Yes
Utilities	UTITransmLines	Transmission lines (power, pipelines, etc.)	polyline	VMAP0	Yes
Utilities	UTITransmNodes	Transmission nodes (plants, pumping, etc.)	point	VMAP0	Yes
Names	NMSAdmReg	Names of administrative region features	table	Geonames	Yes
Names	NMSHydTyp	Names of Hydrographic type features	table	Geonames	Yes
Names	NMSHypTyp	Names of Hypsographic type features	table	Geonames	Yes
Names	NMSLocFeat	Names of locality or area type features	table	Geonames	Yes
Names	NMSPopPlace	Names of populated place features	table	Geonames	Yes

Classes and relative data sources					
Dataset	**Class**	**Description**	**DataType**	**DataSource**	**Primary**
Names	NMSRoadTyp	Names of street, highway, road or railroad type features	table	Geonames	Yes
Names	NMSSptTyp	Names of spot type features	table	Geonames	Yes
Names	NMSUndSea	Names of undersea type features	table	Geonames	Yes
Names	NMSVegTyp	Names of vegetation type features	table	Geonames	Yes

Vegetation dataset has been introduced even if no reliable data sources has been currently identified for that dataset. DataQuality classes should be implemented only after the definition of the final model structure.

5. Geodatabase structure implementation

5.1.CASE tools and UML model

Microsoft Office Visio 2003 has been used as CASE tool to develop a data model in UML format (GlobalGDB_d1.vsd available here).
The model structure can be visualized via browser opening the following link.

In that environment datasets and classes has been modelled, together with all attribute domains.

5.2.Data Model integration

While UML is a useful tool for documenting the relational aspects of a geodatabase schema (such as table layouts and relationships), generally, it is not recommended to solely use UML for geodatabase design.

UML can be useful for relational database design. However, UML has generally not been useful for designing richer geographic behaviour—topologies, networks, terrains, raster catalogs, map layers, map symbols, metadata, cartographic representations, semantic classifications, address locators, cadastral fabrics, linear referencing, and geoprocessing models. These data elements are used to define geographic behaviour and associations. Much of the richness of the geodatabase cannot be universally expressed in a UML design. More important, no special GIS insight is achieved through UML design.

Fore those reasons the base model, developed in UML, needed to be integrated perfoming the below described activities.

As result a complete data model has been defined and made available for download here in the following formats:

- XML (GlobalGDB_d1.xml);
- ESRI Personal Geodatabase (GlobalGDB_d1.mdb).

5.3. Domain updates

In the proposed structure, it has been necessary to analyzeall features of source data in order to define the complete list of unique values for each domain field. An automatic procedure has been developed to accomplish that task, giving as result a set of domain tables. Those domain values has been used to complete the domains defined in UML.

5.4. Spatial reference

Spatial reference cannot be modelled in UML and thus has been defined externally from CASE tool. At the moment it has been defined that all geographic object arestored in a same coordinate system, defined by the following parameters:

GCS_WGS_WGS1984
    Angular Unit: Degree (0.017453292519943299)
    Prime Meridian: Greenwich (0.000000000000000000)
    Datum: D_WGS_1984
        Spheroid: WGS_1984
            Semimajor Axis: 6378137.000000000000000000
            Semiminor Axis: 6356752.314245179300000000
            Inverse Flattening: 298.257223563000030000

5.5. Raster elements

Raster element cannot be modelled in UML and has been added successively.

5.6. Topology rules

Topology rules cannot be modelled in UML and must be define successively. At the actual status of the project, where the actual data model is in a preliminary draft status, only 2 topological rules has been implemented in order to show the potential of those controls during data loading and/or editing procedures:

- BNDCoastlline must be covered by boundary of BNDPolA;
- BNDOceanSea must not overlap with BNDPolA.

6. Data loading

To facilitate the understanding of model behaviour and to test some new functionalities, all geographic elements part of the Boundaries dataset has beenloaded. In particular for the geographic classes BNDPolA and BNDPolL that has multiple data sources (VMap0 and SALB), local data integration has been performed during data loading procedures, exclusively for the territory of Madagascar.

Data representation for all Boundaries classes has been defined; in case of BNDPolL representation based on both feature codes and use has been proposed.

A geodatabase with those features loaded is available here in the following formats:

- XML (GlobalGDB_Feat_d1.xml);
- ESRI Personal Geodatabase (GlobalGDB_Feat_d1.mdb).

7. Metadata implementation

ISO standard 19115:2003 is recognized as reference for metadata description; GeoNetwork Opensource software, on which VAM SIE is based, provides a set of simplified metadata templates that can be used as a guide for metadata compilation.
The data loading process in the geodatabase must include metadata implementation, deriving properties automatically, where possible.

List of acronyms

ACRONYM	DESCRIPTION
DTM	A digital terrain model is a digital representation of ground surface topography or terrain. A DTM can be represented as a raster (a grid of squares) or as a triangular irregular network. DTMs are commonly built using remote sensing techniques, however, they may also be built from land surveying. DTMs are used often in geographic information systems, and are the most common basis for digitally-produced relief maps.
CASE	Computer-Aided Software Engineering tools are software components that assists with the development and maintenance of software, especially the analysis and design. Complex tasks that often require many lines of code are simplified with CASE user interfaces and code generators.
GLCF	Global Land Cover Facility (http://glcf.umiacs.umd.edu/index.shtml)
JRC-IES	Joint Research Centre - Institute for Environment and Sustainability (http://ies.jrc.cec.eu.int/)
NASA-EOS	NASA Earth Observing System (http://eospso.gsfc.nasa.gov/index.php)
NASA-JPL	NASA Jet Propultion Laboratory (http://www.jpl.nasa.gov/)
NGA	National Geostatial-Intelligence Agency (http://www.nga.mil/portal/site/nga01/index.jsp?front_door=true)
NIMA	National Imagery and Mapping Agency (NIMA)
NOAA	National Oceanic & Atmospheric Administration (http://www.noaa.gov/)
XML	The Extensible Markup Language is ageneral-purpose markup language. Its primary purpose is to facilitate the sharing of data across different information systems, particularly via the Internet
NDVI	The Normalized Difference Vegetation Index is a simple numerical indicator that can be used to analyze remote sensing measurements, typically but not necessarily from a space platform, and assess whether the target being observed contains live green vegetation or not.
UML	Unified Modeling Language is a language that uses a series of diagrams to model the objects in a system.
USDMA	United States Defense Mapping Agency