POLITECNICO DI TORINO
Scuola di Dottorato

Research Doctorate Course in Environmental Geoengineering

XX Cycle

Final Dissertation

Geomatics for emergency management purposes: design of a global geodatabase

FRANCESCA PEREZ

Tutor

Prof. P. Boccardo

Co-ordinator of the Research Doctorate Course

Prof. G. Gecchele

Acknowledgements

During the preparation of this thesis a large group of people have supported my personal effort, and I would like to acknowledge their contributions.
First, I would like to express sincere appreciation to my supervisor Prof. P. Boccardo for always giving me encouragement and support through the years since I started my studies in the Geomatics area as well as for giving me several excellent ideas about this work.
This study took place when I participated to the ITHACA *Geodatabase/SDI unit* activities. I would like to express my deep and sincere gratitude to Andrea Ajmar, responsible of the unit, for his continuing support, suggestions and contributions throughout the thesis period. From the very beginning of the study to the end I have collaborate with him. I greatly admire his patience and friendliness which I enjoyed during my work with him. Actually, without him this work wouldn't exist!
My special thanks to Enrico Borgogno Mondino and Fabio Giulio Tonolo for their intellectual contributions to my student life, during the last years. In different ways, they encouraged and helped me to improve my research potential. Especially, I appreciate their enthusiasm and interest as well as their brilliant suggestions and discussions.
I would like to thank everyone at the DITAG Department of Politecnico di Torino for the friendly atmosphere and the stimulating environment. In particular, I would like to thank Tamara Bellone. I can't forget our discussions and conversations about study and life. And my colleagues from ITHACA that always made the work environment very pleasant.
A special thanks to Sabrina Ricauda for his great assistance in correcting the written English of this thesis.
Nevertheless, I will not thank some friends who have been annoying me by asking about the status of my thesis on all occasions during the last year…..
My final acknowledgment is directed to Domenico, which "has been obliged" to share with me the good and the bad of my Ph.D experience. Thanks for your patience!

Francesca

CONTENTS

INTRODUCTION

Background and motivation

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. Scientific predictions and evidence indicate that global climate changes are increasing the number of extreme events, creating more frequent and intensified natural hazards such as floods and windstorms. Population growth, urbanization and the inability of poor populations to escape from the vicious cycle of poverty are conditions to foresee that there will most likely be an increase in the number of people who are vulnerable to natural hazards, with a resulting increase of natural disasters and environmental emergencies.

In recent years, international preoccupation for disasters and their impacts has intensified and risen closer to the top of the development agenda. For many years, response to disasters was largely confined to emergency relief and short-term life-saving actions. But over the last two decades, the critical importance of disaster preparedness, mitigation, and prevention has been widely recognized.

The humanitarian and the United Nations system are therefore called to intensify their efforts to improve their capacity in order to provide support to the countries in need and to be better prepared to intervene. Such request came, amongst others, from the UN General Secretary in various occasions.
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. According to the UN reform, WFP is also the leader of logistics for UN and international bodies during emergency response operations.
WFP initiated a process to reinforce its capacity to be a leading force in the area of emergency response, improving its *Information Management* capacity in support to emergency preparedness and response. To do so, an agreement of collaboration with the recently formed *Information Technology for Humanitarian Assistance Cooperation and Action* (ITHACA) Association has been signed and a joint collaboration started in February 2007. One of the objectives of the collaboration is about the use of *Geomatics* and *Information Technology* instruments in the Early Warning and Early Impact analysis field.
Many worldwide experiences conducted in this area, show that the use of remote sensing and Geographic Information Systems (GIS) technologies, combined with up-to-date, reliable and easily accessible reference base geographic datasets, constitute the key factor for the success of emergency operations and for developing valuable natural disaster preparedness, mitigation and prevention systems. As a matter of fact, the unique characteristics associated with geographic, or geospatial, information technologies facilitate the integration of scientific, social and economic data through space and time, opening up interesting possibilities for monitoring, assessment and change detection activities, thus enabling better informed interventions in human and natural systems.

Besides its proven value, the geospatial information is an expensive resource and needs to be fully utilized to maximize the return on investment required for its generation, management and use. Reuse and sharing of spatial information for multiple purposes is an important approach applied in countries where investment in spatial data collection and in their appropriate management has advanced on the basis of its known asset value. Very substantial economic benefits have been estimated by countries that have moved in the direction of optimizing data reuse. However, it is still relatively easy to find examples of projects and other development activities from around the globe that required expensive recapture of essential spatial data because they were originally captured in unique or non-standard file formats, or perhaps discarded after initial use. Recapture of data has also been undertaken in many cases simply because its prior existence was known only by its originators.
The United Nations has not been immune to this problem, both within and between the multitude of entities that make up the Secretariat and its agencies, funds and programmes. Historically, the production and use of geospatial data within the UN entities has been accomplished by its component organizations, according to their individual needs and expertise. This has resulted in multiple efforts, reduced opportunities for sharing and reuse of data, and a unnecessary cost burden for the UN system as a whole.
Thus, a framework data development approach has been considered necessary. This has resulted in the proposal that implement an UN *Spatial Data Infrastructure* (*SDI*). The term SDI is used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. A SDI hosts geographic data and attributes, sufficient documentation (metadata), a means to discover, visualize and evaluate the data (catalogues and Web mapping), and some

methods to provide access to the geographic data. Beyond this, it will also host additional services or software to support applications of the data.

The concept of developing a *Spatial Data Infrastructure* to fulfil UN data management needs was duly approved by *United Nations Geographic Information Working Group* (*UNGIWG*) members in 2005 at their 6^{th} Plenary Meeting in Addis Ababa, in the context of a UN-specific SDI, or *UNSDI*.

The WFP, like all other UN agencies, has been called to develop a *Spatial Data Infrastructure*, according to the *UNGIWG* recommendations. Therefore, during the last year the different units of WFP involved in the use of geospatial data worked at defining and implementing a WFP SDI with the aim of contributing at the whole *UNSDI* project. This effort was coordinated and supported by the ITHACA association.

Aim of the study

The objective of the conducted research has been to investigate the better solution for collecting and organizing geospatial data within a suitable *geodatabase* with two main purposes:

- to support the WFP SDI effort: the development of consistent reusable themes of base cartographic content, known as *Framework*, *Fundamental* or *Core Data,* is recognized as a main and first ingredient in the construction of a SDI. Therefore, the definition of a geodatabase supporting all the WFP units dealing with GIS and geospatial data can be considered a fundamental and necessary step in the whole complex process of the development of the WFP SDI. Common used data provide key for the integration and, in the context of the SDI implementation, the definition of a *Core Data* geodatabase can be thought as one instrumentality to help improving interoperability, reducing expenses resulting from the inevitable duplications.

  Moreover, the major aim of the planned geodatabase is to supply all WFP users of a "minimum spatial dataset" which assures valuable geographic analyses and mapping, in support to decision makers during emergencies operations.
- to support all activities carried out by ITHACA: the planned geodatabase must constitute a suitable instrument which realizes the integration and the organization of the large geospatial data needed by all ITHACA units in their activities, allowing their effective distribution, sharing and reuse, avoiding any duplication. Moreover, the implemented solution must also guarantee the correct management and updating of the data, keeping their integrity.

  Finally, this instrument must also allow the easy and fast sharing of necessary information produced by ITHACA during Early Impact activities with the WFP's users engaged in the emergencies rescue operations.

In conclusion, the major expected output of the study carried out, described in this thesis, has been the design and the development of a global database and of related rules and procedures in order to correctly store, manage, and exchange geospatial data needed either by WFP humanitarian workers and ITHACA users. The developed database solution allows integrating and updating globally consistent geographic data coming from different sources in many formats, providing each user with the latest datasets, thus avoiding duplications and mistakes.

In methodological terms, the following procedure has been adopted:

- defining requirements, identification of all activities supported by the geodatabase, analysis of the data flows expected in all supported activities, examining existing data sources and relevant standards (particularly those proposed by the *UNGIWG*);
- development of the data model. The data model has been shaped according to specific needs and demands of the involved user groups within the different interested organizations. The adopted design techniques do not wander off the techniques proposed in literature for general database design, even if it has been necessary, in some steps, to consider the specific features of geographic data;
- geodatabase schema generation and implementation of the defined geographic database model as an ESRI ArcSDE Enterprise Geodatabase based on Oracle 10g as DBMS.

Structure of the thesis

The thesis is organised as follows:

- Chapter 1: in this chapter background theoretical concepts that are thought fundamental to understand the context into which this study is situated are reviewed and illustrated. First of all, the natural disaster and emergencies management themes are presented. Moreover, the activities of the WFP (World Food Programme), the major UN agency that is working in this field, are described.
- Chapter 2: in this chapter basic concepts about the UN SDI project are proposed. ITHACA association with its activities is also described.
- Chapter 3: in this chapter the general geodatabase design methodology adopted is presented. Moreover, fundamental concepts about analysis of spatial data, database and GIS systems, and a general overview of data modelling steps are provided.
- Chapter 4: in this chapter the methodology adopted in order to carry out the needs assessment phase with obtained results are presented. Results of the performed survey on available sources of global geographic data are also summarized.
- Chapter 5: in this chapter conceptual and logical database design phases are faced. The proposed conceptual data model is presented, as well as a review of the elements of the *Unified Modelling Language* (UML), used for the data model definition. Aspects of the geodatabase development phase, which covers also the logical design, are also described. Finally, potentialities and features of the adopted geodatabase solution are discussed.
- Chapter 6: this chapter shows a global view of the developed geodatabase content. Moreover, an example showing how the developed solution can support standard maps production activities performed by ITHACA Early Impact unit is also presented.
- Conclusions and outlook: this final part of the thesis discusses and summarizes the main results obtained in the conducted study.

1

NATURAL DISASTERS AND WFP

This study has grown in the disaster management context. Great attention has been paid to the disaster prevention, preparedness, and mitigation (response), with particular emphasis on the food security.
In this chapter general information about topics that are thought fundamental to understand the context of this thesis are provided. First of all, the natural disaster and emergencies management themes are presented. Moreover, the activities of the WFP (World Food Programme), the major UN agency that is working in this field, are described.

1.1 Introduction

Nowadays, the world is facing disasters on an unprecedented scale: based on the data in CRED's EM-DAT database[1] more than 255 million people were affected by natural disasters globally each year in the previous century, on average, with a range of 68 million to 618 million. Only in the last decade, 86% of all disaster-related deaths were caused by natural hazards, with just 14% resulting from technological disasters such as transport or industrial accidents. These disasters claimed an average of 58,000 lives annually, with a range of 10,000 to 123,000.

Scientific predictions and evidence indicates that global climate change will increase the number of extreme events, creating more frequent and intensified natural hazards. Population growth, urbanization and the inability of poor populations to escape from the vicious cycle of poverty are conditions to foresee that there will be most likely an increase in the number of people who are vulnerable to natural hazards, with a resulting increase of natural disasters and environmental emergencies.

There are four main natural disaster types, namely:

- floods and related disasters include floods, landslides, mudflows and avalanches;
- windstorms include storms, typhoons, cyclones, hurricanes, winter storms, tornadoes and tropical storms;
- geological disasters include earthquakes, volcanic eruptions and tidal waves;
- droughts and related disasters include droughts, extreme temperatures and wildfires.

Figure 1.1 displays the polynomial trends over the previous century of the four main disaster types previously presented.

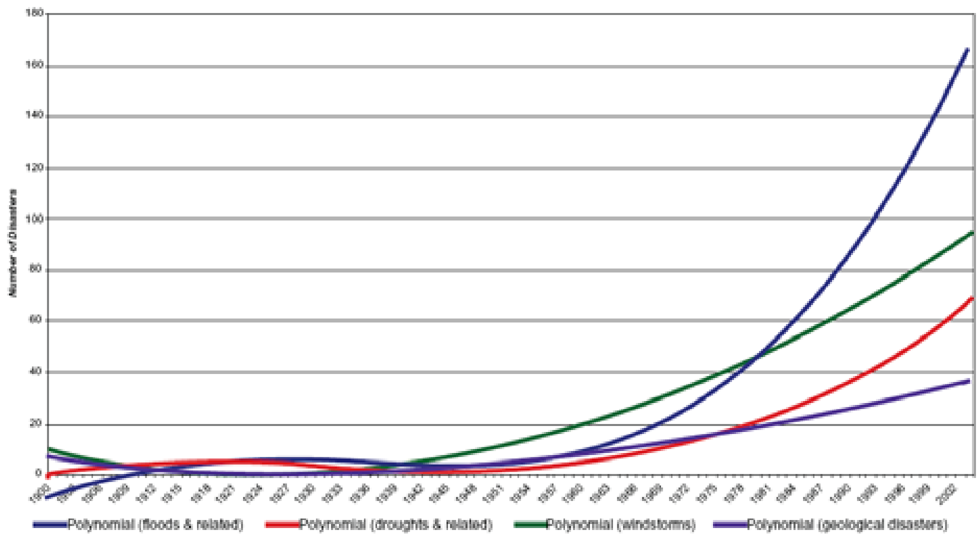


Figure 1.1 – Worldwide polynomial trends for the four major types of natural disasters: 1900-2003 (source: [D. Guha-Sapir, D. Hargitt, P. Hoyois, 2004])

[1] Since 1988 the Centre for Research on the Epidemiology of Disasters (CRED) has been maintaining an Emergency Events Database EM-DAT (http://www.emdat.be/index.html). EM-DAT was created with the initial support of the WHO and the Belgian Government. The EM-DAT contains essential core data on the occurrence and effects of over 16,000 disasters in the world from 1900 to present. The database is compiled from various sources, including UN agencies, non-governmental organizations, insurance companies, research institutes and press agencies.

As you can see, there was an increase in all the four categories, with floods and droughts showing the fastest rate of increase relative to geological disasters and windstorms. This underscores the importance of addressing issues such as environmental degradation, unplanned human settlements and the increasing vulnerabilities of populations.
Over the past 30 years, the number of reported natural disasters has increased steadily, however, the human impact shows very different trends over time (see Figure 1.2). The number of people affected has followed more or less steadily the same pattern of increase as the number of disasters, while the number of deaths has declined. This reduction can be partially attributed to the real effects of awareness and substantial expansion of disaster preparedness, mitigation and prevention activities.

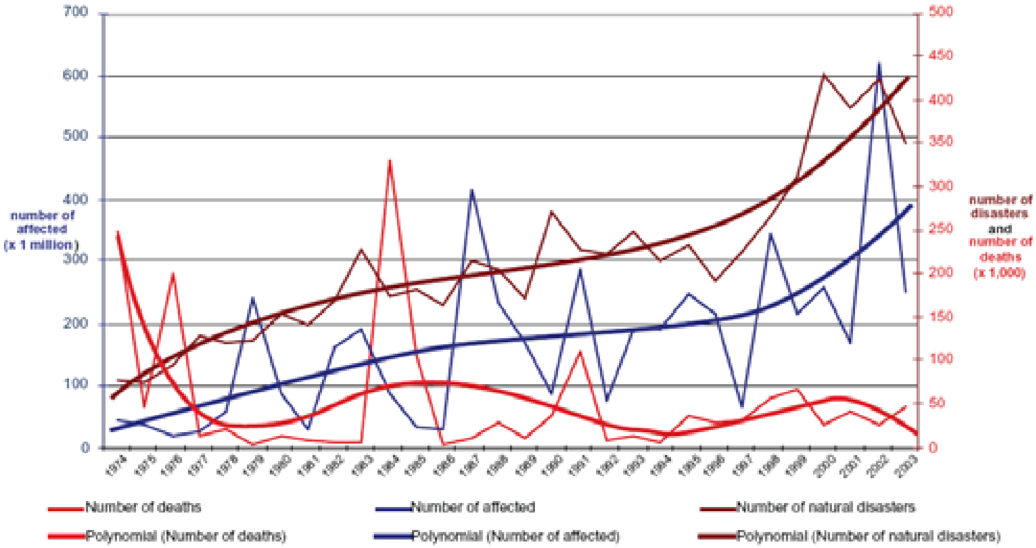


Figure 1.2 – Polynomial trends in numbers of natural disasters, persons killed and persons affected: 1974-2003 (source: [D. Guha-Sapir, D. Hargitt, P. Hoyois, 2004])

It is important to note that the number of reported people killed by a disaster is often limited to those who died as a direct result of the occurrence of the disaster. However, many deaths indirectly result from increases in malnutrition, poverty, diseases and the deterioration of living conditions and of health, sanitation and other basic services. EM-DAT defines the number of people affected by a disaster as "*people requiring immediate assistance during a period of emergency, that is requiring basic survival needs such as food, water, shelter, sanitation and immediate medical assistance*". So, the proposed figures may underestimate the real number of people affected over the longer term after a catastrophic event.

There are great differences in which continents are more affected by the different types of disasters. Asia and Africa bear a disproportionate burden of losses due to disasters. Over the last 30 years, approximately 88% of the total people reported killed and 96% of the people reported affected lived in these two regions alone. Of the total number of people killed by natural disasters worldwide over the last decade, more than 75% were in Asia. This figure rises above 98% for droughts and famines, 72% for earthquakes, 71% for avalanches and landslides and 56% for windstorms. Of the total of those reported killed by volcanic eruptions, Africa takes the lead with close to 62%. Only forest/scrub fire fatalities are more or less evenly spread out across the continents.

The risk of disaster is partially linked to the geophysical and meteorological characteristics of regions. This makes it easier to identify high-risk zones in the world based on their physical and climatic characteristics. Figure 1.3 shows the countries with the highest number of disasters. The classification of disaster occurrences into three classes represents a maximum average of one disaster per year for Class 1, one to four disasters annually for Class 2 and more than four disasters per year for Class 3.

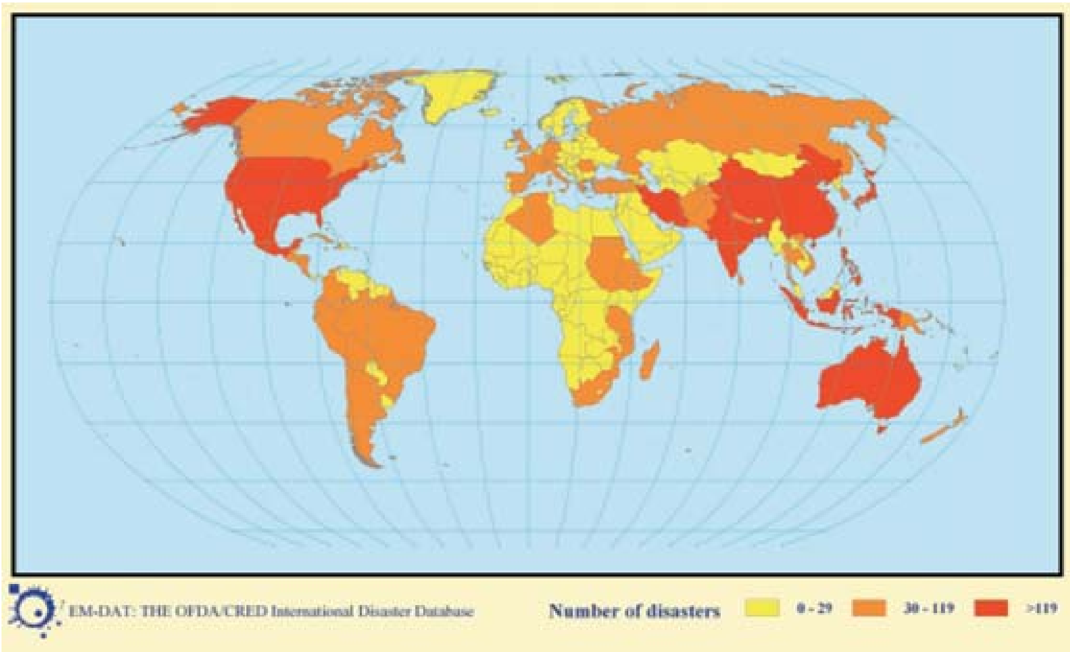


Figure 1.3 –Number of natural disasters by country: 1974-2003 (source: [D. Guha-Sapir, D. Hargitt, P. Hoyois, 2004])

Examining worldwide disaster data by the number of occurrences in each country provides substantial information about which countries suffer more disasters. However, looking at the total number of disasters relative to a country's area or the total number of people affected relative to a country's population can provide a different perspective on natural disaster occurrence and impact.

Considering the classification of countries according to the numbers of people affected standardised by 100,000 population, the pattern differs from the simple distribution of events. This analysis controls for variability among countries with very large and very small populations, although some countries covering large areas still come out on top.

In this way, South and East Asia, particularly India, Bangladesh and China, are in the highest category with a high proportion of their population being affected by natural disasters. All of them have areas of high population density, especially in river basins, and are home to populations whose livelihoods are often based on agriculture. When floods occur, the number of affected communities quickly reach into the hundred thousands, and in some cases, millions. Several countries in Africa are added to this list, although they did not figure in Figure 1.3 based on the ranking by disaster occurrence. In Figure 1.4, Southern Africa, Botswana, Zambia, Mozambique and Zimbabwe are in the highest category, together with Ethiopia and Mauritania.

Countries such as the United States, Mexico or Indonesia in Figure 1.3 were in the highest category based on disaster occurrence whereas in Figure 1.4 they are in the lowest category.

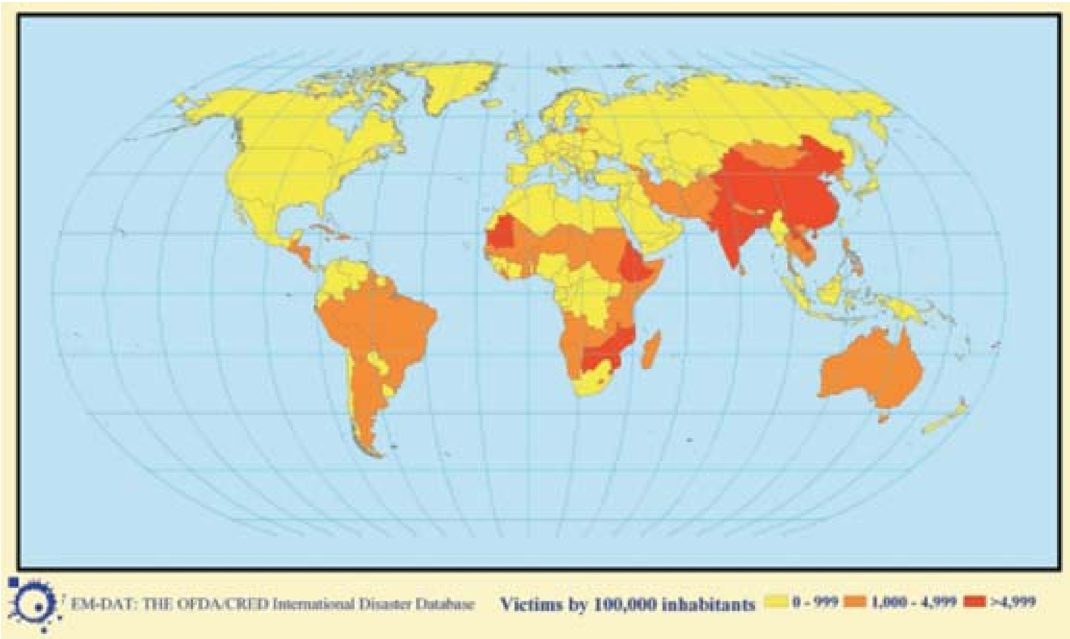


Figure 1.4 –Number of victims of natural disasters by 100,000 inhabitants: 1974-2003 (source: [D. Guha-Sapir, D. Hargitt, P. Hoyois, 2004])

One point that can be noted here, is that three types of natural disasters are recurring in the EM-DAT and they are:

- *drought*: a total of 640 droughts or famines were reported in EM-DAT over the last three decades. They caused the death of more than one million individuals and affected over 1.8 billion.
- *flood*: over the last 30 years, a total of 2,156 floods were reported in EM-DAT, resulting in the deaths of 206,303 people and affecting more than 2.6 billion.
- *windstorm*: over the 30-year study period, a total of 1,864 windstorms caused the death of 293,758 individuals and affected more than 557 million. Windstorms are among the most destructive disasters.

Another point can explain results shown in Figure 1.4. The occurrence of a natural disaster is based on the convergence of two factors. The first is the *hazard factor*, which is the risk of an earthquake, tornado, flood or other natural phenomenon. This factor is based on the geological, meteorological or ecological characteristics of a region, as aforementioned. The second is the *vulnerability factor*. This can be best described as the number of people at risk of being harmed by a hazard occurrence, whether it is through loss of lives or property, injuries or the disruption of livelihoods and economic activity. Several factors influence the vulnerability of a population. These can be mainly classified into four broad groups:

- physical aspects of vulnerability are linked to the exposure of the population to a potential hazard. Some examples are populations living in a flood plain or in a seismically active area;
- social vulnerability includes aspects such as population growth and the existence of conflicts and insecurity;

- economic vulnerability is linked to the population or country dependence on agriculture, the diversification of its economy, its financial assets or debts, as well as its access to basic infrastructure such as water, electricity, communication networks and health care;
- environmental vulnerability includes such factors as soil degradation and erosion, deforestation, chemical or biological pollution and the availability of water, whether for drinking, irrigation or other uses.

Therefore, it is simple to understand that disasters have a greater impact on poorer countries. The key factor in this relationship is the vulnerability of a population to a hazard occurrence. Poverty is closely linked to all four of the above discussed groups. Poor populations often live in high risk or environmentally degraded areas, have the least access to social safety infrastructure and have few savings or available credit. As with any complex situations, these factors do not act independently of each other but are intimately linked. Together they create conditions that increase the vulnerability of population to hazards, and hence a more frequent occurrence of disasters. It should come as no surprise that low-income countries have the highest relative number of victims of natural disasters.

The interaction of poverty and vulnerability is a vicious cycle that can only be broken through appropriate sustainable development mechanisms. Vulnerable populations are economically fragile by definition and are less able to recover from disasters. Savings or assets are few, or for many, non-existent. When a disaster strikes, it destroys not only existing wealth, but also income opportunities and livelihoods, increasing the vulnerability of the already poor population.

In future, natural hazards are not increasing significantly. But the number of people vulnerable and affected by disasters is definitely on the increase.

As Figures 1.5 and 1.6 show, if current trends continue, the number of large disasters will continue to rise in poorer countries, but at a slower rate, while they are likely to decrease in richer countries. At the same time, an increase in the frequency of small disasters in developed countries is a probable scenario, their numbers having risen substantially since 1995. For poorer countries, the number of small and medium-sized disasters also seem set to rise significantly.

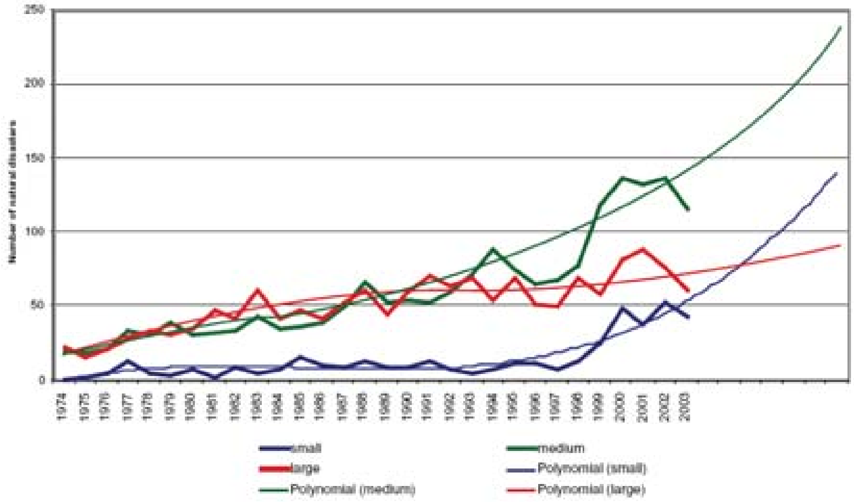


Figure 1.5 – Number of small, medium and large natural disasters: global forecast for years 2004-2010. Low and lower middle income countries. (source: [http://www.emdat.be/Database/terms.html])

In connection with this, international preoccupation with disasters and their impacts has intensified and risen closer to the top of the development agenda. Nevertheless, for many years, response to disasters was largely confined to emergency relief and short-term life-saving actions. But over the last two decades, the critical importance of disaster preparedness, mitigation, and prevention has been widely recognized. It has become increasingly apparent that a relatively small investment in disaster preparedness activities can save lives, reduce the number of people requiring emergency assistance and preserve vital economic assets, as well as reduce the cost of overall relief assistance to countries hit by disasters. In this way, disaster preparedness, mitigation and prevention activities can move out of the humanitarian agenda and become an organic part of the development framework.

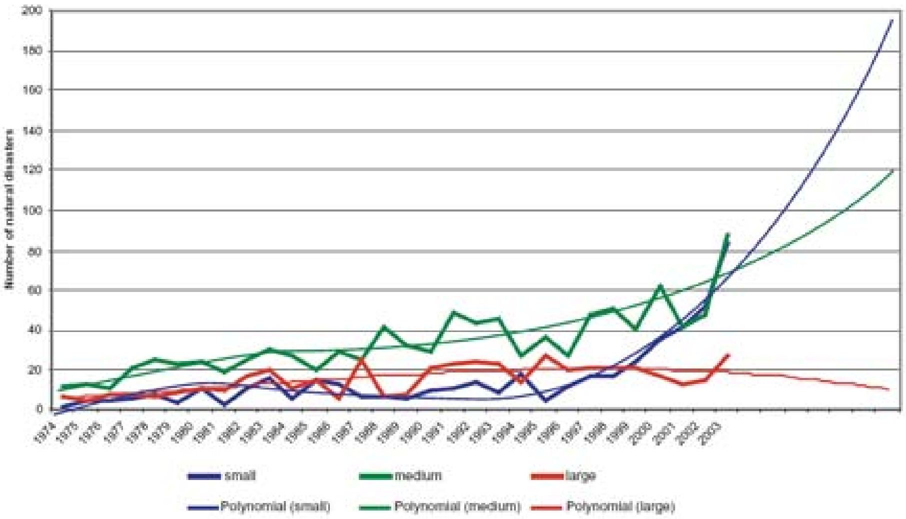


Figure 1.6 – Number of small, medium and large natural disasters: global forecast for years 2004-2010. High and upper middle income countries. (source: [http://www.emdat.be/Database/terms.html])

1.2 The World Food Programme

1.2.1 WFP introduction

The World Food Programme (WFP) is the United Nations' food aid agency and the world's front-line agency engaged in the fight against hunger.

Established in 1961 as a three-year experimental programme by the United Nations General Assembly and the Food and Agriculture Organization (FAO), WFP obtained continuing status in December 1965 "for as long as multilateral food aid is found feasible and desirable"[2].

In 2006, WFP worked in 80 countries with the host governments, United Nations agencies, international organizations and non-governmental organizations (NGOs) to help 16.4 million people through emergency operations. In the same year, WFP delivered a total of 4 million metric tons of food assistance by land, sea and air [*WFP Annual Performance Report 2007*[3]].

[2] http://www.wfp.org
[3] http://www.wfp.org

WFP is governed by the 36 member states that comprise its Executive Board and managed by an Executive Director who is jointly appointed by the United Nations General Secretary and the General Director of FAO. Moreover, WFP is made up of 80 operational Country Offices (CO), 6 Regional Bureaux (RB), 9 United Nations/Donor liaison offices, 5 support offices and its headquarters (HQ) in Rome.

WFP is the food aid arm of the United Nations system. WFP's mission statement is to use food aid to help eradicate hunger and poverty. This means giving food to people whose food consumption is inadequate to help them survive, grow or take advantage of development opportunities. In particular, WFP concentrates its efforts and resources on the neediest people and countries in accordance with the decision to provide at least 90 percent of WFP's development assistance to low-income, food-deficit countries.

Food aid is one of the many instruments that can help to promote *food security*, which is defined as "*access of all people at all times to the food needed for an active and healthy life"* [*FAO/WHO (1992) International Conference on Nutrition*[4]]. Nevertheless, the ultimate objective of food aid is the elimination of the need for food aid.

Finally, the core policies and strategies that govern WFP activities are to provide food aid:

- to save lives in refugee and other emergency situations;
- to improve the nutrition and quality of life of the most vulnerable people at critical times in their lives;
- to help build assets and promote the self-reliance of poor people and communities, particularly through labour-intensive works programmes.

In the first case, food aid is essential for social and humanitarian protection. It is used in a way that is as developmental as possible, consistent with saving lives. To the possible extent, the provision of relief food aid is coordinated with the relief assistance provided by other humanitarian organizations. In the second case, food aid is a pre-investment in human resources. In the third, it uses poor people's most abundant resource, their own labour, to create employment and income and to build the infrastructure necessary for sustained development. In this frame, the WFP plays a major role in the continuum from emergency relief to development. WFP gives priority to supporting disaster prevention, preparedness and mitigation and post-disaster rehabilitation activities as part of development programmes. On the other hand, emergency assistance is used to serve both relief and development purposes. In both cases the overall aim is to build self-reliance.

Overview of WFP's programmes

WFP's programmes fall into four main categories:

- <u>*emergency operations*</u>: response to disasters from natural or human causes;
- <u>*protracted relief and rehabilitation operations*</u>: recovery after a crisis;
- <u>*country programme and development activities*</u>: food aid for social and economic development;
- <u>*special operations*</u>: logistics to speed up the movement of food aid.

<u>*Emergency operations*</u>

When a disaster, from natural or human causes, occurs anywhere in the world and the government of the affected country makes an official request for WFP food aid, WFP considers an allocation of emergency food

[4] http://www.wfp.org

aid if: food aid is an appropriate response to the particular local situation, supply of the required food aid has not already been assured from other sources, and timely delivery of that aid to the intended beneficiaries is possible.

WFP’s emergency operations cover four main kinds of emergency:

- *sudden disasters*: natural or man-made disasters which affect food access and/or cause population displacements;
- *slow-onset disasters*: such as drought and crop failure;
- *refugee crises:* in close collaboration with the United Nations *Department of Humanitarian Affairs* (UNHCR);
- *complex emergencies*: involving elements such as conflict, widespread social and economic disruption, and requiring special United Nations coordination procedures.

Prior to any intervention, WFP first establishes whether food aid is needed and an appropriate response. WFP must:

- assess who needs food, how much and why;
- identify the population in need of assistance and the form that assistance should take;
- design the project and establish: the food ration to be distributed and non-food items required, the time period for assistance, and any security measures needed;
- organize transport and food storage facilities;
- identify an appropriate distribution system;
- develop a monitoring and evaluation plan.

Protracted relief and rehabilitation operations

Once WFP food aid has addressed the immediate needs of people affected by disasters, its operations focus on helping to rebuild their lives and communities.

WFP recognizes that for food-insecure people, the crisis continues after the disaster. For this reason, WFP’s *Protracted Relief and Recovery Operations* (PRROs) deal with the later stages of an emergency. The main objective of a PRRO is to help reestablish and stabilize food security and, if applicable, to contribute to the improvement of the nutritional status of vulnerable groups.

Country programme and development activities

In its country programme and development activities, WFP provides food to poor families for whom hunger is a real threat to health and productivity. WFP’s development projects aim to free people temporarily from having to provide food for themselves, and to give them time and resources to invest in assets such as houses, clinics, schools, new agricultural skills and medicine.

Special operations

WFP carries out special operations to rehabilitate and enhance transport infrastructure when necessary to allow the speedy and efficient delivery of food aid to answer emergency and protracted food needs.

Special operations are short term and usually complement emergency operations. Generally, they involve work on infrastructure and logistics. WFP special operations can cover:

- repairs to roads, bridges, railways;

- repairs to airports, port infrastructure and equipment;
- air operations;
- provision of common logistic services including joint logistics centres and communication initiatives.

1.2.2 Natural disasters and global food insecurity

As aforementioned, natural disasters are a major cause of global food insecurity, particularly in poor countries. In addition to the loss of life and shelter, the very basis of people's livelihoods is often undermined by recurrent natural hazards, leading to a decline in the capacity to meet basic needs, including food.
A *disaster* occurs when a damaging phenomenon affects large numbers of people who are vulnerable to its impact, and causes losses which are beyond the people's own capacities to recover in the short term.
It is important to recognize that:

- it is not the phenomenon itself (the earthquake, storm, flood, etc.) that is the disaster, but the direct and indirect effects that it has on people;
- the magnitude of the effects depend on the *vulnerability* of the people and their society to the particular type of event, as well as on the severity of the event itself;
- the vulnerability varies between different communities, households and individuals depending on the degree of their exposure to a hazard and their ability to resist its impact. That ability depends on the nature and diversity of their livelihoods and their assets.

Therefore, efforts to mitigate the potential impacts and effects of disasters need to consider not only the "hazard" but also the vulnerability of the people, and the determinants of that vulnerability. In some cases it may be possible to modify or reduce the effects of a natural hazard (e.g. through physical structures to reduce the likelihood of floods or landslides), but in most cases, as previously mentioned, it is possible, and may be more effective, to focus on reducing the vulnerability of the people at risk.
The terms "vulnerability" and "vulnerable" are used in many different contexts in different fields, institutions and agencies. Nevertheless, for WFP (and similar agencies) the term is related to food insecurity. In the specific context of analysing disaster vulnerability and planning for disaster mitigation, "*people who are vulnerable to food insecurity following a disaster are people who face a significant risk of being unable to meet their minimum food requirements in the event of disaster, especially those to which the region is prone*".
Moreover, for WFP, disaster mitigation efforts are concerned with the vulnerability of people's food security to particular natural hazards, such as a drought or flood. This is distinct from, and in addition to, the chronic food insecurity or what is sometimes called "structural vulnerability". People who are chronically food-insecure may also be vulnerable to disasters, but problems of chronic food insecurity need to be addressed through long-term development, not mitigation measures.

Impact of disasters on food security

Disasters may directly affect people and their food security by destroying lives, livelihoods, household reserves and assets, or indirectly affect people by weakening economic environment and markets, social environment, services and the natural resource base.
In the following a brief summary of the main hazards with which WFP is concerned:

- Drought occurs in all climatic zones and is a normal, recurrent feature of the Earth's climatic cycles. It is characterized by a prolonged period of dryness and reduced availability of water. High temperatures, unusual wind patterns and relatively low humidity may exacerbate the severity. In climatically dry regions, like some in Africa, high temperatures increase the evaporation rate, so when little moisture reaches the ground is quickly evaporated.

  A shortened rainy season or erratic rains also exacerbate drought-like conditions. The interaction between societal demands on water and climatic patterns of rainfall also play a large role in the intensity of a drought. As societal demands increase and rains decrease, desertification, or environmental and climatic change, appear.

  The four ways in which drought can be defined are:
  - *meteorological*, which is defined by significantly below normal precipitation in a regional context;
  - *agricultural*, which is characterized by below-normal levels of moisture necessary for crop production;
  - *hydrological*, which refers to low surface and subsurface levels of water and usually takes longer to develop;
  - *socio-economic*, which refers to a situation where physical water shortage affects people's livelihoods.

  Due to the longevity of droughts, they are particularly destructive to a community's long-term ability to sustain food production. Secondary effects of drought, such as decreased cloud and vegetative cover, exacerbate the climatic situation and decrease a community's ability to recuperate. Impacts of drought, crop failure and desertification are difficult to mitigate, as the cost of implementing agricultural or infrastructure changes.
- Flooding is the most widespread and common natural hazard in the world, typically causing more deaths annually than any other natural disaster. Heavy rains may cause rivers or lakes to spill over their banks, hurricanes or cyclones may carry copious amounts of water on low-lying coastal lands, and desert may become inundated with water due to heavy rains in nearby mountains.

  Flash flood waters move at incredibly high speeds and can occur in areas where it is not raining. Walls of water can reach metres in height and bring with them a deadly cargo of debris. Low-lying communities, located near areas of water and/or downstream from a dam, are particularly at risk.

  Monsoon rains typically come with wind-pattern changes on the Indian subcontinent during the shift from one season to another. As the different winds blow in opposite directions and air temperature increases, moisture levels drastically increase producing heavy, prolonged rains. This phenomena affects mainly the Indian subcontinent and Southeast Asia, but can also affect northern Australia and western Africa.

  Floods can be slow- or fast-rising, but normally take a few days to develop. Whatever the cause, flooding is responsible for huge amounts of damage annually to communities all over the world. The major effects of flooding on food security are mainly the destruction and loss of crops and livestock, which decrease short-term food availability.
- Tropical storms include cyclones, typhoons and hurricanes. The distinction between these strong storms depends on their locality. Tropical storms occur in the Atlantic Ocean, Caribbean, Gulf of Mexico and the eastern Pacific Ocean typically during the June – November period. They can occur,

however, at any time of year. Tropical storms vary in intensity, from thunder storms to hurricanes. Hurricanes, low-pressure systems with high wind speeds, are the most severe tropical storms.
Effects of tropical storms on food security are the destruction and loss of crops and agricultural lands, and livestock, which decrease short-term food availability.

- Earthquakes, tsunamis, and volcanic eruptions are all considered seismic events. As the Earth's tectonic plates move, earthquakes, tsunamis and volcanic eruptions may occur, depending on where the plates are shifting. Should this occur under a populated area, a disaster will often result.
  An earthquake is a sudden, rapid shaking of the ground caused by the shifting of tectonic plates, the breaking of the Earth's crust due to excessive crust stress, or human-made explosions. This shaking radiates outward from the central source of the earthquake, or epicentre. The waves vary in speed, depending on the material through which they travel. A strong earthquake can be followed in a short period of time by an equally strong aftershock, causing more damage. The majority of damage and death caused by an earthquake is due to the collapse of buildings and other structures.
  Similar to earthquakes, tsunamis are tidal waves, generated by a large scale disturbance such as an earthquake, volcanic eruption or explosion occurring under a body of water. They can cause damage far inland along coastlines. The configuration of the coastline, the shape of the ocean floor and the characteristics of waves all affect the intensity of a tsunami.
  Volcanoes are typically found near tectonic plate boundaries. A volcano forms at an opening in the tectonic plates, which allows the release of pressure through the expulsion of gasses, lava and ash.

1.2.3 Disaster mitigation and WFP

Disaster mitigation for recurrent natural disasters is one of the priority areas in the WFP's activities circle. Disaster mitigation refers to measures taken to reduce the likelihood of disasters occurring and/or to lessen the impacts of those that occur, in order to minimize the losses resulting from natural and other hazards.
For the WFP, disaster mitigation issue includes all measures taken to reduce catastrophic events effects on vulnerable people's access to food. It includes:

- *vulnerability reduction measures* to decrease on a permanent or long term basis:
  - the likelihood and intensity of events that could result in disaster;
  - the vulnerability of people, social and economic systems and infrastructure to the impact of potentially disastrous events;
- *preparedness measures* that ensure a readiness and ability to:
  - forecast and take precautionary measures in advance of an imminent event;
  - respond rapidly and appropriately in the event of a disaster by organizing and delivering timely and effective rescue, relief and recovery assistance.
- *remedial ("crisis avoidance") measures* taken in response to early warnings of an impending slow-onset crisis to prevent further deterioration and/or development of an humanitarian crisis.
- *response measures* taken in order to respond rapidly and appropriately to a disaster event.

Preventive and vulnerability reduction measures and preparedness measures are undertaken in the context of development plans and programmes. They should also be included in post-disaster recovery programmes and emergency operations when social awareness of disaster vulnerability and risk is high and donor funds may be more readily available. In all contexts, they should be planned as part of a comprehensive disaster mitigation strategy based on an analysis of disaster-related risks and the possibilities to reduce those risks, particularly for poor people who are food insecure or particularly vulnerable in the event of a crisis or disaster.

Organisational framework

In order to make the disaster mitigation branch more effective and responsive to the needs of beneficiaries, WFP undertook a decentralisation process aimed at putting decision makers in the field and providing them with the authority to more effectively pursue WFP's mandate.

The basic framework is described below:

- Country Offices have the primary responsibility for implementing emergency preparedness and response activities. Through their own networks of Sub or Area offices, liaison with the Government and other actors present in-country, COs are best placed to provide knowledge on, prepare for and respond to emergencies. When required, and depending on the scale of an emergency COs can call upon the RB or HQ directly for additional capacity and support.
- Regional Bureaux "provide strategic, policy and overall management guidance, direction, feedback, feed forward and support" to country offices. RBs take the lead when an emergency affects more than one country and are responsible for monitoring those countries in their region without a WFP presence. A network of *Emergency Preparedness and Planning Officers* (EPPOs) is being established in all RB. Their main function is to strengthen WFP's system wide capacity to respond to emergencies, especially in the area of early warning, contingency planning and emergency operations planning.
- The role of WFP Rome (and its various offices and units) is to support COs and RBs through provision of normative guidance, technical assistance and activation of a range of different response systems.

  In exceptional cases, the Assistant Executive Director/Director of Operations in Rome may be the decision-maker at the initial stages of an emergency, for instance in situations in which the response involved two Bureaux.

  A Situation Room in Rome is designed to facilitate communication and coordination of efforts to respond to an emergency. Under the management of the *Emergency Preparedness and Response Unit (*ODAP*),* it provides news services (information collation and analysis), geographic information systems (GIS) and mapping support, operational support to field offices and crisis management support. The Situation Room is connected to a network of standby regional operations centres, including meeting and audio-video communication facilities established in each Regional Bureau. The Situation Room also facilitates inter-agency coordination and collaboration through connections to the UN information exchange network, which includes other UN operations centres and the *Geographic Information Support Team* (GIST).

When an alert arises (which can be identified by the CO, RB or a unit in Rome), the *Operations Division* calls Emergency Task Force Meetings in the Situation Room to link the CO, RB with key players in Rome.

The *Emergency Preparedness and Response Unit* (ODAP), in addition to its management of the Situation Room, is responsible for enhancing the effectiveness and responsiveness of WFP operations at the global level, by providing normative guidance and technical support to field offices and Regional Bureaux, as well as by facilitating the sharing of information within the organisation.

The organizational charts for WFP and for *Operations Department* (OD) are shown in Figures 1.7 and 1.8.

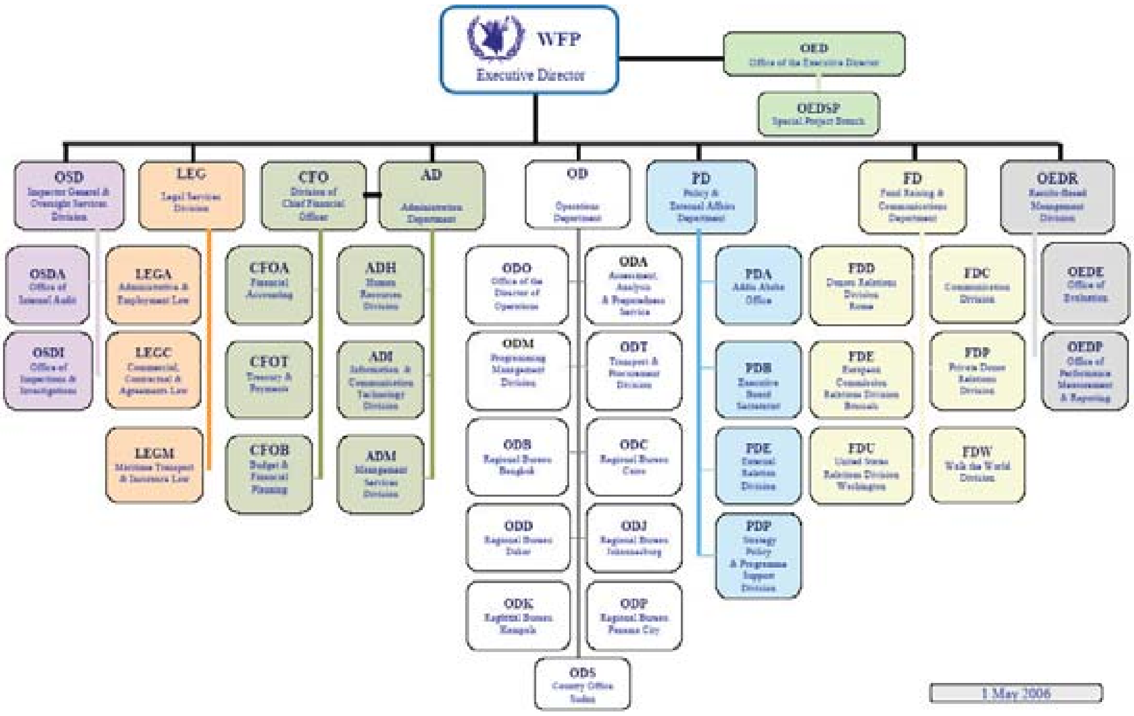


Figure 1.7 – WFP organizational chart – 2006 (source: [www.wfp.org])

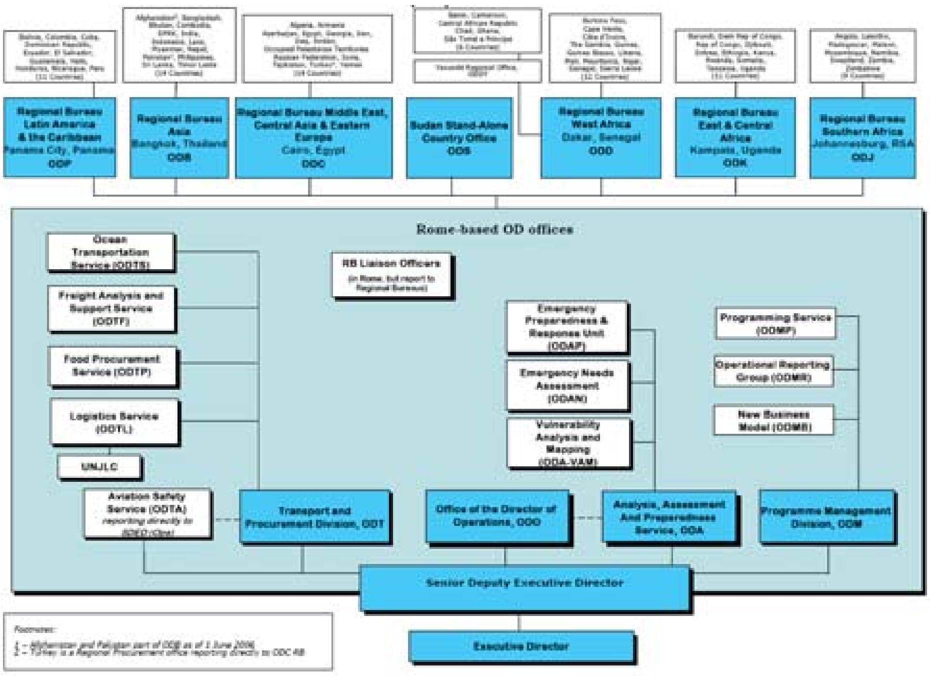


Figure 1.8 – Operation Department (OD) organizational chart – 2006 (source: [www.wfp.org])

Figure 1.9 shows the geographic distribution of WFP's Regional Bureaux.

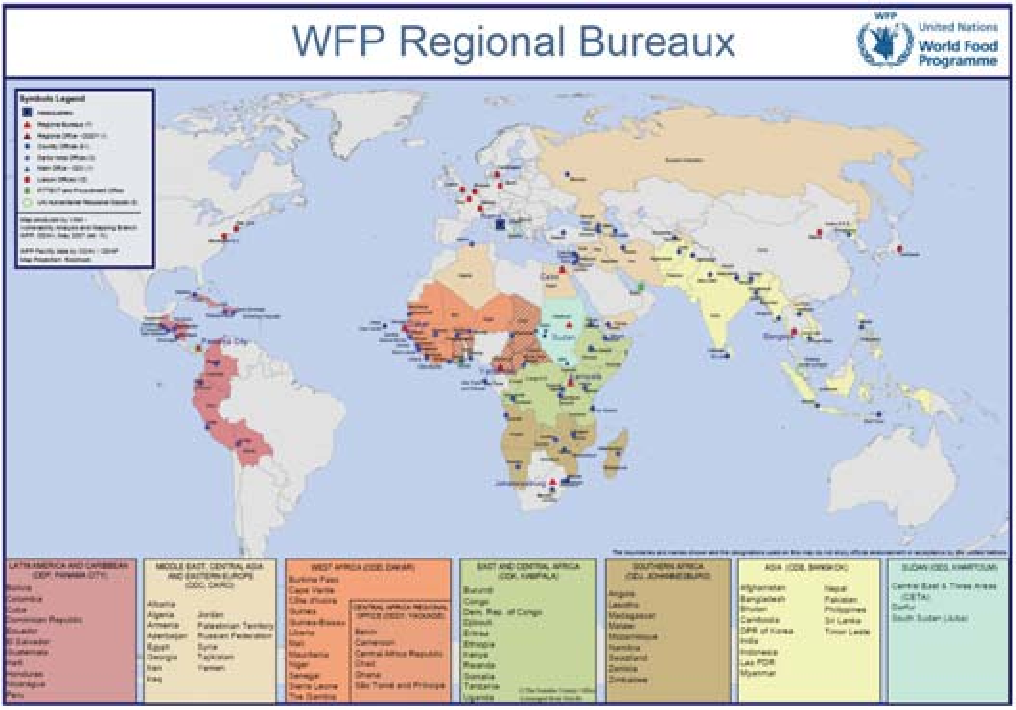


Figure 1. 9 – WFP Regional Bureaux (source: [www.wfp.org])

Elements of Emergency Preparedness and Emergency Response

Emergency Preparedness consists of actions, arrangements and procedures taken in anticipation of an emergency to ensure a rapid, effective and appropriate response.

These actions and arrangements can be broken down into a number of main categories including information preparedness, planning, programmes and mechanisms, and stand-by capacities.

Information preparedness consists of a number of elements, including having baseline information, undertaking early warning, having established public information systems, and good information management systems to manage information flows during emergencies.

Food security baseline information, for example vulnerability profiles provided by WFP's *Vulnerability Analysis and Mapping (*VAM*)*, is essential in determining the effect of a hazard on people's food security.

Anticipation is essential in order to be prepared. Early warning provides a way to identify, anticipate and prioritise emerging crises. It involves the monitoring of key socio-political events, forecasts, and other indicators of potential natural or man-made hazards. WFP does this by collecting and analysing primary early warning information and alerts produced by organisations specialising in conflicts, droughts, floods, volcanic activities, tropical storms and other hazards. Early warning enables early action by informing decision-makers early on about emerging crises so that they take appropriate actions to prepare and respond to developing emergencies. Early warning also enables strategic prioritisation over the use of limited organisational resources for emergency preparedness.

Too much and contradictory information is a common feature of the emergency environment. Management systems must be in place to enable effective flows of both strategic and operational information during emergencies so that the right information flows to the right people at the right time. Communication channels for production and directing of relevant and succinct information for both internal and external purposes need to be adapted to the specific circumstances of an emergency. The WFP Situation Room described above is designed as a tool to facilitate emergency information management for WFP and makes use of systems like the WFP *Emergency Preparedness and Response Web* (EPWeb[5]) and GIS.

Moreover, at the inter-agency level, OCHA's Reliefweb[6] provides the humanitarian community a common information management tool which supports information sharing and information preparedness across the humanitarian community.

An essential element of being prepared to respond to emergencies is having functional mechanisms in place to do so. Both internal WFP programmes and inter-agency mechanisms provide a basis on which WFP and partners can respond. In WFP there are five internal programme types that can be used as a basis from which to respond to emergencies: Immediate Response Emergency Operation (IR-EMOP), Emergency Operations (EMOP), Protracted Relief and Recovery Operations (PRRO), Special Operations (SO), and Country Programmes (CP).

Afterwards, stand-by capacities are a major element of WFP preparedness and enable WFP to mobilise rapidly food and cash, as well as staff and equipment, all of which are fundamental to the success of any emergency response, particularly during the critical initial stages of an emergency response. WFP has an array of both internal and external (other UN agency, NGO or donor) resources, available at short notice through pre-defined arrangements and agreements, through which it enhances preparedness and ensures a rapid response. Stand-by capacities fall in five major groups, including food, human resources, equipment, transport and emergency funding. Particularly, the WFP has a wide range of transport capacities at its disposal, trucking fleets and those on long term lease, and vessels and aircraft under contract to serve existing operations. Additional capacity can also be called upon through donor Logistics Service Packages. These are predefined modular units of personnel and equipment made available to WFP or other agencies by a donor government or major NGO to perform a specific function within a set period.

Emergency Response is an organized set of actions taken to ensure the provision of appropriate emergency food assistance to targeted food insecure populations. When a population faces an acute food shortage and WFP assistance is required, the expansion, adaptation or creation of a food aid supply chain is necessary.

In principle, Country and Regional Directors determine respectively what constitutes an emergency in their country or region and are responsible for activating emergency response mechanisms.

Based on a request from a government of a country or from the Secretary-General of the United Nations, WFP must assess whether emergency food assistance is required and whether it will seek to provide for these needs. WFP's role in responding to emergencies starts with assessment of food aid needs and related requirements.

Based on the results of the assessment, the scale of the emergency and WFP programmes are defined. An operational plan is then adopted. This is a management tool that defines in detail what will be done, how, when and by whom, including decision-making responsibilities and management procedures. It is different from an EMOP or PRRO as it is a living document and contains details on how an operation will be implemented.

---

[5] This is the WFP's corporate system for information management in emergencies (http://epweb.wfp.org/).

[6] http://www.reliefweb.int

2

RESEARCH BACKGROUND

The ITHACA (Information Technology for Humanitarian Assistance, Cooperation and Action) association and its activities are here presented. Recently established, ITHACA, in virtue of a collaboration with WFP, has the main goal to conduct operational and research activities in the field of geomatics for analysis, evaluation and mitigation of natural and manmade hazards.
A fundamental ITHACA's project, which is still ongoing, concerns the supply of support activities to WFP in order to develop and implement a WFP-specific Spatial Data Infrastructure (SDI). The carried out study, which is summarized in this thesis, has been developed in the context of this project.

2.1 A UN *Spatial Data Infrastructure* in support of humanitarian actions

2.1.1 UNSDI concepts

Introduction

At the United Nations Conference on Environment and Development in Rio de Janeiro in 1992, a major resolution was passed to focus on reversing the impacts caused by environmental deterioration. The Agenda 21 resolution establishes measures to address deforestation, pollution, depletion of fish stocks, and management of toxic wastes to name a few. The importance of geographic information to support decision-making and management of these growing national, regional, and global issues was cited as critical at the 1992 Rio Summit, and at a special session of the United Nations General Assembly convened in 1997 to appraise the implementation of the Agenda 21. In 2003, at the World Summit on Sustainable Development in Johannesburg, South Africa, a landmark effort was made to illustrate the capabilities, benefits, and possibilities of using online digital geographic information for sustainable development.

Geographic information, or geospatial data, are vital for the execution of many United Nations activities. These range from peacekeeping and humanitarian assistance in which knowledge of the locality and terrain are indispensable, to development, environment and health programmes in which geographically referenced data are equally critical for planning and coordination. Decision-makers are benefiting from geographic information, together with the associated infrastructures (i.e. *Spatial Data Infrastructures* or SDI) that support information discovery, access, and use of this information in the decision-making process.
In this frame, rapidly evolving technologies include remote sensing, geographic information systems (GIS), global positioning systems (GPS), and an array of environmental monitoring platforms in space that have in common digital data recording, time series data acquisition and data geo-referencing capabilities. Many sources of spatial data exist, and they include, for instance, topographic surveys, satellite imagery and aerial photographs, censuses and household surveys, and environmental inventories.
Spatial data are the fuel upon which the analytical power of GIS depends. Their unique characteristics, associated with geo-information technologies, facilitate the integration of scientific, social and economic data through space and time, opening up exciting possibilities for monitoring, assessment and change detection, thus enabling better informed interventions in human and natural systems. For example, spatial representations of the state of natural resources, sensitive environments, agricultural activities, and socio-economic phenomena are today most often achieved by integrating layers of relevant spatial data in a GIS.
Integrating spatial information in this manner is critical for the long-term sustainable use of resources and it is also of great importance in the design, targeting and implementation of short-term interventions when situations such as wars, food security crises or natural disasters arise unexpectedly. Therefore, the ready availability and cost-effective management of geospatial information is central for the UN in order to raise operational efficiency, in terms of short-term emergency response capacities, long-term risk reduction, development and environmental protection activities, which are the three pillars of sustainability.

The technology to acquire, process, analyse, display and manage massive amounts of geographic data has improved exponentially in recent decades, but this wealth of spatial information has not been matched with the opportunity or ability of users in general to discover, access, evaluate, utilize and share them.

Despite the proven value of spatial information, it is an expensive resource and needs to be fully utilized to maximize the return on investment required for its generation, management and use. Reuse of pooled spatial information for multiple purposes is an important approach applied in countries where investment in spatial data collection and in their appropriate management has advanced on the basis of its known asset value. Very substantial economic benefits have been estimated by countries that have moved in the direction of optimising data reuse. But it is possible to find examples of projects and development activities from around the globe that required expensive recapture of essential spatial data because they were originally captured in unique or non-standard file formats, or perhaps discarded after initial use. Recapture of data has also been undertaken in many cases simply because its prior existence was known only by its originators.

The United Nations have not been immune to this problem, both within and between the multitude of entities that make up the Secretariat and its agencies, funds and programmes. Organizations that could be considered 'significant' users of geospatial data and applications in the United Nations include DPKO/UNCS, FAO, OCHA, UNEP, UNESCO, UNHCR, UNICEF, UNOSAT, WFP, WMO and WHO.

Historically, UN entities acquired or developed a variety of unique software tools and GIS systems in order to solve specific problems that benefited from geospatial data and information. Although created at considerable expense, the resultant systems often turned out to be unique, lacked interoperable standards and contained poorly documented data of unspecified quality. These early, agency-specific actions resulted in a legacy of limited opportunities for sharing the tools and data for uses beyond their originally intended purposes, and reduced the possibilities for growing these applications further in the future.

Information management systems of some national governments also evolved in similar ways in the past allowing only limited interoperability and data sharing between internal authorities, and even less with external bodies such as the UN. Consequently, the sharing of spatial information both within and between United Nations authorities and within and between many of its member states has also been hampered historically by the lack of information infrastructures that promote data access, sharing and interoperability through agreed policies and standards.

In order to maximize the potential benefits of geospatial data for UN operations and business processes, greater coherence in their system-wide management is essential.

The SDI definition

The term *Spatial Data Infrastructure* (*SDI*) is used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. A SDI provides a basis for spatial data discovery, evaluation, dissemination, coordination, use/re-use, sharing and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general.

The word infrastructure is used to promote the concept of a reliable, supporting environment, analogous to a road or telecommunications network, that, in this case, facilitates the access to geographically-related information using a minimum set of standard practices, protocols, and specifications. Like roads and wires, a SDI facilitates the conveyance of virtually unlimited packages of geographic information.

The establishment of a SDI is driven by: a) the spatial data itself characterized by high generation costs, duplication, digital availability, inconsistency, increasing volume; b) the expanding Internet capabilities and capacity; and c) the need to easily access and use the geospatial data and information for knowledge-based

decision makers at various levels of government. A SDI is therefore more than a single data set or database; a SDI hosts geographic data and attributes, sufficient documentation (metadata), a means to discover, visualize, and evaluate the data (catalogues and Web mapping), and some methods to provide access to the geographic data. Beyond this, it requires additional services or software to support applications of the data. To make an SDI functional, it is also necessary the organisational agreements needed to coordinate and administer it on a local, regional, national, and or trans-national scale.
Although the core SDI concept includes within its scope neither base data collection activities or myriad applications built upon it, the infrastructure provides the ideal environment to connect applications to data, influencing both data collection and applications construction through minimal appropriate standards and policies.
Relevant aspects in a *Spatial Data Infrastructure* implementation are:

- the development of consistent reusable themes of base cartographic content, known as *Core data* is recognized as a common ingredient in the construction of national and global SDIs to provide common data collection schemas;
- metadata are a key ingredient in supporting the discovery, evaluation, and application of geographic data beyond the originating organisation or project.
- geospatial data catalogues are a means to publish descriptions of geospatial data holdings in a standard way to permit search across multiple servers. Geospatial data catalogues are discovery and access systems that use metadata as the target for query on raster, vector, and tabular geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among SDI communities.
- the primary view of geographic data has historically been through maps. In the context of SDIs, it is increasingly useful to provide mapped or graphical views of geospatial data through online mapping interfaces. This can satisfy many of the needs of novice or browse users of data without requiring download of the full data. Although it is not a replacement for direct data access, it satisfies a broad requirement for public interaction with geospatial information. Assuming that data are being used for their correct purpose and at an appropriate scale, maps can quickly portray a large amount of information to the inquirer. Current best practice in on-line mapping deals with the results of the OpenGIS Consortium.
- once spatial data of interest have been located and evaluated, using the catalogue and online mapping techniques, access to detailed geospatial data in its packaged form is often required by advanced users or application software. Access involves the order, packaging and delivery, offline or online, of the data (coordinates and attributes according to the form of the data) specified.
- the application of special services hold great promise in realizing true Web-based GIS interactions on data in support of decision making.

The UNSDI

The will of a coherence in the management of geospatial information in the UN system was acknowledged at an operational level in March 2000 with the establishment of the United Nations *Geographic Information Working Group* (UNGIWG). UNGIWG comprises a voluntary network of professionals working in the fields of cartography and geographic information science that seeks to address common geospatial issues: maps, boundaries, data exchange, standards, naming conventions, and location. It works directly with non-governmental organizations, research institutions and industry to develop and maintain common geographic databases and geospatial technologies that enhance guidance and operational capabilities.

UNGIWG also coordinates activities and formulates policies concerning geographic information that affect the work of UN Organizations and Member States. As a result, the working group identified the need for an enterprise-wide solution that ensures the coherent use and management of geospatial data for UN activities. The concept of developing a *Spatial Data Infrastructure* (SDI) to support these data management needs was duly approved by UNGIWG members in 2005 at their 6th Plenary Meeting in Addis Ababa, in the context of a UN-specific SDI, or UNSDI. The development of a UNSDI has been considered essential for increasing system coherence using and exchanging geospatial data and information for UN activities. So, the UNGIWG has been laying the foundations for the UNSDI construction, which is still ongoing.

At the moment, among the users community, four primary business cases drive the need for a UNSDI:

1. provision of spatial data and information including:
   - cartographic data, satellite imagery and GIS services;
   - thematic data;
   - data from, and for, global and regional environmental observations and assessments;
   - data to support emergency response and disaster preparedness.
2. development of common data services to:
   - increase sharing and potential reuse of data internally and externally for immediate partners such as member states;
   - adopt/develop data standards, metadata and the provision of technical infrastructure.
3. capacity building:
   - internal, UN capacity building to increase efficiency and effectiveness;
   - external capacity building in spatial information related subjects, primarily with member states and regions to strengthen abilities in sharing and utilizing spatial data.
4. promotion of partnerships and cooperation:
   - strategic partnerships promoted to spatial data access and support capacity building.

During the last years, UN bodies dependent upon geospatial data for normative activities have steadily grown to accept the need to structure information management more effectively within their own agencies, and with their partners, to improve data reliability, exchange, and utilization.

Significant initiatives have accordingly been initiated by individual agencies or groups of UN partners to address common data sharing issues. OCHA (*United Nations Office for the Coordination of Humanitarian Affairs*[1]) for example, proposed the development of modular, service-oriented, and standards-compliant web based information architectures while UNEP (*United Nations environment Programme*[2]) has proposed a preliminary SDI. FAO (*Food and Agriculture Organization of the United Nations*) and other agencies steered the development of GeoNetwork[3] open source, a software tool that facilitates interoperability and constitutes a single point of access to geospatial data and systems. UNOSAT (*United Nations Institute for Training and*

[1] http://ochaonline.un.org/
[2] http://www.unep.org/
[3] http://www.fao.org/geonetwork/

Research Operational Satellite Applications Programme[4]) and UNOOSA (*United Nations Office for Outer Space Affairs*[5]) also facilitated operational access to satellite data coverage to assist in responding effectively to natural and manmade disasters. These are examples of inter-agency cooperation supporting the evolution of a UNSDI.
A complete description of all the UNSDI aspects is beyond the purposes of this work. For in-depth information on this topic see the documents specified in the bibliography.

Other SDI initiatives and implications for the UNSDI

In order to define the major issues involved in the SDI development process, the UNGIWG gave great importance to the analysis of other national, regional and global SDI initiatives. Considered initiatives have been reviewed and discussed in a "UNSDI Compendium" document, proposed in the bibliography.
Among these, two in particular are here briefly presented, since their knowledge is necessary to better understand topics proposed in Chapter 4. These are:

- *The Global Spatial Data Infrastructure (GSDI) Initiative*

  The *Global Spatial Data Infrastructure* (GSDI) Association[6] was formed to promote the SDI concept and support local, national and international spatial data infrastructure developments. It is an inclusive organization of organizations, agencies, firms, and individuals from around the world that promotes international cooperation and collaboration concerning spatial data issues, data and systems that make up an SDI. For these purposes, GSDI conferences have been held annually since 1996 and have done much to promote the SDI concept globally.

  The Global Spatial Data Infrastructure (GSDI) community is trying to develop a global spatial data infrastructure characterised by international standards, guidelines and policies to enhance data management and access, and support global economic growth, and associated social and environmental objectives. In this regard it has developed an impressive support network, and useful tools such as the publication of "Developing Spatial Data Infrastructures: The SDI Cookbook" that is cited in bibliography.

  Certainly, the UNSDI is profoundly influenced by the cooperative developments underway in the GSDI.

- *The Global Map initiative*

  The Global Map project is an international collaborative initiative based on the voluntary participation of national mapping organizations of the world, aiming to develop a globally homogeneous geographic data set at a scale of 1:1 million. The Government of Japan and the Geographical Survey Institute (GSI) of Japan conduct this initiative.

  The primary objective of the Global Map project is to contribute to sustainable development through the provision of a base framework geographic data set. About 130 countries and regions, that correspond to more than 80% of the Earth's land area, are participating in the project. Data for completed countries are downloadable through the Internet from the ISCGM website[7]. Regional and national mapping centres participate in this initiative through joint training courses and workshops.

[4] http://unosat.web.cern.ch/unosat/
[5] http://www.unoosa.org/oosa/index.html
[6] http://www.gsdi.org/
[7] http://www.iscgm.org/cgi-bin/fswiki/wiki.cgi

Activities of the United Nations are emphasized in the Global Map Project, to advance its contribution to international society ideals. Moreover, the Global Map project supplies an important framework dataset of digital cartographic information that conforms to international mapping and data standards which support is wide dissemination and use in the future.

2.1.2 A Spatial Data Infrastructure for the WFP

The WFP, like all other UN agencies, has been called to develop a *Spatial Data Infrastructure*, according to the UNGIWG recommendations, with the aim of contributing at the whole UNSDI project. This effort was coordinated and supported by ITHACA association.
Three units, operating inside WFP and dependent upon geospatial data and use of GIS instruments for their activities, are identified:

- the *Emergency Preparedness and Response Unit* (ODAP);
- the *Vulnerability Analysis and Mapping Unit* (ODAV - VAM);
- the *UN Joint Logistic Center* (UNJLC).

Different GIS services and cartographic data support the operations of these units. As a matter of fact, these different units of WFP have developed, either internally or through partnerships, thematic datasets that support their normative and project related activities on a variety of operational scales. These data have also potential value for other UN agencies, partners and member states dealing with issues of environmentally sustainable development and support for human health and well-being. However, awareness of, access to, and interoperability regarding these data have historically remained sub-optimal.

The *Vulnerability Analysis and Mapping* (*VAM*) plays a dual role in the design and management of WFP's emergency and development activities. As an information tool, VAM provides up-to-date information about who the hungry poor are, why they are vulnerable, and where food aid should be targeted. When WFP draws up its emergency and development operations in any country, its first priority is always to identify the hungry poor: who is vulnerable, where they live and why they are hungry.
VAM activities aim to assess the food security and vulnerability situation in a country before, during and after a crisis; they also provide a framework for continually assessing the food security and vulnerability status of WFP beneficiaries. Food security and vulnerability usually reflect the complex interplay of several factors: agriculture, markets, employment opportunities, food policies, international commodity prices, weather and climate, health and nutrition, etc. The final products of the VAM analyses are: analytical reports and assessments, maps and databases. Particularly, VAM assessments and maps are the final expression of a longer and systematic process of data collection and analysis often making use of satellite images showing agro-climatic conditions.
Although VAM output is specifically designed to assist vital WFP programming decisions, the unit's information products also assist the wider international community, from national and local governments to UN agencies and non-governmental organisations (NGOs).

Moreover, VAM is adopting an "enterprise" approach in managing geographic information. This is referred to the VAM *Spatial Information Environment* (SIE)[8], using the GeoNetwork software that FAO has developed. SIE enables WFP Country Offices (CO), Regional Bureaux (RB), and HQ to access and exchange geo-referenced food security databases and cartographic products from a variety of sources. SIE includes tools for standardization, infrastructure to support the appropriate use of spatial information and collaborative efforts to increase accessibility to original and derived information within WFP and with information partners. SIE users in WFP-Rome will be connected directly to VAM spatial data resources and tools by a high-speed network, while those outside Rome will access the same data through standard Internet protocols and an Internet Map Server (IMS). Finally, the SIE will also provide access to a worldwide network of shared public domain data sets.

In the field of *Disaster Preparedness and Response* (see also 1.2.3), to mitigate the impacts of disasters on populations, geospatial technologies increasingly play a role in helping WFP activities, in order to anticipate, prepare and meet the challenges of disasters in a more timely and effective manner. *ODAP* is required to respond to emergencies at a variety of scales, from regional down to national, sub-national, but also local scales in the field. Here, it employs or utilizes the spatial outputs of GIS applications to plan interventions as evolving circumstances dictate, and the effectiveness of this approach depends also on the availability and quality of geospatial data on site. Information, for instance regarding the local infrastructure, population, and the resources that are impacted or threatened by a crisis, is critical to develop appropriate mitigation strategies. Rapid access to the most recent satellite data is important in this regard for assessing damage to infrastructure and for estimating the affected population.
ODAP products, as cartographic data showing the impact of the disaster event on the population and on the environment, are very useful in the decision making process in order to grant the correct management of emergency actions. These products also help to answer key questions as how much food aid is needed and how to deliver it to the hit population. For this purposes, the use of remotely sensed data, combined with updated and reliable reference base datasets, allows to perform accurate and timely assessments.
This activity raises several questions including: how the required thematic and cartographic data are made available in the field and at short notice, and how to assess the quality of the data itself. The adaptability and interoperability of applications tools that support the necessary analyses within and across the users needs to be considered too. Partners outside the WFP system, such as national governments and their agencies, NGOs, donor agencies, regional and international organizations that participate in many of the field operations, also need fast and reliable access to the same geospatial information. Also, they may want to share the information they hold with UN bodies for reasons of combining forces to address a problem. Common standards that increase interoperability and ease use of data are an issue of concern in these circumstances.
As one of the most difficult aspects of the emergency environment is the complexity, quantity and urgent need for good information and analysis upon which to make decisions, the importance of information management should not be under-estimated. To facilitate this important function, the ODAP unit is also responsible for the management of the WFP *Emergency Preparedness and Response Web* (EPWEB)[9]. This

[8] The VAM SIE portal is accessible at http://vam.wfp.org/vamsie/srv/en/main.home
[9] http://epweb.wfp.org

portal constitutes an important tool which allows easy and fast exchange of and access to the data and information necessary during emergency situations.

These are the map repositories accessible through the EPWEB portal:

- *the ODAP Map Repository*: searchable repository that gives users access to a large collection of printable geographic maps and types of maps relevant for emergency operations;
- *the VAM-SIE*;
- *the Topographic Maps repository*: repository that provides very detailed maps for locations worldwide. Maps can be downloaded in tiles and printed with a small printer. The map tiles can be used separately or stitched together, as larger maps and wall charts.

The *United Nations Joint Logistic Center* (UNJLC) is an inter-agency Humanitarian Common Service. Its mandate is to provide Logistics Information Management support and services. The UNJLC operates under the custodianship of WFP which is responsible for the administrative and financial oversight of the Centre. A UNJLC Core Unit is established in Rome.

The UNJLC is activated when intensified field-based inter-agency logistics coordination is required at the onset of natural and/or man-made disasters. Once mobilised, the UNJLC seeks the widest possible participation among all humanitarian logistics actors (UN and NGO). So, the UNJLC aims to facilitate and support the coordination of logistics capabilities among cooperating humanitarian agencies through the provision of Information Management, GIS and commodity tracking services.

As a matter of fact, the UNJLC acts as an information sharing platform for the gathering, collection, analysis and dissemination of information required for planning and managing humanitarian logistics operations. This includes:

- *Logistics Information*: gathering of general logistics information with particular attention to sea and airlift capacity and availability, transport procedures and schedules, infrastructure assessments and updates;
- *Mapping*: providing UN agencies and NGOs with accurate and constantly updated maps and geographic data. These provide operational overviews and logistical information to assist all humanitarian actors in their decision-making;
- *Commodity and Pipeline Tracking*: through the Joint Supply Tracking (JST) for prioritization and consolidation of cargo movement;
- *Information Sharing Tools*: the gathered information is published on the UNJLC webpage[10] to enable all relief actors to easily access strategic logistics information. A variety of dissemination mechanisms are employed, including interagency meetings, situation reports, surveys, bulletins and CD-ROMs.

Finally, the UNGIWG has committed to the UNJLC the definition and the implementation of the Transportation part of the UN global database.

[10] www.unjlc.org

In conclusion, during the last year, the different units of WFP involved in the use of geospatial data are working at defining and implementing a WFP SDI with the aim of contributing at the UNSDI project. As aforementioned, this WFP users' effort is supported by ITHACA association.
The development of a WFP SDI must allow:

- the collection and the reorganization of the whole complex of geospatial data currently used by the different units of WFP interested, with particular emphasis to the access to public domain geographic resources;
- the definition of a "minimum spatial dataset" which assures valuable geographic analyses and mapping, in support to decision making during emergencies operations;
- the sharing and the distribution of various datasets, outputs and information among WFP actors (also HQ Rome / field exchange) and between WFP agencies and partners also outside the WFP system;
- the normalization of data content;
- the data reliability and integrity, allowing at the same time instruments for a distributed environment aimed at data managing;
- the definition of new web based tools.

2.2 The ITHACA association

2.2.1 Introduction

ITHACA (*Information Technology for Humanitarian Assistance, Cooperation and Action*) is a non-profit association, founded in November 2006 by the Politecnico of Torino and the Higher Institute on Innovation Territorial Systems (Si.T.I.).
In this association the Politecnico of Torino operates through the Land, Environment and Geo-Engineering Department (DITAG), whose main aim is developing research and education in environment and territory sectors, in order to develop new technology instruments devoted to the protection and sustainable use of natural resources.
ITHACA focuses mainly on the development of new technologies in Information Technology (IT) and Geomatics fields. The different skills characterizing ITHACA are related to the acquisition, management and elaboration of geographic and cartographic data. The operational team has, in fact, experience in the various areas pertaining Geomatics, such as Remote Sensing, Photogrammetry, Geographic Information Systems, digital mapping and GPS.

2.2.2 ITHACA and WFP

By virtue of a specific agreement with the World Food Programme (WFP), ITHACA works closely with the United Nations system by providing a concrete and operational cooperation in research and development fields, in order to encourage the adoption of new processes and technologies in the humanitarian aid sector.
Particularly, working closely with the WFP's *Emergency Preparedness and Response branch* (ODAP), ITHACA is delivering technological and methodological support and services to WFP, and potentially to other UN Agencies, in order to:

- develop tools to improve the collection of emergency preparedness and response data;
- develop technologies and systems for Early Warning and Early Impact analyses for emergencies arising from natural hazards;
- support Information Management in the realm of emergency preparedness and response;
- develop procedures and services to provide decision-makers in humanitarian aid with real-time information and data generated by the Early Warning and/or the Early Impact developed systems;
- support the WFP SDI implementation effort;
- support the use of new methodologies by developing suitable tools;
- enable co-operation with other research centres and private entities.

Afterwards, in order to constitute a service support team for the WFP, ITHACA:
- maintains disaster databases;
- supports the maintenance of web based tools and systems;
- provides remote sensing analysis;
- provides GIS support pre- and post- natural disaster impact;
- manages and maintains global spatial data, making them accessible.

2.2.3 ITHACA's activities

In its first year of life, ITHACA carried out many activities which fall within the humanitarian sector. The first developed projects deals mainly with the thematic maps production and related products and services supply, useful to correctly plan and manage natural disaster mitigation operations. The main application fields for these products are the Early Impact and Early Warning. Early Impact activities are environmental and social analysis and evaluation activities devoted to the identification of effects caused by natural and manmade disaster events and to the correct planning and support of necessary relief operations. On the other hand, the main aim of Early Warning activities and analyses is to foresee catastrophic events and their effects on population and territory.
These activities are here briefly presented; they will be described in detail in the Needs Assessment phase as well, proposed in Chapter 4.

Early Impact activities

In this frame, the main goal of ITHACA activities is to supply WFP's users with cartographic products showing the effects of a catastrophic event, in a very short time after it happened. As a matter of fact, Early Impact analyses are based on well-established procedures which allow to obtain final results and elaborations within 24/48 hours after the event. Usually, ITHACA's Early Impact activities are carried out according to WFP's requests.
At the moment, procedures related to flood events were defined and tested. The main aim of these procedures is to rapidly produce georeferenced information on the Early Impact of flood events, especially data on affected areas and population. Infrastructure damages are also considered, with particular emphasis to road conditions. This information would avoid time-consuming procedures preferring rapid mapping methodologies, crucial in emergency situations. The information provided would be extracted from the data

acquired during and after the event, usually radar and optically satellite imagery. Geographic data are used both as comparison data and as cartographic basis. Remote sensing techniques and GIS spatial analyses are performed in order to obtain the final products.

The main outputs consist of small scale raster thematic maps, covering the area(s) of interest, which are delivered to the UN regional/country offices at the first state of alert, and just after the event. Data about identified flooded areas are also provided in a vector format as well as in KML files (suitable for Google Earth). During the first analyses of flood events, experiences in data exchange and sharing were also conducted mainly devoted to supply the final information and products to WFP's users, but also devoted to share intermediate outputs with other entities involved with humanitarian response.

Up to today ITHACA has brought into action for about 15 flood events and about 50 maps have been produced. Figure 2.1 shows the geographical distribution of catastrophic events analysed by ITHACA (Table 2.1).

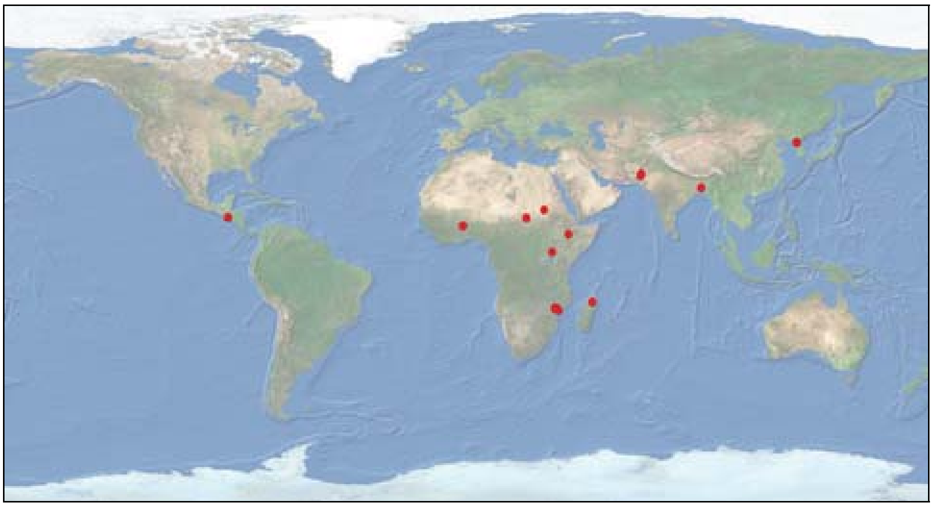


Figure 2.1 – Geographical distribution of ITHACA Early Impact activations

Table 2.1 – Flood events analysed by ITHACA

Event	*Number of produced maps*	*Date*	*Type of analysis*
Santa Ana Volcano - ELSALVADOR	2	13/03/2007	Potentially affected population
Mutarara and Caia areas - MOZAMBIQUE	5	February - March 2007	Flooded areas (Radarsat data - Charter), Potentially affected population, Analysis of damage to settlements, roads and other infrastructures (Formosat data - Charter)
Mopeia, Chinde and Marromeu areas - MOZAMBIQUE	1	28/02/2007	Flooded areas (Radarsat data - Charter), Potentially affected population
Mutarara, Caia and Mopeia areas - MOZAMBIQUE	3	February - March 2007	Flooded areas monitoring (MODIS data)
MOZAMBIQUE	6	February - March 2007	Rainfall monitoring (Precipitation forecast by NOAA - CPC Climate Prediction center), February and March flood affected area analysis: persistence of water (MODIS data)
Cyclone Indlala and Favio - MADAGASCAR (NE area)	2	March 2007	Flooded areas (MODIS data), Potentially affected population

Balochistan Province - PAKISTAN	5	2-5 July 2007	Definition of Charter priority areas, flooded areas (MODIS data and Radarsat data - Charter), Potentially affected population
Larkana and Dadu districts - PAKISTAN	6	5-10 July 2007	Flooded areas (Radarsat, ALOS, DMC data - Charter), Potentially affected population
Flood alert - ETHIOPIA	3	5-7 July 2007	Preliminary analysis of potentially floodable areas
Flood alert - SUDAN (Darfur area)	1	17/07/2007	Preliminary analysis of potentially floodable areas
SUDAN	0	July - August 2007	Flood monitoring (MODIS data supply to WFP for visualization in EPWEB - GDACS points)
BANGLADESH - NEPAL - INDIA	4	07/08/2007	Flooded areas (MODIS data), Potentially affected population
GHANA (North) - TOGO (North) - BURKINA FASO (South)	3	Sept - Oct 2007	Flood monitoring: Flooded areas (MODIS data), Potentially affected population
Soroti area - UGANDA	6	Sept - Oct 2007	Flood monitoring: Flooded areas (MODIS data), Potentially affected population
Cyclone Wipha - NORTH KOREA	2	20/09/2007	Preliminary analysis of potentially floodable areas
Cyclone Sidr - BANGLADESH	2	12-19/11/2007	MODIS Change Detection Analysis
MOZAMBIQUE	6	January - February 2008	Flooded areas (MODIS data), Potentially affected population
MOZAMBIQUE	3	January - February 2008	Simulated floodable areas, Precipitation forecast and considerations (Cabora Bassa Dam)
ZAMBIA	2	January - February 2008	Flooded areas (MODIS data), Simulated floodable areas

Figure 2.2 shows an example of produced thematic map.

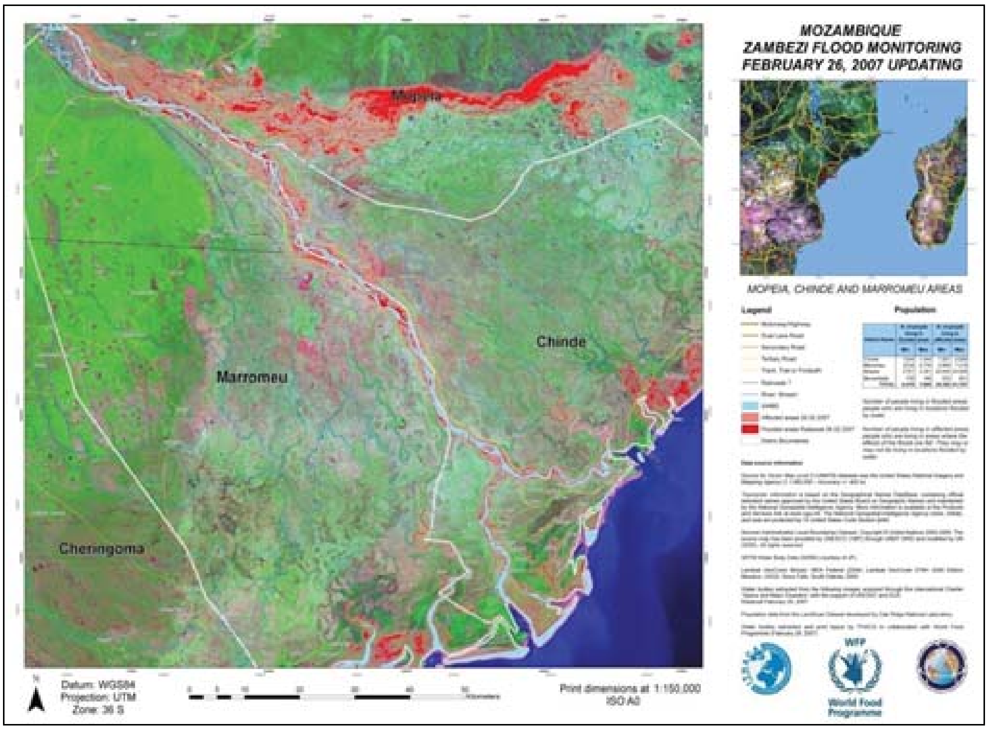


Figure 2.2 – Mozambique Zambezi flood monitoring, February 2007

Flooded areas delimitation activities require valuable reference data on water bodies and also are helped by the knowledge of historical areas affected by floods. For these purposes, a systematic inventory of past records that offer information about the extent of areas affected by major historical floods from the last 10 years is in the building stage. Such an inventory will also provide information on the areas that may be stricken again, highlighting dynamics and magnitude of the events.

Snow cover monitoring

The aim of this project is to provide daily updated information on snow covered areas to support WFP logistic activities. The information provided enable to perform road accessibility analysis, which allows WFP Logistics staff to select the optimal route when making food aid deliveries in areas where access is potentially adversely affected by the presence of snow. The analysis is achieved through the implementation of an automated process based on cost-free MODIS (*Moderate Resolution Imaging Spectroradiometer*) satellite data.

The expected outputs of this procedure consists of a daily updated ITHACA Web GIS application (open source) for the monitoring of particular areas of interest, and an on-request service for other areas (the expected map scale is 1:1000000, global coverage). The Web GIS interface displays a synthesis of MODIS Snow Cover data and also allows spatial analysis in order to know road conditions.

In addition, the on-request service can provide snow vector data almost in real-time. The files refer to snow cover information generally updated 24 hours before the request. A metadata file containing reliability information of the data is provided too.

At the moment, snow coverage is monitored in Nepal, Afghanistan and Central Asia (Himalayan region), in accordance with a WFP request. The snow cover data obtained for these areas is stored in a proper archive in ITHACA in order to allow future searches.

Using the same approach, it will be possible monitoring other useful environmental parameters. Research activities are at present ongoing in order to define the effectiveness of the use of satellite extracted vegetation index (NDVI, *Normalized Difference Vegetation Index*) historical series in drought early warning and prevention systems.

Floodplain modelling

The aim of the project is the development of a method to predict flood effects. The expected results are the delineation of floodable areas for different rain rates and the delineation of flood extent related to precipitation forecast. These results may be global provided that the input data for the model are available. They can be supply to the users in raster or vector format.

The outcomes of the model may also allow to identify the available access to the area and the potentially affected population. The development of this kind of flood forecast system can provide enough time for the appropriate actions, being a valid tool for emergency preparedness.

Currently, the research team is working on defining a methodology based on the analysis of historical precipitation time series data and the definition of various scenarios extracted from satellite images acquired during historical events.

The final goal of this research is the development of an automated procedure for the whole workflow including input data collection and pre-processing.

Web application for geo-data sharing

The main aim of the project is to distribute and share georeferenced information for both Early Warning and Early Impact activities, by means of suitable Web GIS applications, based on commercial or Open Source (OS) platform, highly customized and implemented using different software and hardware architectures, granting data reliability and integrity.

The project is aimed to test different frameworks in order to offer the most suitable solution to a specific data dissemination need.

Web applications may be based on several types of software architectures and servers, both Windows and Unix. The software architecture can be made of commercial or OS components. Developed applications can be accessed by means of a simple web browser without any plug-in. Published data may be stored on the same server the applications run on or may be reached on the net. Developed web services, besides map viewer functions, can also provide some analysis tool for data querying, editing and downloading.

At the moment, the research team is working on test applications. The main components under testing are OS tools such as Geoserver, Mapserver and Open Layers. A working solution is available on-line for the diffusion of Snow Cover project data outputs. New features under development and testing are the visualization of GeoRSS (georeferenced feeds) from different sources in a single disaster map and a buffer analysis around a GeoRSS for affected population assessment.

Survey airborne device development

ITHACA is also working on a project aimed at developing and constructing remotely controlled mini aircrafts (UAVs or "drones") for the acquisition of digital data necessary to plan emergency and relief food aid interventions. Each drone is equipped with photographic and multispectral sensors suitable for digital photogrammetric shootings. The aerial platform is capable of autonomous flight except for landing.

The UAVs are easily transportable and user-friendly. They produce images which will allow rapid map production and updating, and thematic maps generation, highlighting damages to buildings, infrastructures and flooded areas. They may be useful, for instance, for analysis of events when no suitable satellite data are available. In addition a micro-camera can acquire sequences documenting the investigated area that can be downlinked in real-time.

Currently, the research team is working on:

- the assessment of the autopilot software to allow autonomous photogrammetric flights;
- testing other types of sensors (IR, radar, etc) whose weight and dimensions are sufficiently miniaturized to be compatible with the characteristics of the UAV;
- developing a direct photogrammetric approach able to perform aerial triangulation and orthoprojection as soon as the data are downloaded.

The Global Geodatabase / Spatial Data Infrastructure for UN WFP project

As from the first experiences conducted, the following general needs were pointed out:

- availability of valuable geographical base data with global coverage useful to referencing and map production;
- definition of valuable and easily searchable archives for different data (externally or internally produced data, both in raster and in vector format);

- integration of many types of data obtained from different sources with many different formats as well as their organization and structure;
- availability of tools and instruments allowing the easy and rapid sharing of data and produced information among ITHACA research groups and between ITHACA and external agencies involved in the humanitarian sector (e.g. WFP units or UNOSAT);
- availability of tools and instruments allowing all the users to exploit the same data in a common repository and/or services in real time.

In order to meet these needs, the design and the implementation of a database containing all the geographic data globally consistent, used by the different research groups within ITHACA, has been considered as a very priority objective. In this way, the ITHACA *Global Geodatabase project* was born.
The main purpose of the geodatabase was to make effective the spatial data storing, querying and processing operations, constituting a suitable support tool for all analysis and visualization operations needed by ITHACA activities.
Subsequently, a natural growth of this project has been produced by the request submitted by the WFP to support their effort in developing the WFP *Spatial Data Infrastructure*.
Because the development of consistent reusable themes of base geographic content or *Core Data* is commonly recognized as a common ingredient in the construction of a SDIs, the *Global Geodatabase project* converged in the major *Spatial Data Infrastructure for UN WFP project* (also the *WFP SDI project*).
The different GIS users, in WFP but also in ITHACA, employ a limited number of common geospatial data. These common data provided key for the integration. Therefore, the concept of "*Core Data*" is here used as *a set of Geographic Information that is necessary for optimal use of different applications and activities*. In the context of SDI implementation, the definition of *Core Data* is one instrumentality to help improving interoperability, reducing expenses resulting from the inevitable duplications.
The main aim of the new *WFP SDI project* was the development and implementation of a global database and of related rules and procedures in order to store, manage and exchange geospatial data for WFP and other actors in the humanitarian sector, ITHACA included. The data model had to be shaped according to specific needs and demands of the user groups within the different interested organizations. Moreover, the developed database had to allow integrating and updating data coming from multiple sources, providing each user with the latest datasets, thus avoiding duplications and mistakes.
In this thesis the different stages which have been necessary in order to accomplish this first phase of the project, mainly the development of the data model for the global geographic database, are described. The general adopted workflow for the geospatial database design is presented in the next chapter. Moreover, a description of the features of the first version of a prototype (an ESRI ArcSDE Enterprise geodatabase based on Oracle 10g as DBMS) which implements the geodatabase model is also presented in Appendix.
At the moment, a complete test version of the geodatabase, including all information for Africa continent, is accessible through web services hosted by ITHACA server.

The project of the data model and the implementation of the geodatabase constitute only the first step of the more complex *WFP SDI project*.
The SDI, currently in development, has to be an efficient tool for storing, querying and manipulating large amounts of global geographic information and spatial data for analysis and visualization purposes. The major

expected output of this project is a framework of spatial data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way granting data reliability and integrity, standardization and metadata integration.

As a matter of fact, the concept of SDI concerns not only spatial data structure, but also data management (the technology to allow multiple users to query, retrieve, manipulate, edit and save data in a common repository) and supply of services (what the users can get from the data repository). Therefore, the complete system will contain catalogues, search engines, data processing tools, map generating tools, user identification and security measures.

Finally, the services which will be provided include:

- direct data access to a common data repository, including grants and rules for data management (the data will be accessed through user-friendly web-based platforms; the data will be mainly available to GIS practitioners in-house, that is VAM, ODAP and UNJLC, and partners and for ITHACA activities);
- web applications and portals designated mainly for data consultation and analysis;
- on-line archive data catalogues (mainly for past events data);
- simultaneous access to the geodatabase for support to disaster mitigation operations.

The presentation of the first SDI solution adopted, including the definition of the system architecture and configurations is out of the goal of this thesis; nevertheless they are briefly presented in the document in Appendix, produced by the ITHACA research group that manage the *WFP SDI project*.

3

DATABASE DEFINITION AND DESIGN

The adopted methodology for the global geographic database design is here presented. As aforementioned, this database is essential for applications and activities developed by ITHACA, and it also constitutes the first and basic step in the WFP SDI development.

The adopted design techniques do not wander off the techniques proposed in literature for general database design. Nevertheless, in some steps, it was necessary to consider the specific features which make geographic data different from the other data types.

Moreover, some basic concepts on database and geographic database theory are presented.

3.1 Basic concepts

3.1.1 Database and Database Management System

Each organization, for its own functioning, needs to dispose of information; this information represents one of the organization resources, therefore each organization houses a subsystem that deals with the management of this resource. This subsystem is called *information system* of the organization. The information system of an organization is a combination of human and material resources, and of organization procedures for the collection, storage, processing and exchange of information necessary for the operational activities.

The information of an organization, once reduced to data through a process of interpretation, quantification and formalization, can be automatically processed by computers. In this case we talk about a *data processing system*. A data processing system is a set of data processing tools used for the automatic elaboration of the information of an organization in order to facilitate its information system functions.

The data processing system, in order to provide the expected services to the users, includes three main components:

- basic *software* and *hardware*;
- an *information base*;
- the *application programs*, that provide services to the users by running a certain set of operations on the information base.

In this section the attention will be focused on the information component of the data processing system.

A *database* is a collection of permanent data that can be managed by a computer. These data are:

- the *metadata*, which is to say the *schema* of the database *(database schema),* is a collection of definitions that describe the data structure, the restrictions on the data permissible values (integrity constraints), the existing relations between the sets, and sometimes also some possible operations on the data. The scheme should be defined before creating the data and it is independent from the other applications using the database;
- the *data*, representations of certain facts complying with the schema definitions, with the following characteristics:
  - they are organized into homogeneous groups, among which the relations are defined. The data structure and the relations are described in the schema with appropriate mechanisms of abstraction that depend on the *data model* applied;
  - they are permanent, that is, once created, they continue to exist until they are explicitly removed; therefore their life doesn't depend on the duration of the applications that are using them;
  - they are protected both from the access of unauthorized users, and from the corruption due to hardware and software malfunctioning;
  - they can be used simultaneously by different users[1].

---

[1] The term "user" here indicates both a person who accesses to the data interactively from a terminal using an appropriate language, and has also the meaning of an application program that contains instructions for accessing data.

With the previous definition, the following facts are highlighted:

- the data are structured, that is they have a default format, and the number of present data types is relatively small compared to the number of samples for each one;
- the data are grouped into homogeneous sets, in relation to each other; it is expected to have operators to extract elements from a set and to know the ones that, in other sets, are in connection with them.

According to a traditional vision, the data of a database and its related schema correspond respectively to a set of variables (which denote changeable sets of permanent values) that many applications can read and modify, and to the definition of the type of such variables.
Finally, a *data model* is a set of abstraction mechanisms dedicated to define a database, with associated a predefined set of operators and integrity constraints. The data models adopted by the commercial systems and on which we will not go into depth, leaving as reference for more information the indicated bibliography, are: the *hierarchical*, the *reticular*, the *relational*, the *object - relational* and the *object - model*. The relational model and the objects-relational nowadays are the ones more widespread in the commercial system.

The databases characteristics, above defined, are guaranteed by a *DataBase Management System*, (DBMS), that has data control and makes them accessible to the authorized users. Technically, a DBMS is a centralized or distributed system that allows: (a) to define databases schemas, (b) to choose the data structures for the data storage and access, (c) to store, retrieve and modify data interactively or by programs, for the authorized users and respecting the constraints defined in the schema.
Figure 3.1 shows a simplified version of a DBMS structure.

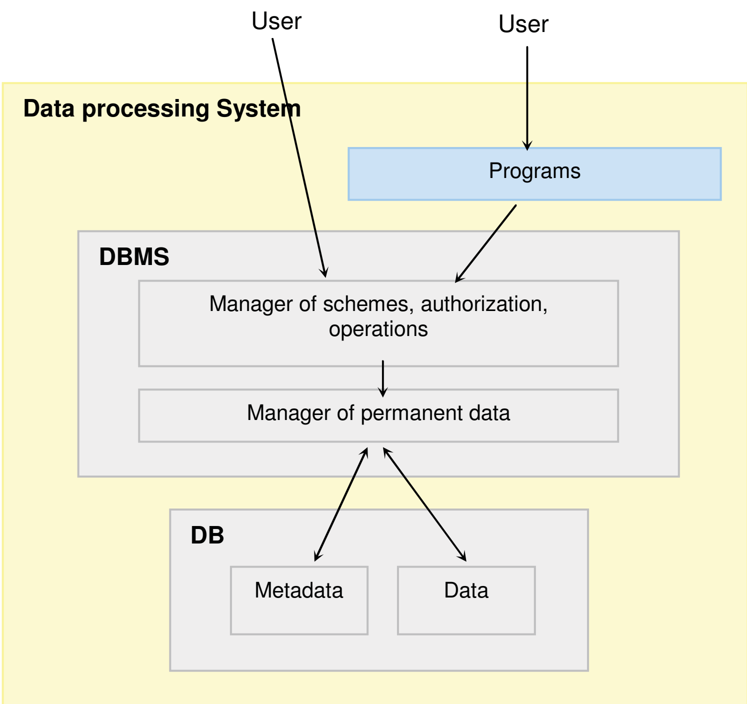


Figure 3.1 – Database and DBMS

The DBMS is the system, hardware and software, which allows the database definition and use. In other words, it is the abstract machine whose language is *the language for databases*. Therefore, it is similar to the compiler, or interpreter, of a certain language, with the addition of all the active modules during the program execution (*run time system*). This type of system represents the most appropriate technology in developing applications that use, in a concurrent way, persistent data, in contexts where it is necessary to access to the same information from multiple applications and several areas of the same organization.
The databases technology, in addition to the advantages of data integration and flexibility, also offers other benefits, in particular it establishes standards concerning the information structure and nomenclature, and it significantly reduces the development time of applications. The issue of establishing standards is strongly felt within large organizations, where the variety of needs requires a terminology and a common way to define the data in order to facilitate communication and cooperation among users.
To summarize, in a databases system it is guaranteed:

- data sharing and integration: data are organized to be used by different applications;
- data consistency: the updates are seen immediately by all applications, due to the data sharing;
- data integrity;
- physic independence: you can access to the data regardless of their physical organization, which is chosen among the organizations provided by the DBMS;
- data standardization: programs use the names defined in the scheme;
- data easy to be used: data are accessible through programs and interactively;
- data security and reliability: guaranteed by the DBMS.

3.1.2 Geographic database and GIS

The main aim of the work described in this document has been the design of a global geodatabase. The term geodatabase is used here to indicate a structured collection of geographical data. In a simple way, *a geodatabase is an instance of a database, one that contains a structure for representing geographic data.*
A geodatabase is the foundation for a Geographic Information System (GIS). GIS are automatic systems used in order to store, analyse, manage and handle geographic data. The key idea to grasp about GIS software is that it is, in fact, a software application which accesses to geographic database and carries out database applications.

GIS concepts

There are many definitions of what GIS is and most of them have some common parts. Due to their heritage in cartography or map-making, Geographic Information Systems may be viewed as the result of the union of Computer Assisted Cartography (CAC) and database technology.
Here are some definitions of GIS that are in compliance with the purposes of this discussion:

GIS is a special-purpose digital database in which a common spatial coordinate system is the primary means of reference.

A geographic information system (GIS) is a computer-based tool for mapping and analyzing things that exist and events that happen on earth. GIS technology integrates common database operations such as query *and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of applications*.

Three observations should be done about these definitions:

- GIS are related to other database applications, but with an important difference. All information in a GIS is linked to a spatial reference. Other databases may contain location information (such as street addresses), but a GIS database uses georeferences as the primary means of storing and accessing information.
- GIS integrates technology. Whereas other technologies might be used only to analyze aerial photographs and satellite images, to create statistical models, or to draft maps, these capabilities are all offered together within a comprehensive GIS.
- GIS, with its array of functions, should be viewed as a process rather than as merely software or hardware. GIS are for making decisions. The way in which data is entered, stored, and analyzed within a GIS must mirror the way information will be used for a specific research or decision-making task.

One of the main features of a GIS is its ability to manipulate *geographic data* or *geometric* and *descriptive data* in an integrated way, providing a consistent method for the analysis and the consultation. *Geographical entity* is any phenomenon of the real world that possesses attributes associated with its location on the hearth surface at a given time. The GIS allow to access to descriptive information of a geographical entity from its geographic location and vice versa.
More specifically, *geographical data* within a GIS are usually composed of three major components:

- *location*. This is the most important component, the one that justifies the essence of geographical data. Every object or phenomenon has a position, described in a concrete reference system;
- *attributes, type, appearance* of the object or phenomenon considered. For example, attributes of a forest could be its density, the species of the trees it is composed of and their average height;
- *spatial o topological relationships* with others entities. For example, a building may belong to a city.

What differentiates a GIS from other information systems is the manipulation of spatially referenced data: this implies that a GIS must deal with spatial relationships among the objects and phenomena it handles. This spatial analysis is unusual in other classical information systems.
In fact, there are several systems that manipulate spatial data, such as CAD (Computer Aided Design) systems. The GIS differ from these systems into two aspects:

- for their ability to represent the spatial or topological relations between the geographical entities. In fact GIS provide tools to store spatial relations that exist between the geographical entities.
- because they permit the realization of complex spatial analysis on geographic data.

Moreover, a system of computerized graphic doesn't take into consideration non–graphic attributes of the described objects, while in GIS these are essential for the implementation of spatial analysis. A graphics system deals exclusively with the visualization of the spatial information contained in the data, condition necessary but not sufficient for a GIS.

The basic software modules of a GIS are at least five and perform the following functions (Figure 3.2):

- data introduction and validation (input);
- data management (management of the database);
- data diffusion and presentation of results (output);
- data processing;
- interaction with the user.

The data are introduced into the system, in order to be processed. The system must provide a way to store, organize, and give access to the information. This is known as data management. The relevant data for solving the problem are extracted from the system and analyzed (data transformation). The results are communicated to the user, usually by means of maps and graphics.

The data management takes place within a DBMS associated with a GIS for the database management. This DBMS must ensure quick data access and easiness in the execution of update and correction operations. In fact a GIS is a tool that permits the integration, in a single database, of geographic information coming from different sources. Moreover, the correct data management needs a preventive definition of their storage structure, so that they can be easily processed in the variety of applications.

The information base is the core around which all the GIS software modules of application and management turn. In general, it is necessary that the design of an information base takes into account all the users' needs, so that each of them can dispose of all the information required to satisfy their needs. The condition that all data are available inside the database is not enough to guarantee the system functioning: in fact, it is required to proceed to their organization, inside the memory, in order to connect different types of information, in particular the *geographical entities*, such as points, lines and areas, with all the *attributes* related to them.

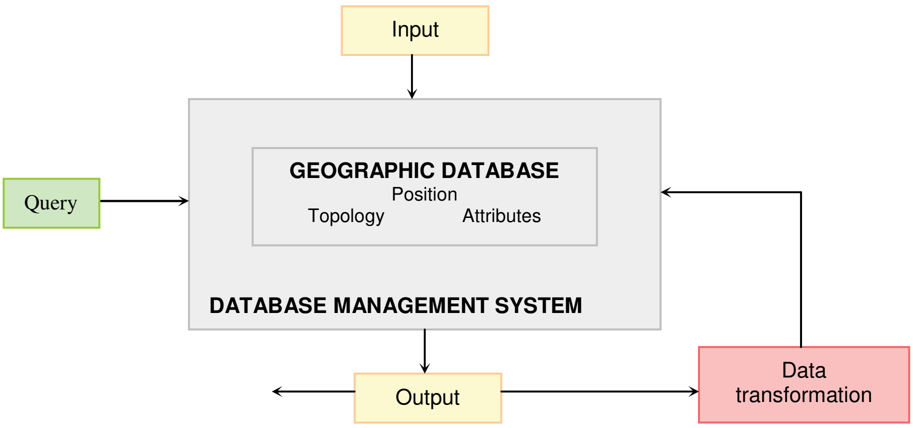


Figure 3.2 - Software modules in a GIS and their interactions.

Finally, data transformation refers to several operations useful to retrieve, summarise, selectively display and analyse both the alphanumeric and graphical data which are contained in a GIS. These data transformation capabilities of the GIS instruments make them useful tools for decision making support and environmental analyses (Figure 3.3).

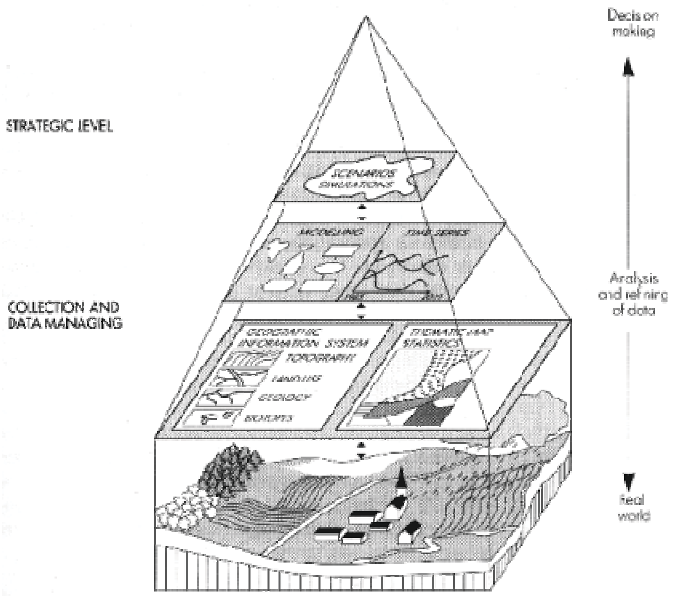


Figure 3.3 - By use of geographic information systems, a "simplified" world can be brought and analysed into the computer in order to support decision making activities. (source:[UNESCO, "Training Module on the Applications of Geographic Information Systems (GIS)"])

They have been here briefly described some introductory concepts concerning GIS because they are a powerful tool for territory analysis and study, which is the basis of the work carried out by ITHACA. The potentialities of these tools, the functions of analysis that they make available, and their practical use in these activities context will be presented in Chapter 4. Here, we wants only to highlight how the design of a geographic database can not set aside the certainty that the data it contains will be taken mainly through the use of GIS tools. That is why in the following paragraph the spatial data models used in the current GIS will be briefly presented, the knowledge of which is necessary for a proper design of a spatial database.

3.1.3 GIS spatial data models

A *spatial data* is any data describing phenomena to which it is associated a spatial dimension. The *geographic* or *georeferenced data* are spatial data in which the spatial dimension is associated with their localization on the terrestrial surface, in a specific moment.
The term *spatial* has a broad meaning, while the term *geographic* refers specifically to the Earth's surface. The term geographic can be applied in a subset of the situations where the term spatial can be applied.
The geographic data have three key features: *spatial, non-spatial and temporal characteristics.* The spatial characteristics provide information on the location and geometry of the phenomenon; the non-spatial characteristics describe the phenomenon; the temporal characteristics indicate the period of validity of the geographic data.
The geographic data have also *geometric* and *topological characteristics*. The geometric properties are metric properties. Starting from primitive geometric features, such as points, lines and polygons, which represent the geometry of the entities, the metric relations are established. These relations reflect the metric of the spatial characteristics in reference to a coordinates system. The topologic properties (non-metric),

such as connectivity, orientation, adjacency and containment, are based on the relative positions of the objects in the space.
The geographic data in a database are the result of the reality observation and its subsequent modelling and, therefore, they constitute the elements needed to represent that part of the real world that involves the GIS users. In order to properly model the reality we have to define:

- the geographical entities of interest. A *geographical entity* is any entity identifiable in the real world, which has spatial characteristics and spatial relations with other geographical entities.
- their spatial representations.

Below, there are presented some fundamental theories for modelling, inside a GIS, objects belonging to the real world.

Geographic data representation

There are two opposing types of models to gather, store and manipulate geographic information: *field-based* and *object-based* models. These two types of models are used during the definition of the high level model of the database, namely during conceptual or logical design phases, as presented in next chapters. Although the field and object models are concepts independent of implementation, in a GIS the field model is more naturally implemented by raster data structures and the object model more suitable to be implemented using a vector format.

Field versus object model

Field-based models treat the information as continuous geographic phenomena distributed over the space. Patterns of rainfall and temperature are examples of geographic phenomena that can be represented appropriately by the field model. In this approach, the information is captured and stored as a collection of fields. Each field may be formalized as a mathematical function from a set of locations to a set of attribute values that characterize the geographic phenomena. The set of locations is created, for example, by dividing a given region into a finite grid of squares where each individual square is a location approximated by a point. Then, the phenomenon under consideration is sampled at the determined locations.
In particular, the spatial fields are constructed as functions mapping each location to the sampled attribute value. So, a field is a total function $f : R \rightarrow V$, where R is the finite set of sampled locations (expressed as coordinates) that cover a region and V is the set of values that each location may assume. Determining the value of the field for a location that has not been sampled requires an approximation. This approximation depends on the method by which the region R is partitioned in different cells or by which the phenomena are sampled in the region R. This method also defines a more refined classification to describe the type of the field. For each type, the discrete locations may have different names, shapes and characteristics:

- Grid of Cells (figure 3.4a): the region R is partitioned in rectangular cells. The value of the field for each location within one cell is constant. The value of the field may change abruptly at the edge of the cells.
- Adjacent Polygons (figure 3.4b): the region R is partitioned into irregularly shaped polygons. The value of the field within a single polygon is defined as a constant and changes abruptly at polygon edges.

- Triangular Irregular Networks (TINs) (figure 3.4c): the region *R* is partitioned into triangles of different sizes. The value of the field is defined by measurements only at triangle nodes. For any other location on a triangle face, the value can be calculated directly from the values at the nodes. The value of the field is defined by a linear function. There is no abrupt change in value at the edges of the triangles. However, there is an abrupt change in slope at the edges, because each planar triangle face has a different linear function.
- Grid of Points (figure 3.4d): the phenomenon is measured at the intersections of a rectangular grid over region *R*. Unlike the previous field types, the interpolation function is not clearly defined. The map designer is responsible for defining the interpolation function for any given location that does not belong to the sampled points.
- Irregular Points (figure 3.4e): the phenomenon is measured at some selected locations of region *R*. The interpolation function should also be defined by the map designer. The locations may be selected according to their relevance for the phenomenon (e.g., representative locations for measuring rainfall) or they may be selected not accounting the phenomenon under consideration but other factor.
- Contour Lines (figure 3.4f): this type of field differs from the others, in that the value of the phenomenon is what determines the locations represented by the field. Adjacent locations which have the same value are connected by a line called a contour line. All contour lines are continuous and they do not intersect each other. Each contour line explicitly identifies all adjacent locations that exhibit its constant and desired value. The value of the phenomenon is defined only along the contour lines. The locations of the contour lines are determined by the phenomenon (e.g., rainfall) and the selected value for drawing the contour line (e.g., 5mm). The interpolation function for locations lying between two contour lines is not necessarily linear. However, its range is constrained by the values of the two bounding contour lines.

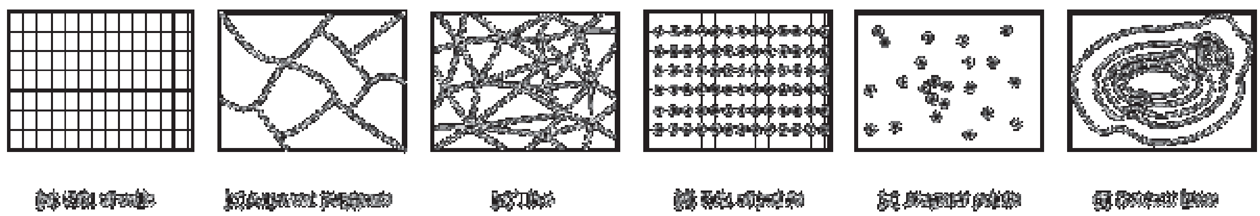


Figure 3.4 - The different types of fields.

Object-based models treat the geographic information as collections of discrete, identifiable entities that populate the space. Each of these entities is georeferenced and has attributes describing the geographic phenomenon. For example, cities can be represented in the object model as discrete entities with associated location, average temperature, average rainfall, population, etc. Instead of identifying the attribute value of the phenomena for each location as in the field model, the object model clumps together all the attribute values for one specific object. It also maps this discrete and identifiable object to its location.

Object-models have two subcategories depending on whether or not the relationship between objects is captured and stored in the model:

- *Isolated object model*: in this model, objects are described using simple geometric primitives such as point, line and area. Each object is described as a geometric whole, since it occurs in isolation from the others. The object contains also any non-geographic attributes.
- *Connected object model*: in this model objects are described in relation to each other. For example, a point may be the end point of a line or a point of intersection between two lines. A line may divide an area into two polygons. These relationships are not recognized at the isolated model. The connected model is, however, mostly concerned with the description of these relationships which are called *topological*.

  As used in cartography, topology is concerned with "the relative location of geographic phenomena independent of their exact position". The topological relationships are relations that show the topological link between two spatial entities (for example, the Tevere crosses Rome). To determine the minimal set of topological relationships that may occur between two spatial entities, Egenhofer and Franzosa [M. J. Egenhofer, R.D. Franzosa, 1991] have studied how the edges (or contours), the inside and outside of the geometric figures can intersect. If we assume that the possible values for the intersections are empty and not empty, we can generate a matrix (known as *9-intersection matrix*) able to encode every possible combination of the borders, the inside and outside intersections, of two spatial entities in $\Re^2$. Not all the possible configurations of this matrix have a topological meaning: for example, a configuration where the contour, inside and outside of both considered spatial entities have no intersection is not possible. However, it can be identified a minimal set of seven topological relations between two objects in $\Re^2$. The seven relations are: disjunction, adjacency, coincidence, inclusion, covering, overlapping with disjoint edges, overlapping with intersection of edges (Figure 3.5).

  As we will see, at the geometry or implementation level, topologies are about simple relationships such as coincidence, covering and crossing between the geometries (point, line, polygon) that make up spatial objects. To capture the topological relationships, features need to be stored using node, edge and polygon primitives. Nodes represent the beginning, the ending and any intersection of linear objects which are, in the connected model, represented by edges. Nodes bound edges and edges bound polygons. From these simple relationships, more complex ones can be built. For example, if two polygons share a common edge, then they are said to be adjacent.

(a) disjunction **(b) adjacency** **(c) coincidence** **(d) inclusion** **(e) covering** **(f) overlapping with disjoint boundaries** **(g) overlapping with boundaries intersection**

Figure 3.5 - Topological relationships.

<u>*Raster versus vector representation*</u>

The two basic forms for representing the geographic data inside a GIS are the *vector* form and the *matrix* form or *raster* (Figure 3.6).

In *raster data representation*, space is regularly subdivided into cells. Raster data are two-dimensional arrays of cells (or pixels). Each cell represents a surface of the real world. This surface area defines the resolution of the model. Usually the cells are rectangular and the height and the width of each cell are fixed and the same.

Each cell has a value. This value can represent many qualities of the surface they define, including reflectance, colour, precipitation and elevation.

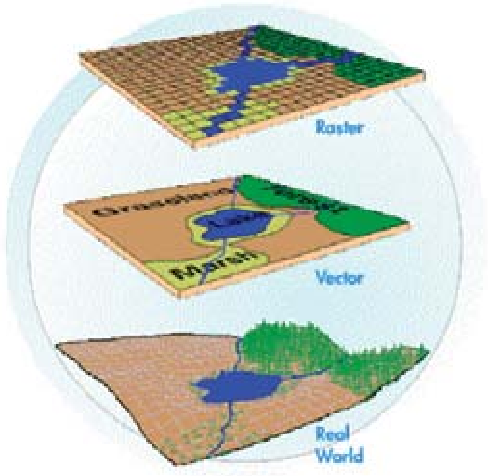


Figure 3.6 - Raster and vector representations

Raster data have an integer coordinate space. You can determine the coordinate of a cell by counting columns from the left and rows from the top.

In this case, the topological relations in the space are implicitly determined by the closeness of the cells, while the geographical coordinates (longitude, latitude), or planar (x, y), can be indirectly obtained by starting from the position of the cell inside the matrix (row, column).

Common sources of raster data in GIS are satellite imagery, aerial imagery, scanner maps or pictures.

If the phenomena in the real world are perceived as distinguishable objects, then these objects should be described using the *vector representation*. The vector representation needs to make a distinction between the different classes of objects. Their location is also crucial. However, there is another important phenomenon that should be described about the objects: their geometry. Vector data representation is best applied to discrete objects with defined shapes and boundaries. The representation in vector format use points, lines and polygons to represent the geometry of the geographic entities. Each of the geometric primitives assumes that the location of the object is provided through a number of coordinates. The points are represented by a couple of coordinates, the lines by a sequence of points and the polygons by a sequence of lines where the coordinates of the initial points and the ones of the final point are the same.

The linear geographic entities, such as roads, politic-administrative divisions and road networks are naturally represented in vector format.

If the objects are described by class of the object, location, and geometry, then the map is said to follow the isolated object model. No relationship between the objects is required in the description. This description of the reality may be enough for some applications but may be too simplistic or may not talk about the necessary phenomena when considering another application.

In the connected model, objects need to be described as distinguishable entities as in the isolated model, but their relationships with other entities are also required. There are important geographic relationships between entities that may need to be considered in the modelling process in order to solve a specific problem:

- <u>topological whole-part relations</u> are a special type of geographic relations where each part has a non-empty intersection with the whole entity. These relations are further divided into: **–** *covering*: in a covering relation the geometry of the whole is covered by the union of the geometries of the parts.

The parts may overlap and may exceed the boundary of the whole. The whole and the parts must belong to the same geometric type. – *containment*: in a containment relation the geometry of the whole contains the geometry of all its parts. The parts may overlap but must not exceed the boundary of the whole. There may be regions of the whole that are not covered by any of the parts. – *partition*: in a partition relation the geometry of the parts forms a partition of the geometry of the whole. In this case the parts must not overlap and there is no region of the whole that is not covered by a part. The whole and the parts must belong to the same geometric type.

- topological network relations are concerned with two distinct classes of objects called nodes and edges. A network or graph is a collection of nodes and edges where objects of the node class are connected to each other through objects of the edge class; in other words, each edge is incident to two nodes. If the arcs are oriented, flux and direction information inside the network can be represented.
- other topological relations may describe how two or more entities may be related to each other geographically. For example, given two polygons, *X* and *Y* (closed and with no holes), the following relations (planar topology) are defined:
  - *X* is *disjoint* from *Y* if the intersection between their boundaries is empty and the intersection between their interiors is also empty.
  - *X is inside Y* (*Y contains X*) if *X* is a subset of *Y* and *X; Y* do not share a common portion of their boundaries.
  - *X equals Y* if their boundaries are the same.
  - *X meets Y* if *X* and *Y* touch externally in a common portion of their boundaries and their interiors do not intersect.
  - *X covers Y* (*Y* is *covered by X*) if *Y* is a subset of *X*.
  - *X overlaps Y* if the intersection between the interior of *X* and the interior of *Y* is non-empty.

  The definition of these topological relations depends on the dimensionality (0,1 or 2: nodes, arcs and polygons) of the objects involved in the relations. However, they may all be defined in terms of the intersection between the boundaries, interiors and exteriors of the objects, as aforementioned.

Note that all the relationships presented here can be calculated and built from simple relation rules and the location (coordinates) of the objects.
It is important to describe how the classes of objects in a map are related to each other because this is an important topic in data modelling, which is a fundamental step of the spatial database design process. There are many geographic notations and models (conceptual and logical) specialized to describe geographic objects, their geometries and locations, and the relationships among them. Most of the contributions in this area come from the general database theory. For instance, in the conceptual model field, some notations available are specializations of the Entity-Relationship model, others rely on some object-oriented modelling language. This topic will be faced with more details in Chapter 5.

The *field-model* and *object-model* are implemented into systems for the geographic information management, in *raster* and *vector* formats. In most cases, the fields are represented in matrix format, while the objects are represented in vector format. However, as already noted, in general, views based on fields and those based on objects can be mapped in either structure. The example of the contour lines well shows

this fact: despite being a model based on fields approach, the vector structure allows to represent them better than the matrix one.
The design activity of a geographic database requires an effort of modelling the reality, including that the appropriate representation form is chosen for each object of the modelled real world. The choice between the two representations, vector and raster, generally depends on the kind of data sources and on their real availability, on the degree of spatial resolution required, and on the requirements of algorithmic nature needed for the development of applications that use the data contained in the spatial database (for example, overlay operations are more efficient with data in a raster format). In the chapters that follow, you will see how we have kept into consideration these factors for modelling the entered data into the designed geodatabase.

3.2 Geodatabase design

3.2.1 Introduction

Designing a database means to define its structure, its characteristics and its content. The design of a database represents only one of the components of the development process of a complex information system, and it must be therefore framed in a broader context, *the life cycle* of the information systems.
As described in Figure 3.7 below, the life cycle of an information system generally includes the following activities:

- *collection and analysis of needs*. It consists in the identification and study of the properties and features that the information system will have to have. This phase requires an interaction with the system users and produces a complete description, but generally informal, of the involved data and of the operations done on them. Also, the software and hardware requirements of the information system are defined.
- *planning and design*. It is generally divided in data design and applications design. In the first one we will identify the structure and organization that the data should have, in the other one we will define the characteristics of the applications programs. The descriptions of data and of the applications produced in this phase are formal and refer to specific models.
- *implementation*. It consists in the realization of the information system according to the structure and characteristics defined in the design phase. The database is constructed and populated and the program codes are produced.
- *validation and testing*. Used to verify the correct functioning and the quality of the information system.
- *activity*. At this stage the information system becomes operational and performs the tasks for which it was originally designed.

It must be specified that the process is never strictly sequential since often during the execution of one of the mentioned activities, it is necessary to review decisions made in the previous activity. What you get is an operations "cycle". Moreover, sometimes it is added to the mentioned activities the *prototype* one, which consists in the rapid realization of a simplified version of the information system, with which to experiment its functionality.

The databases are in fact just one component of an information system that typically includes also application programs, interfaces with the user and other service programs. However, the central role that the data have in an information system widely justifies the particular interest generally dedicated to the databases design.
The first two activities of the presented process were developed during the geodatabase design work, which is the object of this thesis, according to the methodology presented in the next.

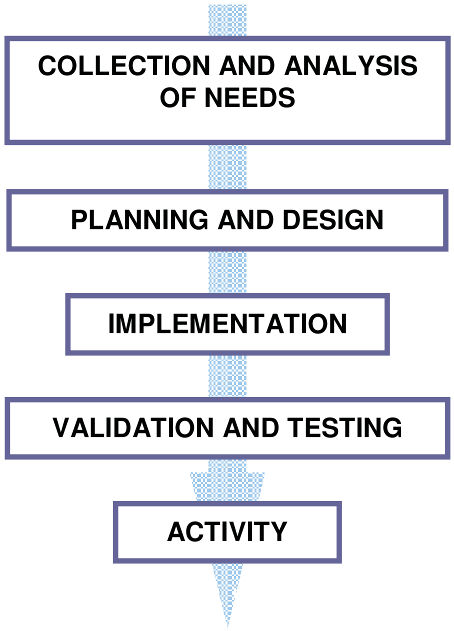


Figure 3.7 - Life cycle of an information system.

3.2.2 Design and modelling

When designing a database it is useful to imagine that the interested organization wants to collect information concerning the status of a portion of reality, *the universe of discourse*. In this situation, the database must be structured in such a way that it represents an *abstract* or *symbolic model* of this universe of the discourse.

An abstract model is the formal representation of ideas and knowledge related to a phenomenon.

This definition highlights three key aspects of an abstract model:
– a model is the *representation* of certain facts of a phenomenon;
– the representation is given with a *formal language*;
- the model is the result of a process of a phenomenon *interpretation*, driven by the *ideas* and *expertise* already possessed by the person who interprets.

The design of a database occurs in several steps, as described in the next section, each of which produces a more and more detailed project, in order to reach the realization of the system. It is useful to describe these phases by saying that each of them constructs a model of the universe of discourse by refining the model shown in the previous step. As you will see, in the first stages we will use more abstract formalisms, closer to how the reality is perceived by humans, while in the final stages we will use formalisms in which it is possible to specify a higher number of details, which are directly interpreted by a processor.

Basic modelling concepts

In this paragraph we propose some definitions related to the following aspects of modelling: *ontological, linguistic abstract and linguistic concrete*. The concepts here introduced are considered necessary prerequisites for understanding the next chapters.

Ontological aspect

The ontological aspect (study of what exists) is related to what we suppose existing in the universe of discourse and therefore that is to be modelled. About this point, several hypotheses are possible, and the one normally preferred when modelling a database is that the following facts are modelled: *the concrete knowledge, the abstract knowledge and the procedure knowledge.*

<u>The concrete knowledge</u>

The concrete knowledge concerns the specific phenomena that we want to represent. By adopting a simplified approach, we assume that reality consists of *entities*, which have a *type* and some *properties*. It is also assumed that these entities are classifiable within homogeneous entities *collections*, and that there are *associations' instances* between these entities.
Here are some definitions:

An entity is something concrete or abstract of which we want to represent some facts.

Examples of entities are concrete objects (a book, for example, or a library user), abstract objects (the bibliographic description of a book) and events (a loan). What it is necessary to know first about an entity is the fact that it exists, and then its *properties*.

The value of a property is a fact that describes a quality of an entity.

Examples of properties are the user name and address or the author, the title and publisher of a book. The difference between a property and an entity flows from the different interpretation of their role in the model: the property values are not facts interesting by themselves, but only as a characterization of other concepts interpreted as an entity. An entity does not match with the values taken by its properties, because:

- the property values of an entity change over time, although the entity remains the same;
- two entities may have the same properties with the same values still being two separate entities.

The concept of property is closely linked to the concept of *type*.

A type is an abstract description of something in common within a set of homogeneous entities (of the same kind), existing or potential.

For example, person is the type of Giovanna, Mario, etc.
A type is not a collection of existing entities, but describes the nature of all "possible" entities with this kind. A type can be seen as an infinite collection of possible entities. To a type, we associate the properties that interest the entities belonging to that type, and the *characteristics* of these properties. For example, the type user has the properties name and address meaning in this way that each user has a name and an address, but with a different value than all the others. With regard to the *characteristics* of the properties, each property has an associated *domain*, which is a set of all the possible values that this property can take.
Another important concept is the one of *collection*.

A collection is a set of interesting homogeneous entities in the universe to be modelled, variable through the time.

For example, the books of a library are the collection of books owned by the library. The set of the elements of a collection, at a specific time is called *extension* of the collection.
Therefore, while a type describes all the possible entities with certain qualities, a collection is a finite and variable set of entities of the same type, which are actually part of the universe to be modelled.
It has to be taken into consideration the fact that, while realizing a model, neither all the possible collections of entities nor all the entities that exist in the universe to be modelled are represented; in the model, only those entities and collections that are considered interesting in relation with the objectives for which the model is built are represented.
Two or more entities may be connected by an *istance of association*.

An association is a fact that correlates two or more entities, establishing a logical link between them.

An association may also correlate more than two entities and can also have properties. An association can be characterized not only by its domain and its properties characteristics, but also by the following *structural properties*: the *multiplicity* and the *totality*, which are defined, for simplicity, only for binary associations.
The *multiplicity* of an association between X and Y concerns the maximum number of Y elements which may be related with an X element and vice versa. It is said that the association is *univocal* from X to Y if each element of X can be put in relation with maximum one Y element. If there isn't such a bond, it is said that the association is *multi-value* from X to Y. In the same way it is defined the univocity bond from Y to X.
It should be noticed that the univocity bond from X to Y is independent from the univocity bond from Y to X, giving origin to four possible combinations of presence and absence of the two bonds. These combinations are expressed in a compact way as follows.
The *cardinality* of an association between X and Y describes at the same time the association multiplicity and of its inverse. The *cardinality* is *one to many* (1: N) if the association is multi-value from X to Y and univoque from Y to X. The cardinality is *many to one* (N: 1) if the association is univoque from X to Y and multi-value from Y to X. The cardinality is *many to many* (N: M) if the association is multi-value in both directions, and it is *one to one* (1:1) if the association is univoque in both directions.

For example, in a universe of the discourse populated by students, departments, courses and professors, the association *Attends (Students, Courses)*[2] has cardinality (M:N), the association *Teaches* (*Professors, Courses*) has cardinality (1:N) and the association *Manages* (*Professors, Departments*) has cardinality (1:1).

The *totality* of an association between two collections X and Y concerns the minimum number of Y elements that are associated with each element of X. If at least one entity of Y is associated with each entity X, the association is defined *total* on X, and vice versa replacing X with Y and Y with X. When the bond of totality doesn't exist, it is said that the association is *partial*. For example, the association *Manages* (*Professors, Departments*) is total on *Departments*, since each department has a director, but not on *Professors*.
In summary, the *structure of the concrete knowledge* concerns the knowledge of the following facts:

- existence of some collections, collections name, type of the collections elements (hence, name and characteristics of the properties of these elements);
- existence of certain associations, associations name, collections related from each association, structural properties of the associations.

<u>*The abstract knowledge*</u>

The *abstract knowledge* concerns the general facts that describe: (a) the structure of the concrete knowledge, (b) the restrictions on the possible values of the concrete knowledge and on the ways in which they can evolve over time (*integrity constraints*), and (c) the rules to derive new facts.
It is useful to classify integrity constraints in static and dynamic integrity constraints.
The *static integrity constraints* define some conditions on the concrete knowledge values that must be fulfilled. The conditions may concern:

- the values of a property. For example, a user has the property tax code, name, address, with values of type alphanumeric characters string, and year of birth, with values of integer type; a student must be at least 17; the salary of an employee is a positive number;
- different properties values of the same entity;
- property values of different entities of the same group. For example, the students' matriculation numbers are all different. A set of properties is said *key*, as regards to a collection of elements, if (a) its values uniquely identify an element of the collection and (b) if each property of the key is necessary for this purpose. An example of key for users collection is the tax code;
- features of sets of entities.

The *dynamic integrity constraints* define some conditions on the way the concrete knowledge can evolve in the course of time. While a static constraint concerns each single state of the universe to be modelled, a dynamic constraint concerns transitions from one state to the other.

<u>*The procedural knowledge*</u>

The universe to be modelled is an evolving reality that interacts with an environment. While the abstract knowledge concerns the structure of the universe of discourse, the *procedural knowledge* concerns

[2] It is indicated with the notation A (C_1, …, C_n) an association A between the collections C_1, …, C_n with domain C_1x …x C_n (Cartesian product).

operations which the concrete knowledge may be subjected to, and in particular the effect of those operations and the way they develop.

It is useful to distinguish two types of procedural knowledge:

- the operations through which the interactions with the external environment take place, in order to provide the planned services (*knowledge of the users operations*);
- the elementary operations which involve entities in order to produce certain effects, in particular the operations for their creation, modification and cancellation that must satisfy the conditions illustrated by the integrity constraints (*knowledge of the basic operations*).

For example, the basic operations that may affect a university student are: succeeding an exam, switching to another course, etc… The basic operations related to an entity complete its meaning and characterize its behaviour.

While the basic operations are related to a single entity, a users operation aims to provide a service to the information system users according to preset modalities, coded within the organization procedures.

Abstract linguistic aspect

The abstract linguistic aspect concerns the conceptual tools, or abstraction mechanisms, adopted to model the universe of the discourse. A mechanism of abstraction is the fundamental tool for understanding one aspect of the situation to be described, omitting details considered at that time of little significance (the abstraction as a simplified force). The main mechanisms of abstraction generally used are: classification, generalization, aggregation and projection.

Concrete linguistic aspect

For the definition of the model you can use different types of formalism, as illustrated also in Chapter 5, which differ first of all for the data model they support, that is for the abstraction mechanisms offered to model and for the ways they represent the possible entities states, associations and property and for the possibility or not of offering to represent the procedural aspects.

Data models commonly used are: the *semantic data model*, the *Entity-Relationship model*, the *reticular model*, the *hierarchical model*, the *relational model* and the *object model*.

The concrete linguistic aspect concerns the characteristics, and the definition, of the formal language used to construct the reality model. Once the abstraction mechanisms to be used for modelling are set, we can use formal languages with very different characteristics that support the chosen abstraction mechanisms.

3.2.3 Geodatabase design: the adopted methodology

Designing a database means to define the overall schema of data, the integrity constraints and the applications operations, in order to prepare for the next phase in which the database and the operations will be realized[3].

[3] An application is a portion of a data processing system (see Figure 3.1) finalized at satisfying the informative needs of one organization sector. An application interacts with only one part of the database (it is said that it owns a “vision” of the database). A group of users operations (the application operations) is associated to an application.

The design of a database is a complex task, for which many methodologies and development environments have been proposed. Here, we will illustrate the methodological approach chosen, used for the design of the geodatabase studied.

Introduction

The conducted geodatabase construction process can be subdivided in three major steps:

1. *requirements analysis*: in this phase, the specific objectives of the geographic database to be realized have been identified;
2. *system design*: in this phase, it has been designed a system answering to the needs and objectives highlighted in the previous phase;
3. *realization* of the designed system.

The phases of requirements analysis and system design can be collectively called *design*, since their purpose is the production of a project of the system to be realized. The description of the results of these steps is the object of this study.
In particular, the design problem of the geographic database has been faced taking into account certain aspects:

- the design methodology has been organized, on the whole, in a sequence of phases during which a limited number of decisions at a time were taken. Each stage produced a definition of the solution at a different level of detail and requested the use of appropriate tools. For example, in the methodology that will be presented, in the first phase (requirements analysis), the information to be discussed and the external behaviour of the system were specified; in the second phase (conceptual design) this description was refined, providing in particular a global data schema; in the third phase (logical design) this description was implemented using the abstraction mechanisms of a specific data model, and in the fourth phase (physical design) additional details concerning the physical data organisation were added.
- the definition of methodological tools to be used during each stage to accomplish the methodology steps, in order to evaluate the phase results and to move from the results of a phase to the next one. These instruments were of different kinds (methods of documentation, graphical representations, various sets of forms, etc.) and aimed at different purposes depending on the phase in which they were used. They were used to produce a code able to comply with the specifications defined in the previous steps, to help the coordination between the several groups involved in the implementation or interested in the project, to facilitate the documentation and finally to ensure the correct management of the process through which the result was produced.
- generally, the approach of database design can be separated into two main categories; *data-driven* and *process-driven*. With the data-driven approach the entire focus of the design process is on data and its properties. After identifying user data requirements, a conceptual data model needs to be designed and to be implemented in a database, and then applications that use the database are developed. With the alternate process-driven approach (also called *process analysis*), any working activity within an enterprise is determined and individually analysed, then application programs are identified according to user requirements and, finally, the necessary data identified are organized in a

complete structural schema. Both approaches have advantages and disadvantages during the design of a complex system, such as the proposed one. With the data-driven approach, a complete view of the system exists, although there is a possibility to not efficiently response to specific application requirements of single users. With process-driven approach, due to consideration of single repositories separately, application requirements are identified in detail. However, there is a risk of not recognizing the data exchange or common activities, due to the fact of not taking a complete view of the system. A joint data/process-driven method has been used here in order to benefit from both approaches and to close the gaps of each one. During the study this goal was achieved by surveying user requirements and analyzing existing models as well as existing systems. This is the main topic of Chapter 4.

Database design

The adopted methodology for the geodatabase design required several steps. As already said, for complex situations it is in fact inevitable the adoption of a multiple phases methodology in which the parameters of the project are gradually considered and the various phases are coordinated in order to achieve a satisfactory realization. This approach is based on the principle that complex applications are developed through abstraction levels progressively more detailed. The most consolidated guidance foresees at least four phases (see Figure 3.8), which have been considered theoretically necessary for the design of the studied system: one for the analysis, *the requirements analysis* or *needs assessment,* and three for the design in its strict sense, the *conceptual design*, the *logical design* and the *physical design*. Indeed, as we will better see in Chapter 5, dedicated to the geodatabase modelling, this distinction between the different levels of the scheme definition hasn't been strictly respected.

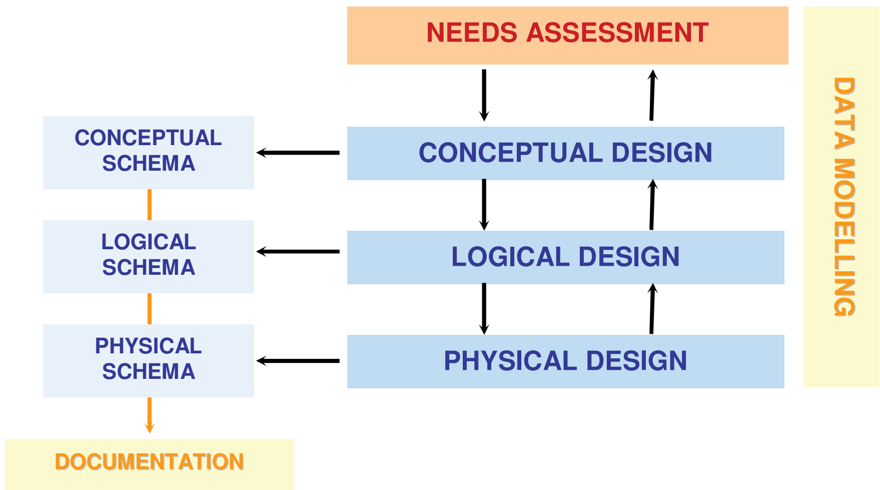


Figure 3.8 – Database design steps.

The purpose of *the requirements analysis* is to define the users needs. During the requirements analysis we establish, together with the customer organization, which are the interactions that the several organization sectors have with the organization information system, and which are the new interactions that the users wants to make possible. The result of this phase is a document, the requirements specification, which describes, in a non-ambiguous and understandable way, the users information needs.

The *conceptual design* is used to translate the requirements specification in a project of the data conceptual structure. The data structure is described by using a formalism that is as natural as possible. The product of this phase is called *conceptual schema* and describes in a formal way the information to be represented in the database. The conceptual schema is usually expressed by using a high level data model such as the Entity-Relationship model or the object model.
The *logical design* is used to translate the conceptual schema in the *logical schema*, which is expressed in the data model of the system chosen for the implementation of the database. The product of this phase, the database logical schema, refers to a logical data model. A logical model allows to describe the data according to a representation still independent from physical details, but concrete since available in database management systems. At this stage we need to know the logical model adopted, but it is not yet necessary to know the particular DBMS chosen (only the category it belongs to: hierarchical, relational, etc..).
Finally, the *physical design* is used to produce the *physical schema*, which enriches the logical schema with information related to data physical organisation (physical parameters of data storage, organization of files and indexes). The physical schema refers to a physical data model, which depends on the chosen management database system.

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*, as defined above. 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.
In table 3.1 there is a summary of the aspects on which the different stages are focused.

Table 3.1 – Different aspects considered during design steps. (source: [A. Albano, G. Ghelli, R. Orsini, 2007])

	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*

We described here the general methodology adopted for the geodatabase realization. The results of the needs assessment phase will be discussed in Chapter 4, while the specific techniques used for the conceptual and logic design phases, with their results, will be described in Chapters 5. Hints about the choices made for the system implementation phase will be proposed in the Appendix.

4

NEEDS ASSESSMENT

In this chapter the main features of the needs assessment phase are presented. Particular attention will be paid to the structural aspects (data design) and to the dynamic aspects (procedures analysis).
In this phase, the geodatabase will be considered as a fundament for ITHACA's activities and as a relevant ingredient, or Core Data container, in the construction of the SDI in the WFP SDI project. Therefore, the needs of geospatial data that are necessary for the optimal work of different applications and activities of the GIS users, in ITHACA and in the WFP, will be considered, with particular emphasis on common data.

4.1 Basic concepts

4.1.1 Introduction

The needs assessment is the first, necessary step for implementing a successful geodatabase. A needs assessment is a systematic look at how involved organizations function and the spatial data needed to carry out their activities correctly. In addition to the final reports that are generated, the needs assessment activity itself serves as a learning tool. As a matter of fact, during this phase, potential users in each participating organization learn about the geodatabase and the different applications and operations, involved with geographic data, which can be considered necessary in ongoing and planned activities.

The needs assessment phase presupposes that a suitable *preliminary analysis* of organizations interested in the project and their objectives has been carried out, with particular attention to:

- the *objectives* of the geographic database to be planned;
- the *activities* which must be supported by this database;
- the organization *units*, or functional areas, that will exploit the data contained in the database (in the following called *user groups*);
- the database *development plan*.

Therefore, the result of this preliminary study concerns fundamental choices of the involved organizations and of their way of working, presumably invariant in the medium term, which must be taken into consideration while doing the requirements analysis. The first results of this survey have been presented in Chapter 2, and briefly summarized in paragraph 4.2 below.

Subsequently, the *needs assessment* phase refines the preliminary analysis results by specifying more clearly what the organizations involved in the project, and the several sectors related to them, expect from the system being designed, arriving in particular to delineate the information structure to be treated (choice of data and their static characteristics) and the procedures (dynamic characteristics) that should be supported, whether already existing or still to be realized.

In general, the result of the needs assessment phase consists in a first specification:

- of the information exchanged or shared between sectors, setting the meaning of the terms;
- of the structure of concrete and abstract knowledge;
- of the procedural knowledge, placing the emphasis on services that the system has to offer to its users to define what operations should do, rather than how they do it.

The result of the analysis phase of the requirements may be given in different ways, that can also be: (a) informally, with natural language, (b) with appropriate tables complemented by diagrams or (c) with not executable formal language. This work will in part take into consideration the second approach.

The necessary information for a proper development of this phase are usually collected by doing two types of activities which can proceed in parallel, but that often are intertwined and affect each other: the *data analysis* and the *processes analysis.*

With the *data analysis*, we initially focus on the understanding and specifications of the structural aspects, in particular on those facts to be stored in the database, trying to give a description independent from their use

in the operations. Afterwards, we specify the operations, checking that they can be performed with the available information.
With the *analysis process* we initially specify the operations of the applications in order to consequently get a description of the information needed to execute them.
The data analysis and the processes analysis can be performed in any order, as long as at the end you check the consistency between the results of the two phases. In this work, these two analysis have been conducted with a certain degree of parallelism, taking into account the intermediate results of both analysis, according to a methodology that provides that for each examined area of the organization, at least the entities collections and the essential associations are identified (postponing any possible refinements to the conceptual modelling phase) and the operations of the studied applications are specified. Details of the adopted methodology are described in the next section. The formal instruments used to perform processes analysis will be then described in 4.1.3.

4.1.2 Needs assessment: the adopted methodology

The objective of this step is to ensure a common understanding between the design team and those users that have some interests in the implementation of the geodatabase. The main aim of the activities carried out during this phase has been to identify:

- the *applications,* which would have been supported by the geographic database: the geodatabase design has been influenced by the structure of the organizations involved in the project. As a matter of fact, distinct user groups of organizations may have responsibility for different segments of the geographic data. At basic level, we began by identifying the providers and consumers of geographic information. Then, the key data flows have been modelled. This was the starting point for defining logical groups of data.
- the *data required* to support the identified applications: for each of them, we have identified all types of data that were necessary to fulfill its requirements. For each data type, we have also identified the possible source of data. Specification of necessary data capture and processing procedures has been also documented.
- the suitable organization of the data into *logical sets of themes*: finally, we have organized data into logical groupings. From an inventory of all the necessary types of geographic data, we identified a set of groupings that comprise all of defined geographic data.

One point which can be observed here, is that, during the analysis of users' applications, we have limited our analyses to the only specification of interactions between users and interested data stores. Moreover, using suitable notations, or formalism, for each user group we have identified, at a first level of detail, the most significant entities collections and their associations. The term *entity* is used here to represent objects or things to be included in the database, and *attribute* will be the term for representing the characteristics or measurements to be recorded for the entities. The term *theme* usually refers to a collection of one or more entities, organized in a useful way which is specific to the GIS software in use. On the whole, the information collected during this phase constituted the basis for the subsequent conceptual design phase.

In particular, for each identified geodatabase user group, the following information has been collected in a specific project documentation:

- *description of the context:* a brief description of the activities of the group and their expected interactions with the geodatabase;
- *application descriptions:* an evaluation of the responsibilities and work-flows of the user group. In order to carry out some of the applications identified, certain GIS functions are required. These include standard operations such as queries and displays, spatial analysis functions (such as overlay analysis or buffering), and possibly advanced analysis requiring special programming. Descriptions of current and desired applications (GIS and not) that involve maps and geographic database have been proposed as part of the needs assessment.
- *data needed:* identification of needed geographic data and their sources. These data would have been included in the geographic database. Particular attention has been paid to:
    - the identification of "internal data", that is data produced internally by considered user groups, and "external data", that is data acquired by external providers. By looking at the work-flows and processes within and between user groups, responsibility for data creation, updates and maintenance became apparent. We indicated who is the responsible for the definition, collection, storage and distribution of considered data;
    - the identification of specific needs of sharing data across different user groups of the organization and across the different organizations here considered (mainly, ITHACA and WFP);
    - the identification of data accuracy requirements according to the adopted map scales.
- *data creation and maintenance procedures*.

Needs acquisition

The most significant aspect of a needs assessment is to document the findings in a standard and structured manner. It is very important to adopt (or develop) a standard methodology to be used for the description of all the tasks, processes and data that are included in the needs assessment.
The methodology that we used in the needs assessment phase provided for the identification of the requirements in the following ways:

- *activities of the user groups:* these complex activities or processes, where information needs to be used, distribute and kept, were described using *Data Flow Diagrams* (*DFD,* see 4.1.3);
- *cartographic activities*: map-making outputs produced by ITHACA were analysed and described using suitable *Map Product Sample Forms*;
- *spatial data*: data that are managed and stored in order to carry out the activities of the different user groups were identified and synthesized into suitable *Data Lists*.

The needs acquisition activities have been performed completely only for ITHACA's user groups. On the contrary, the major outputs of the needs acquisition activities carried out for WFP's units interested by the geodatabase design, were only Data Lists. As we will see, they have been defined according to the large documentation produced in the context of UNSDI development activities, and to the results of several official meetings and document exchanges with involved actors in the WFP.

4.1.3 Data Flow Diagrams concepts

A *data flow diagram (DFD)* is a graphical representation of the *flow of data* through an information system. Data flow diagrams are used in order to visualize how the system operates and what the system accomplishes. A data flow diagram can also be used for the visualization of system data processing operations. Therefore, DFD constitutes an excellent communication tool for analysts to model systems processes and functional requirements.

In our application, Data Flow Diagrams have been used in order to identify the fundamental operations and data flows required in all the ITHACA activities supported by the planned geodatabase.

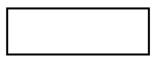
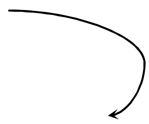
It is common practice to draw a *context level* DFD first, which shows the interaction between the system being modeled and outside entities. This context level DFD is then "exploded" to show more details of the system.

The methodology used in order to develop DFDs counts several steps. For each user group considered we:

- have made a *context level DFD* showing the interactions (or data flows) between the system (represented by a unique process) and the system environment;
- the system has been then decomposed, using lower level DFD, into a main set of processes, data stores, and data flows;
- finally, each identified process has been decomposed into an even lower level DFD diagram containing its sub-processes. This approach in theory can continue on the subsequent sub-processes, until a necessary and sufficient level of detail is reached. Each process or operation has been linked (with incoming data flows) directly with other processes or data stores. The reaction of each process or operation to a given event has been modeled by an outgoing data flow instead.

Table 4.1 shows the DFD formalisms, used in the following phases.

Table 4.1 – The Data Flow Diagram notation.

Constructions	Graphical representation
Process (or function): depending on the level of the diagram it may represent the whole system, as in a context diagram, or a process or a sub-process in lower level diagrams.	
*External Entity***:** a person or group which interacts with the system.	
Data Store: a repository of information or data. In a database logical or physical model, this represents a file, table, etc., while in a database conceptual model, a data store is an object or entity.	
Data Flow: the directional movement of data to and from External Entities, Processes and Data Stores. In the physical model, flows into a data store mean storing, updating, deleting operations, etc., while flows out of data stores mean data reading, querying, displaying, or using operations.	

4.2 Preliminary analysis

In the preliminary analysis phase the organizations involved with the geodatabase design project and their aims have been identified and described. Information discussed in Chapter 2 of this thesis (see in particular sections 2.1.2 and 2.2.3) constitutes the results of this phase. Here we provide only a summary of the major interesting issues.

The main goals of the geographic database implementation project are (Figure 4.1):

- to support ITHACA activities and to respond to needs that they require;
- to constitute a container of the *Core Data* needed in the context of the major UN WFP *Spatial Data Infrastructure* (SDI) project.

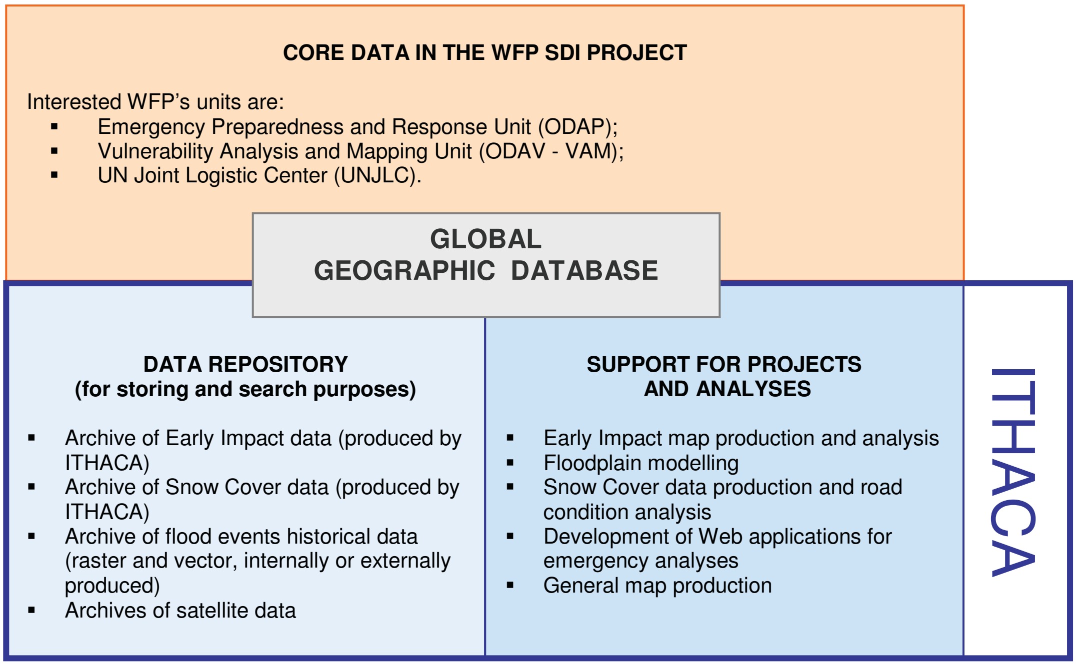


Figure 4.1 – The main implications of the development of the global geodatabase.

As you can see in Figure 4.1, in the ITHACA context, the global geodatabase implementation meets mainly the need of disposing of valuable base geographical data with global coverage, useful to referencing and map production, and of valuable and easily searchable archives for different data (externally or internally produced). Moreover, the implemented database solution must support the future development of applications and services devoted to the easy and rapid sharing of data and produced information among ITHACA user groups, and between ITHACA and external agencies involved in the humanitarian sector (e.g. WFP units or UNOSAT). As a matter of fact, many user groups, involved in exploitation of the projected geographic database, have been identified. They are shown in Figure 4.2. The major activities of these user groups have been already introduced in Chapter 2 and will be analysed more in details in the subsequent paragraphs.

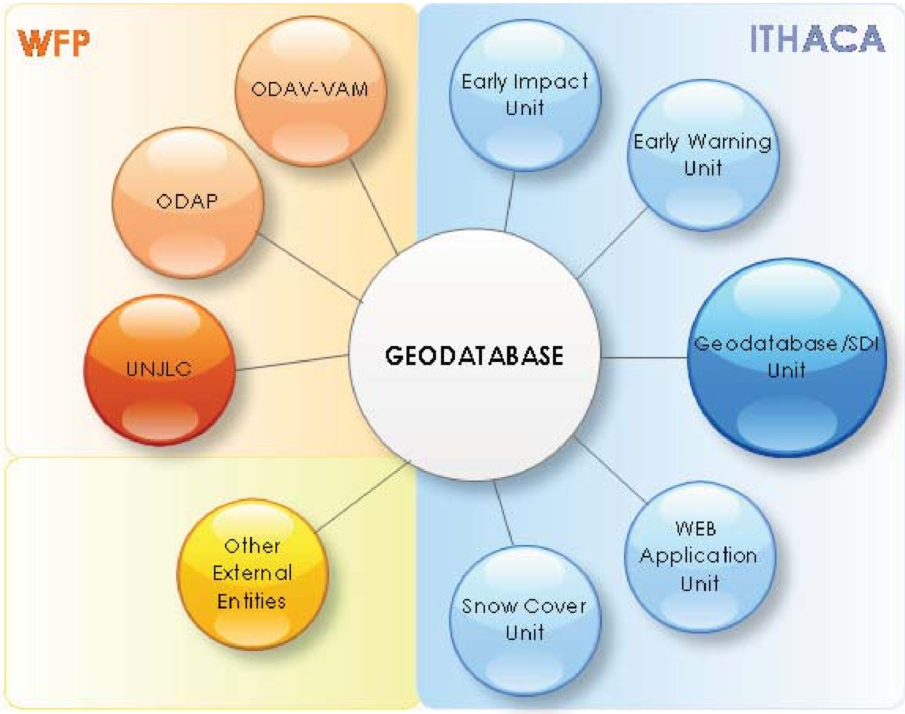


Figure 4.2 – User groups involved in the development of the global geodatabase.

As a result of implementing of a central global geodatabase, ITHACA expects to realize a number of tangible and intangible benefits. As a matter of fact, several of the potential benefits of a corporate database, that is a single organization-wide resource, include:

- to have rapid access to current, updated, digital information, with more efficient performance of routine activities;
- the reduction or the elimination of data redundancy among units, resulting in decreased operating costs and increased operating efficiencies;
- the maintenance of data integrity and quality. Shared rather than independent databases reduce problems of inconsistencies in stored information: data must follow prescribed models, rules, and standards;
- increased data query, reporting, and map production capabilities;
- increased decision support capabilities;
- increased organization and integration of geographic data and related information.

During the preliminary analysis, two important issues have been highlighted:

- according to the geographical extension of areas of interest considered in ITHACA and WFP units' activities, the planned geographic database must contain **globally consistent data** (very little scale, about 1:1,000,000) or data concerning at least the main WFP's action areas (Asia, Latin America & Caribbean, Middle East, Central Asia & Eastern Europe and Africa);
- the major outputs of ITHACA and WFP's activities is constitute by maps. The information contained in these maps needs to be distributed or shared in order to produce benefits. Therefore, during the geodatabase development activities, particular attention must be paid to identify and use **geographic**

data released into the public domain: these data cannot be copyrighted, restricted or licensed once they are released, allowing their right use and re-use for humanitarian aims.

4.3 WFP needs assessment

4.3.1 The geodatabase as a Core Data collection for the WFP SDI project

The development of consistent reusable themes of base geographic content, or *Core Data,* is generally recognized as a common ingredient in the construction of a SDI (see Chapter 2, section 2.1.1). Therefore, the data model design and the implementation of a global unique geodatabase have been considered as the first and main step of the complex project dedicated to SDI development for the WFP, conducted by ITHACA. As a consequence, in the WFP needs assessment phase, the adopted decisions taken about *Core Data* in order to implement the general UNSDI (United Nation *Spatial Data Infrastructure*; see Chapter 2) have been considered as a fundamental issue. Afterwards, the specific needs of WFP's units dealing with geographic data and GIS instruments, have been also analysed in order to identify further non-*Core Data* to be included in the planned geodatabase (section 4.3.2).

The Core Data concept

Geospatial data provide the content in a *Spatial Data Infrastructure*. The value of these data is one of the main factors for establishing a SDI.

The different UN GIS users employ a limited number of common geospatial data. Geographic data are defined as collections of georeferenced datasets. Common datasets consist of data layers, such as geodetic control, road networks, hydrography, geographical names, administrative boundaries, elevation data, and so on, upon which other sector-specific geographic data are aligned. These common data provide keys for the necessary integration to implement a UN SDI. As aforementioned in Chapter 2, the concept of "*Core Data*" is used in this work as *a set of Geographic Information necessary for optimal use of different applications and activities*. In the context of a SDI implementation, the definition of *Core Data* constitutes an instrumentality that helps to improve interoperability, reducing the expenses resulting from the inevitable duplications.

As a matter of fact, the concept of the core aims at sharing the core datasets between UN users in order to facilitate the development of GIS applications. Each data item may be supplied by a different data provider. Such data providers produce or simply collect data in order to carry out their specific activities. Although there may be many data providers, the datasets they provide must be integrated to develop core datasets. Data integration complications exist at different levels, and they can be found in four main types:

- cross-border: edge matching between different datasets;
- cross-sector: datasets created for different sector-based applications;
- cross-type: e.g. raster data vs. vector data;
- overlap: same features coming from different sources and process.

Once these core datasets are shared between data users, each user does not have to develop the core data by himself, and can avoid duplicated efforts of core data development or collection. Consequently, by sharing the cost of developing or collecting the core data, data development or collection cost can be minimised and shared between users.

Moreover, the benefits of the *Core Data* concept will be revealed when updating. Since some core datasets are developed by those who produce the data through their activities, they are updated most frequently. Therefore, the users are assured of using always up-to-date core datasets.
The proposed *Core Data* in the UNSDI project constitutes an interesting starting point which has been considered in order to correctly identify the *Core Data* in the WFP SDI project, that is, the major content of the planned global geodatabase.

Adopted approach in the UNSDI development

In Chapter 2 some examples of both developed and developing SDIs, with different degrees of success, in recent years, are presented. In the UN approach to an UNSDI implementation these results are considered very instructive. As a matter of fact, the SDIs examples provide lessons learned and help to define the essential ingredients of what constitute a successful SDI for both the United Nations and their member states.
For example, the Global Map Project is one illustration of *Core Data* sets conceived in a global or at least multi-national environment. As aforementioned in Chapter 2, the Japanese Geographical Survey Institute (GSI) took an initiative in 1992 to develop a suite of global geospatial data (Global Map) to cope with the global environmental problems. The goal was to involve national mapping organisations to collaboratively develop global geospatial datasets from existing international and national sources. This project provided a public set of reference data from trans-national to global scales to assist decision-makers and society in depicting global environmental concerns. The Global Map could be considered as an initial implementation of the concept of a suite of *Core Data* for global SDI in concert with similar framework datasets at regional and national levels.
At present, Global Map is a group of global geographic datasets which are open to the public. Global map thematic datasets provide coverage of the whole land area on earth at 1 km resolution in a consistent manner. Each thematic dataset is released to the public in a digital format for easy handling by computer and its content is equivalent to conventional maps at scales of 1 : 1,000,000.
These geographic datasets are composed of eight thematic layers (see Figure 4.3): elevation, vegetation, land cover, land use, transportation, drainage systems, boundaries and population centres. These general purpose spatial data layers are originally based on VMap Level 0 data (also known as Digital Chart of the World) for vector themes, Global Land Cover Characteristics Database from the U.S. Geological Survey (USGS) for land cover, land use and vegetation, and the 30-second GTOPO30 elevation product also hosted by the USGS. Current plans are that the data will be frequently updated by national mapping organisations to facilitate the monitoring of changes occurring in the global environment.
The geographic layers proposed in the Global Map project may be considered as candidate global data categories for other global SDI implementation projects. For instance, the Global SDI (GSDI) Cookbook authors [Global Spatial Data Infrastructure (GSDI), 2004] recommend that Global Map specification be adopted for trans-national applications requiring land cover/use, vegetation, transportation, hydrography, administrative boundaries, populated places, and elevation data. Of course, the Global Map content specification defines a simple content model with a small number of feature types and attributes suitable for the construction of base cartography at regional scales. To evaluate the level of details with respect to the

desired GIS or mapping applications it is necessary and may require extension to suit other base requirements.

There are two major reasons of the attention paid by the UN in existing *Core Data* initiatives at the sub-national, national and international scale:

- an organisation interested in implementing spatial data that will be compatible with local, regional, national, and global datasets, must identify, and potentially reconcile different *Core Data* designations across their geographic area of interest. There is a need of linking UNSDI with national SDI capacities, both in developed and developing countries. The first dialogue meeting between UNGIWG-UNSDI and national and regional-level SDI partners took place on 1 and 2 March 2007, hosted by the European Space Agency (ESA) at its ESRIN facilities in Frascati, Italy. UNSDI development discussions are also ongoing with Spain and the Federal Republic of Germany. Furthermore, SDI authorities in the following countries and regional organizations responded positively to an open UNGIWG invitation to join the UNSDI development process: Australia, Austria, Brazil, Chile, Hungary, Jamaica, India, Japan, Mexico, Morocco, Mongolia, Nigeria, Spain and South Africa.
- the *Core Data* represents a foundation on which an organisation can build by adding their own details and compiling other datasets. Existing data content may be enhanced, adjusted, or even simplified. By attaching their own geographic data, which can cover innumerable subjects and themes, to the common data in existing frameworks, users in the organization can build their applications more easily and at less cost.

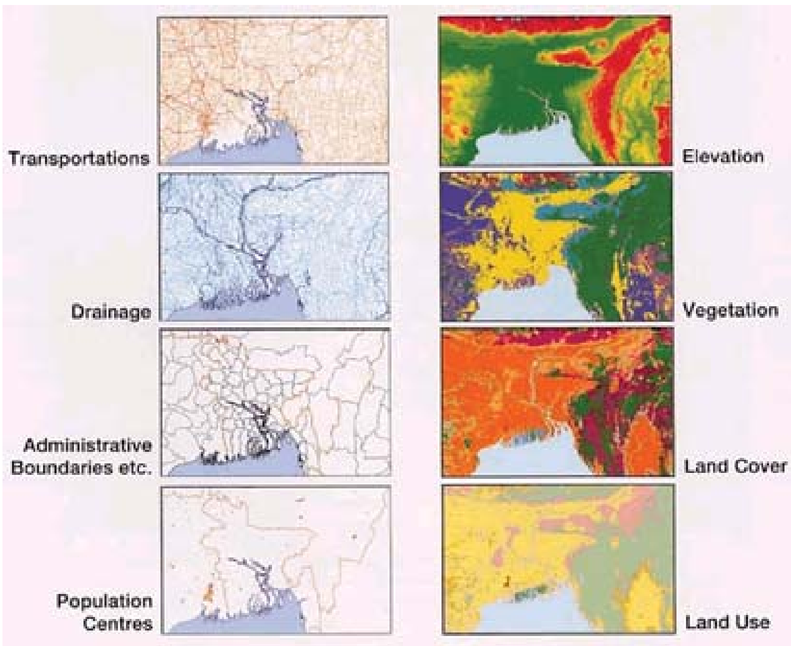


Figure 4.3 – Global Map Database at 1 km resolution.

Another important topic considered dealing with the UNSDI implementation issue is *standardisation*. Relevant elements of the United Nations agencies, funds and programmes recognize the essential requirement to adopt (or develop) widely accepted standards describing geospatial data, metadata and other

components of an SDI. Standards ease access to and sharing of data, and increase the interoperability of associated information management systems. Adoption of suitable standards is the most cost effective way in which to achieve the necessary interoperability of the data and systems in an SDI.
The GSDI Cookbook authors (see [Global Spatial Data Infrastructure (GSDI), 2004]) recommend that *Core* and non-*Core data* are modelled and shared using emerging ISO (International Organization for Standardization) standards. Within ISO there is a technical committee for geographical information, ISO/TC211, setting a wide range of (more than 20) standards. The ISO TC 211 Geomatics standardisation activity is working on several areas that greatly assist in the global specification of content models and feature models for framework and non-framework data. In particular, ISO/TC 211 adopted UML (Unified Model Language) as the conceptual schema language for specifying implementation neutral models (according to ISO 19103). This topic will be faced in Chapter 5 of this document. Other interesting standards include ISO 19109 - Rules for application schema, and 19110 - Feature cataloguing methodology. National and global framework data, as well as non-framework data, will be made more accessible and semantically correct through such technologies.

The UNGIWG and OCHA functions

As aforementioned in Chapter 2, the single most important development in the evolution of a *Spatial Data Infrastructure* to support the United Nations has been the creation of the United Nations Geographic Information Working Group (UNGIWG) in March 2000. Since its establishment, UNGIWG has stimulated improved geographic information management within and between UN Organizations and among its partners as well and, finally, has become a point of reference for UNSDI building.
In the context of the UNSDI implementation, UNGIWG and ISO are working together to align UN technical specifications for geographic information with agreed international standards. UNGIWG also cooperates with the Open Geospatial Consortium (OGC) to create open and extensible applications and interfaces for geographic information systems technologies.
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;
- build mechanisms for sharing, maintaining and assuring the quality of geographic information;
- provide a forum for discussing common issues and emerging technological changes.

The structure of UNGIWG comprises a Secretariat and five working Task Groups (TGs) that tackle priority geospatial issues impacting on the activities of UN bodies and their member states (see Table 4.2). The Secretariat, of which FAO and WFP were Co-Chairs for 2005-2006, oversees the work of the Task Groups. Chairmanship rotates between participating United Nations organizations and membership has grown to 35 UN agencies and a number of industry and not profit partners from around the globe.
The framework structure of UNGIWG includes two main areas: *Data* and *Services*, where the activities of the TGs would be focused. Both of these areas needs a common ground dealing with *Standards*.

Table 4.2 – The Task Groups of the UNGIWG.

Task Group	Thematic	Activity Area
1	Core Geo Database	DATA
2	Remote Sensing	DATA
3	Interoperable Services	SERVICES
4	Map Production Guidelines	SERVICES
5	Global Navigation Satellite Systems	SERVICES

The main activities of the different Task Groups are:

- TG 1- Core Geo-Database: this Task Group inventories and selects global datasets for adoption as UN *Spatial Data Infrastructure* core datasets. This is needed for standardizing thematic maps. A comprehensive *Inventory UNGIWG Report* has been prepared on core thematic datasets and published as a FAO document.
  Among the different core data layers at the global level, special priority has been given to produce a consistent, worldwide coverage of:
    - International Boundaries at 1:1, 1:5, 1:10 and 1:25 million;
    - Second Administrative Level Boundaries (SALB), including historic changes.

  The aim of the International Boundaries initiative is to provide the UN community with a standardized, GIS compatible, dataset of international boundaries. The dataset comes with related metadata and boundary information that reflects the cartographic practice of the United Nations. The Second Administrative Level Boundaries (SALB) initiative aims to improve the availability of information about administrative boundaries down to the second sub-national level.
  These core datasets are of particular value for reuse in the applications of many UN agencies and their partners.
- TG 2 - Remote Sensing: satellite imagery and other remotely sensed data are now an integral part of most UNGIWG members' geospatial analyses applications. The goal is to integrate remote sensing within GIS to have a better tool for decision-making.
  This Task Group facilitates the finding and assessment of very high (VHR) resolution imagery. It promotes sharing and producing data through the UN Systems Contract and other agreements, to improve their availability and use. Requests have been made for example to imagery suppliers, to ease data access for humanitarian applications. Ongoing TG initiatives also focus on improved sharing information on dataset updates, provision of compressed Landsat mosaics and assessment of their accuracy, and population of metadata on high resolution imagery. Investigations into the degree of duplication among UNGWIG members in purchasing VHR imagery are planned.
- TG 3 - Interoperable Services: international standards are needed to facilitate the sharing and use of geospatial information within the UN *Spatial Data Infrastructure*. Adoption of appropriate standards allows members to develop their respective geospatial capacities in a coherent and interoperable

manner. As a result, substantially more geospatial data can be shared between agencies avoiding, or at least minimizing, duplication. In this regard UNGIWG has adopted the ISO/TC211 standard number 19115 on Geospatial Metadata as its baseline. Overall, TG 4 aims to:

- develop guidelines that advocate greater interoperability;
- encourage the use of Open Geospatial Consortium (OGC) and ISO/TC211 standards based software.

▪ TG 4 - GIS Map Production Guidelines: this Task Group is developing guidelines for making maps using GIS software with the intention to standardize maps, scales, smallest mapping units, file interchange formats, map projections, pixel size of raster maps, etc. It has proposed technical guidelines for high resolution and low resolution scale maps and the most generalized cartographic forms regarding layout, content, marginal information, coloration, fonts, line types and symbology for scales of interest for the working group members.

▪ TG 5 - Global Navigation Satellite Systems: the Task Group works to harmonize and facilitate field data collection activities undertaken by the different UN agencies.

Some results of activities performed by these TGs have been considered relevant for carrying out the ITHACA *geodatabase/WFP SDI project*. Therefore, in the needs assessment phase, the major reports and documents produced in the last years by the UNGIWG Task Groups have been collected and analysed, with particular attention to activities of TG 1, TG 2 and TG 4. The most interesting topics identified are:

▪ **The comprehensive *Inventory Report* on core thematic datasets prepared by TG 1**

This report presents an inventory of global data sources which can be used, according to the opinion of its authors, to provide consistent geospatial baselines for selection of *Core Data* layers, suitable for generalized base mapping, emergency preparedness and response, food security and poverty mapping. This inventory was prepared by the TG 1 in conjunction with the *Poverty Mapping Project Group* (PMPG) of the FAO (*Food and Agriculture Organization* of the UN) with the main aim of developing a standard list of *Core Global Databases (CGDB)* for use across all UN agencies.

In the report, in order to narrow the results of the inventory, two restrictions were applied to the reviewed and presented data. The first of these restrictions is that the data should be globally consistent in regards to the data source, scale, and methods used for data capture and processing. The second restriction considers either: the actual scale of vector data, including a minimum scale of 1:5,000,000 and, given data availability, a maximum scale of 1:250,000; and, for raster data, a maximum nominal cell size of 5 km, but more commonly 1 km.

The sources of data presented in the inventory were identified through a review of on-line Internet resources, conducted in the first quarter of 2004 and updated in January 2005. The inventory does not cover country or project specific data which might be available from sources such as UNEP-GRID, FAO's GeoNetwork, or the regional and country offices of various UN agencies.

According to the report, the TG 1 adopted as a baseline some core data layers identified and further categorized them into a topical index covering ten areas of data specialization. These ten topical areas, shown in Table 4.3, can be considered the minimum CGDB Data to meet the evolving needs of UN database.

Another important aspect of this work is that the UNGIWG identified and proposed, for each of the defined CGDB, the potential global baseline framework data libraries and other sources, in both the public domain and from commercial sources. The potential sources identified in the report will be presented in paragraph 4.5 of this Chapter.
Evidently, the conclusions proposed by this report have been considered an important point of reference for facing the design of the global geodatabase containing the *Core Data* needed for the WFP SDI project, conducted by ITHACA *Geodatabase/SDI Unit*. Therefore, during the needs assessment phase, the defined ten topical areas, or themes, have been refined (some themes have been added, while other rejected) where necessary, in order to correctly respond to the specific needs of WFP's units involved in the WFP SDI project.
Moreover, the results of this report date back to 2005, so it has been necessary to carry out further searches in order to define, for the chosen final themes, a more updated list of data sources consistent with features described in the Preliminary Analysis. This topic is the object of paragraph 4.5.

Table 4.3 – The topical areas of Core Global Databases (CGDB) identified by UNGIWG TG1. (source: [J.F. Dooley Jr., 2005])

Topic	CGDB Data Layer
Boundaries: coastal, administrative, and areas of special interest	Coastline and Maritime Boundaries Country, Political, and Area of Dispute Boundaries Country and Sub-national Boundary Data Layers Areas of Conflict, and Landmine Dispersal Parks, Conservancies, and Protected Areas
Human health: boundaries and facilities	Human Health Infrastructure and Statistical Databases
Human population: population centres and distribution	Population Centres and Census Databases Population Density 2015 Population Census and Distribution Databases Rural Population density
Transportation: roads, railways, airports, harbours, and navigation routes	Roads Databases Railway Line, Station, and Marshalling Yard Databases Airport Databases Harbour Databases Navigation/Routes
Bathymetry and terrestrial elevation	Bathymetric Databases Terrestrial Elevation
Geophysical: geology, geomorphology, seismic, hydrogeology, and soils	Geology or Minerals Databases Geo-Morphology and Physiographic Databases Earthquake, Tsunamigenic and Volcanic Databases Hydro-Geological / Aquifer Databases Soils Databases
Surface hydrology: surface water bodies, water points, drainage and watersheds	Drainage and Flow Routing Databases Surface Water bodies (SWB) Databases Watershed and River Basin Databases Water Points and Limnological Databases

Satellite imagery, orthorectified mosaics, land cover and vegetation data	Satellite image mosaics and orthorectified imagery Satellite derivative land classification and vegetation databases
Climatic data: temperature, rainfall, and atmospheric emissions	n/e

- **TG 4 set of standards applicable to most mapping situations**

  The UNGIWG Task Group 4 has defined a *Map Production Guidelines* documentation that covers most generalized cartographic topics. It covers layout, content, marginal information, coloration, fonts, line types and symbology. Due to the vast range of products, types and uses, the proposed standards cannot be rigorous but can be considered only as guidelines.

  The activities carried out by the TG 4 in order to develop and distribute *Map Guidelines* include:
  - identification of standard elements of a print map;
  - collection of map symbol sets from various UN GIS shops to compile a digital symbol library;
  - multiple map layout design;
  - creation of ArcGIS map templates and their PDF versions for easy viewing;
  - creation of an ArcGIS Symbol-set Library;
  - generation of a set of individual graphic files for each map symbol;
  - definition of colour gradients and line styles for a limited set of feature classes;
  - generation of a set of example maps for download to show the map templates in action;
  - development of a *Map Production Guidelines* distribution prototype web-site for UN OCHA[1].

  Also, a subset of norms for map scales, minimum mapping units, cell sizes and projections has been developed and evaluated for adoption as a standard. Some of the recommended standards for UNGIWG use have been considered in the needs assessment phase. They are presented in Table 4.4.

Table 4.4 – Considered UNGIWG Core Spatial Standards Set - GIS Guidelines (source: UNGIWG documents)

Argument	Recommended standards for UNGIWG use	Notes
Data formats	Maintain as much as possible compatibility with ESRI data formats. ESRI Geodatabase recommended for central database. For data distribution, and in absence of a geodatabase schema, ArcInfo coverages and GRIDs are currently recommended; Shapefiles when ArcInfo coverages are not supported. GeoTiff format for images.	ESRI data formats are the most commonly used in the UN system; they are often use as formats for data exchange at international level.
Map scales	Datasets for distribution should be suitable for displays analysis at four map scales: 1:1,000,000; 1:5,000,000; 1:10,000,000; 1:40,000,000	Commonly used scales; international datasets at global level are built on this scales reference.
Cell-size for raster data	Given according to map scale it is recommended to use: 1:1,000,000 200m 0.001666° (6 arc-sec) 1:5,000,000 1000m 0.008333° (30 arc-sec)	These are the most commonly used cell sizes on global maps developed by the UN, NGOs and other international organizations.

[1] http://ochaonline2.un.org/mapguidelines

	1:10,000,000 2,000m 0.016666° (1 arc-minute) 1:40,000,000 4,000m 0.083333° (5 arc-minutes) or at least an exact multiple or divisor of the cell sizes above	
Datum	WGS84	
Metadata	Compatible with ISO 19115 metadata standard	ISO 19115 is being adopted by several UN agencies.

4.3.2 WFP specific needs assessment

In order to achieve the needs assessment phase for the WFP, the specific needs of its units dealing with geographic data and GIS systems have been also analysed in order to identify further necessary non-*Core Data*. For these purposes, it has been necessary to know the different kinds of geographic data used by WFP's units interested in the *Global Geodatabase/SDI project* and analyse how these data were normally managed. The conducted activities required several meetings with the units and a large exchange of documentation. Figure 4.4 shows an example of digital form developed by ITHACA and distributed to the interested users in the WFP in order to collect information about spatial data exploited during their activities.

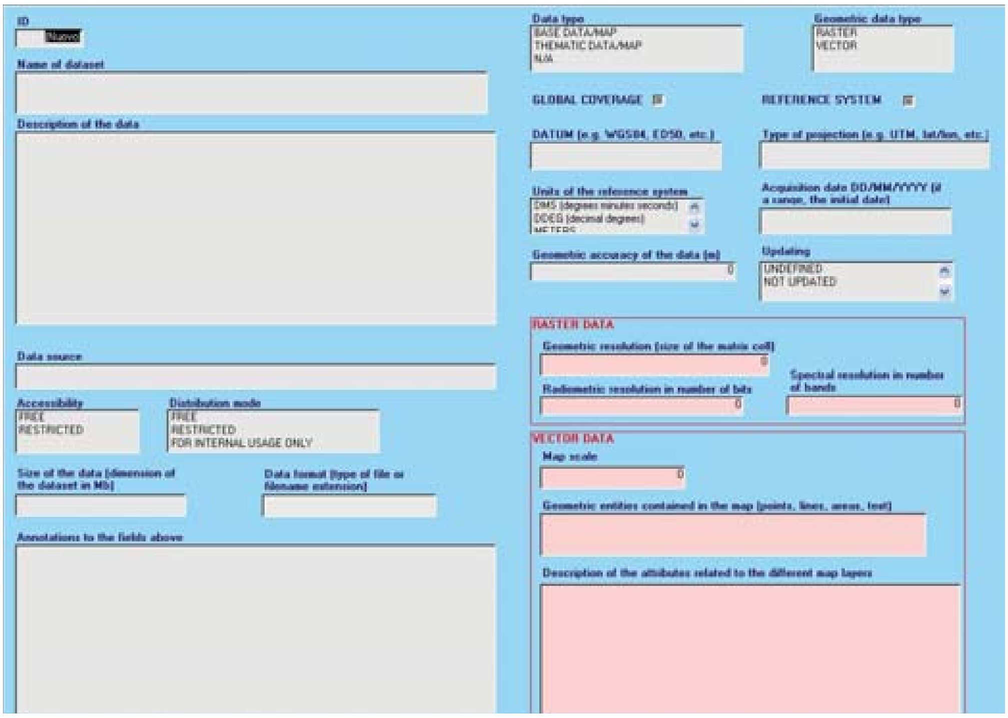


Figure 4.4 – Form used for geographic data information exchange.

In WFP organization, three units are potentially interested in the *Global Geodatabase/SDI project*. These units, dealing with geographic data stored in GIS, are:

- the *Emergency Preparedness and Response Unit - ODAP;*
- the *Vulnerability Analysis and Mapping Unit - ODAV / VAM ;*
- the *UN Joint Logistic Center – UNJLC*.

Activities of these units have been already introduced in Chapter 2, section 2.1.2. Here we discuss briefly only their spatial data needs.

ODAP unit

The activities of this unit deal with the rapid response to natural disasters in order to mitigate the impacts on hit populations. In this field, the ODAP unit, working in collaboration with ITHACA, attends to production of cartographic works (at a small scale, 1:1M to 1:20M) in support to WFP’s rescue operations. Moreover, as already discussed, ODAP manages a web-site dedicated to necessary data and information distribution to involved users during emergencies.

ODAP has and maintains a geographic data repository for the cartographic production. The primary data used by ODAP, according to the information exchanges, are:

- road and railroad network;
- populated places (settlements);
- population density data;
- Point of Interest (POI): elements including food distribution points, schools, and hospitals;
- administrative boundaries;
- transportation data;
- elevation data;
- land cover / land use data;
- hydrography data (river/stream networks and water bodies);
- world ecoregions;
- additional data derived from Country Offices (COs) and Regional Bureaux (RB): these data are country specific and not globally consistent;
- other data types, mainly aerial or satellite images, are considered interesting. Thus, the possibility of international agreements with satellite data providers (mainly high resolution data providers) is examined.

According to their corresponding data sources (presented in the paragraph 4.5), some of these data suffer for problems of accuracy, updating, completeness, name spelling and attribute coding.

Usually, the used spatial data are organized and subdivided on a country basis. All data are stored in geographic coordinates (datum: WGS84) and in ESRI compatible formats on a shared data server for *internal users*. *External users*, as Regional Bureaux or Country Offices, need data export activities to access to the central repository data and, sometimes, the volume of data exchanged is a problem, due to low speed and performance internet connections.

Data updating operations are frequent. All updates are managed internally by headquarter offices personnel (in Rome), according to local agents reports. Therefore, the main problem is to grant reliability and integrity (data formats and projections, up-to-date sources and workflow, metadata) of data contained in the ODAP repository. Finally, for data updates, satellite images are considered a relevant data source but no systematic images use is made.

During the needs assessment phase some subjects have been proposed:

- the possibility to increase updating efficiency and data integrity using disconnected editing with versions reconciliation (only for local offices having GIS knowledge and instruments);
- the possibility to design valuable Web GIS solutions for data visualization and editing, keeping in consideration the limitations imposed by low speed and quality internet connections;
- finally, the possibility to use open source platforms for GIS data management.

ODAV-VAM unit

The main field of work of the VAM unit concerns the monitoring of vulnerable areas before, during and after a catastrophic event or a crisis, with particular emphasis on the food security topic. During its activities, the VAM produces various reports and map products, often based on satellite derived data and indicators. Understanding the nature of food insecurity and monitoring emerging food security problems are the major objects of VAM analyses. Primary data used in analyses are, for example, household surveys, nutrition surveys, and market price surveys. Additional used data are remote sensing imagery, agricultural production forecasts, market information, political and insecurity information.
In particular, global geographic data used by VAM, which had to be included in the planned geodatabase are (according to the information exchanges):

- population data;
- elevation data;
- land cover / land use data;
- administrative boundaries;
- global map of irrigated areas: percentage of land equipped for irrigation;
- global faming systems: these data define broad farming system categories at a global scale;
- populated places;
- hydrography data (river/stream networks and water bodies);
- transportation data;
- NDVI: Normalized Difference Vegetation Index data (satellite derived);
- RFE Rainfall Estimates (satellite derived);
- WFP Programme Presence: presence of WFP School Feeding programmes, Mother & Child health and nutrition programmes, HIV/AIDS programmes at sub-national unit level;
- WFP Facilities: location of WFP offices and warehouses.

Moreover, the VAM is adopting an "enterprise" approach in managing geographic information. It is referred to the VAM *Spatial Information Environment* (SIE), using the GeoNetwork software that FAO has developed (see also Chapter 2, section 2.1.2). SIE enables WFP Country Offices (CO), Regional Bureaux (RB), and headquarters to access and exchange georeferenced food security databases and cartographic products from a variety of sources. The VAM SIE is a Java data repository application which allows the data management (based on a file system approach) and the metadata management (following ISO19115 standard), supporting dynamic mapping too. VAM SIE network has a central node in Rome, where a common repository is established. This node make regular harvesting of metadata from distributed nodes, leaving data resident on those nodes.

UNJLC

The UNJLC is a inter-agency unit that deals with logistic data. At the moment, it operates in the custody of the WFP agency. The UNJLC, collaborating directly with other UN units, mainly supplies the logistic support needed for the WFP operations during natural disasters emergencies.

The 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 enriched and updated locally in case of emergency activations. In those occasions, the interested portion of geodatabase is checked out from the main one and provided to local officers 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 and the new acquired information is inserted in the main geodatabase, after a necessary validation procedure.

When we started the needs assessment phase, the UNJLC geodatabase conceptual model, the terminology definition and data collection forms and specifics were available. ITHACA support has been requested only for the conceptual model translation in the UML (this topic will be faced in the subsequent Chapter 5) and for the implementation (the data sources, with reference scale 1:1M, were also identified by the UNJLC). Moreover, when we started the needs assessment phase, considered future developments of the geodatabase project conducted by the UNJLC dealt with the definition of Web services with mapping functionalities.

Final remarks for this phase are:

- as the majority of UN organizations seem to have adopted GIS software from the Environmental Systems Resources Institute (ESRI), we considered as primary need the development of an architecture compatible with ESRI ArcGIS suite, available and well known in UN GIS structure. This choice was perfectly in conformity with UNGIWG GIS *Guidelines* (see Table 4.4). The proposed default solution has been to use Oracle as RDBMS and ArcSDE as access gateway (this topic will be faced in Chapter 5);
- together with WFP's users, we defined a time plan for the realization of the geodatabase. According to the defined timeline, we decided to concentrate initial efforts in developing a first version of the geodatabase which would have contained only fundamental, globally consistent, well-established and tested data. The data needed in more specific applications and the local ones provided by COs and RB, which are not fundamental, have not been included in the planned

version of the geodatabase. Therefore, those kinds of data have not been considered in subsequent steps of the geodatabase design, described in this thesis. These data will be analysed, modelled and stored only in a second phase of the geodatabase project.

4.4 ITHACA needs assessment

The major aim of the planned geodatabase is to constitute a central data repository and system used by all ITHACA's user groups (see Figure 4.2) for their activities. Through the planned system, each group must:

- have access to central geographic data used by different applications (reuse of data, avoiding any duplication and controlling their quality);
- have access to partial results of other groups or sub-groups activities (sharing of data).

ITHACA works closely with WFP's *Emergency Preparedness and Response branch* (ODAP) in order to deliver technological and methodological support and services to WFP, and potentially to other UN Agencies. Besides to support the WFP SDI implementation effort, ITHACA pays great attention to the development of technologies and systems for early warning and early impact analyses for emergencies arising from natural hazards with the major aim of providing WFP decision-makers with real-time information and data generated by these developed systems. In particular, during emergencies, timely and valuable maps produced by ITHACA are required by WFP's aid workers to locate and identify affected areas and to implement and correctly manage effective disaster mitigation operations. For these purposes, the planned central geodatabase must provide a useful way to store, manage and maintain information and data produced in ITHACA emergencies analyses, making them more easily accessible to WFP's users.

4.4.1 ITHACA's Early Impact Unit

This unit deals mainly with supporting supply of flood disasters mitigation operations conducted by WFP. Standard procedures related to flood events have been defined and tested.

ITHACA support in early impact activities for flood events is conceived mainly in provision of *map products* showing the impact of the flood on population and on infrastructures, with particular emphasis on road network. Furthermore, additional analysis is requested to facilitate the set up of priorities in undertaking field emergency needs assessment and finally in supporting midterm food security analysis.

In operative terms, map-supporting is required at two different stages: at the first state of alert and just after the event. In both cases the response time after help request triggering is short, lasting generally from 24 to 48 hours.

In the first state of alert, before the flood event, small scale maps indicating the areas that will be potentially interested by the floods (the so called 'floodable areas') are required: in this phase the floodable areas are defined and reference satellite data (Radar, MODIS, etc) about the zone of interest (before the flood) are collected.

In the second phase, after the flood hits, the following four points must be fulfilled:

- which are the areas physically hit by the flood (the so called '*flooded areas*');

- which are the areas that could be hit directly or indirectly by the flood effects (the so called '*affected areas*');
- which are the population suffering the flood effects (the so called '*affected population*');
- which are the *damages* occurred to the infrastructures.

In order to derive this information, it is necessary to receive updated satellite data and to apply classification, photo interpretation techniques and GIS procedures, as it is shown in the next sections. After its definition, this information needs to be mapped in a suitable format, respecting rules of clarity and easiness of distribution defined by the WFP's users.

In order to correctly manage its activities, the Early Impact Unit strove to define standard work-flows including tested procedures obtained using an approach that required the definition, the collection and the analysis of large "*lesson learned*" documents, exchanged in the last year with WFP's users involved in disaster mitigation activities and with WFP's field teams.

Two standard work-flows have been defined according to the different types of satellite data available (Figure 4.5). Usually, monitoring activities and small scale analyses on the flood effects are carried out using free low spatial resolution optical satellite data (250m – 500m), easily available in internet. These analyses may be invalidated because of the cloud cover presence.

Otherwise, for some flood events, specific rush satellite imagery acquisition can be requested in order to obtain radar data and/or optical medium (2.5m – 30m) and high (0.6m - 2.5m) spatial resolution data allowing medium and large scale analyses over specific areas of interest. Acquisition requests are generally addressed to the *International Charter "Space and Major Disaster"* system[2]. The *International Charter* is a unified system of space data acquisition and delivery to those populations affected by natural or man-made disasters through authorized users. If the Charter is triggered, normally, in a few days, different kinds of satellite images are available, such as radar images and high resolution ones, suitable for the definition of the flooded areas and damage assessment operations.

[2] www.disastercharter.org

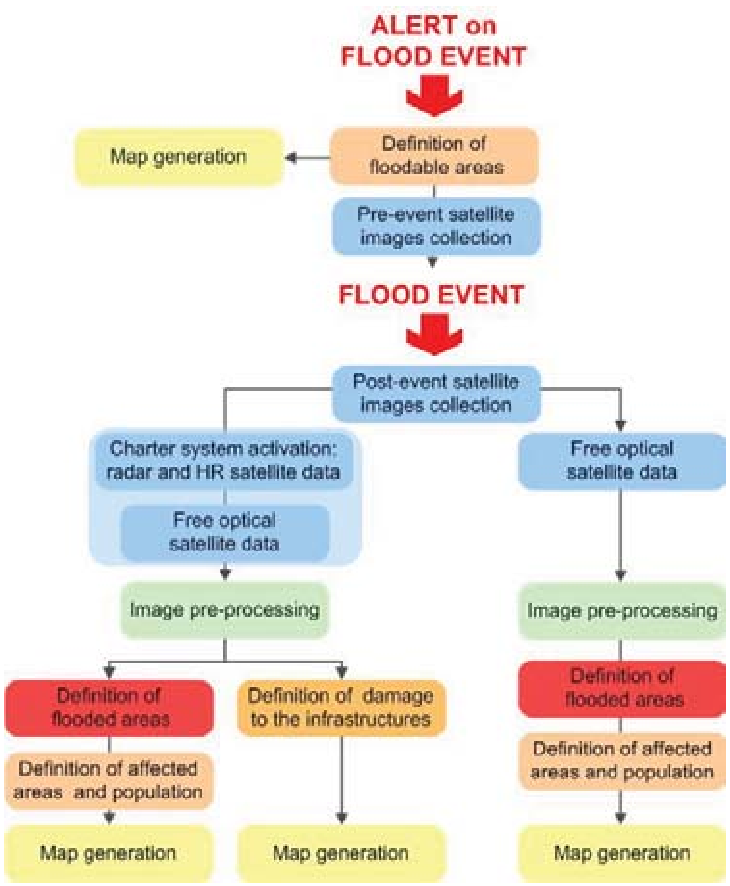


Figure 4.5 - ITHACA Early Impact Unit activities: possible work-flows during a flood emergency event.

Unit's needs assessment

In the needs assessment phase we have performed a *processes analysis* of the Early Impact Unit, that is, a complete identification of all the processes carried out by the users in their activities. First of all, this analysis has required the decomposition of the identified "event management" general process (see Figure 4.6) in some sub-processes which then have been analysed in order to identify all necessary geographic data and entities (*Data Lists*), which had to be included in the planned geodatabase, and data flows among users. As aforementioned, in order to carry out this phase in a standard manner, we turned to DFD formalisms, previously described.

As you can see in Figure 4.6, in order to correctly face the activities required when a flood emergency starts up, during the general process two sub-units need to produce, obtain and modify several data (flows coming and leaving) and exchange each other (through the geodatabase). Final products (generally map products) are supplied to WFP's users and to other external users (flows leaving).

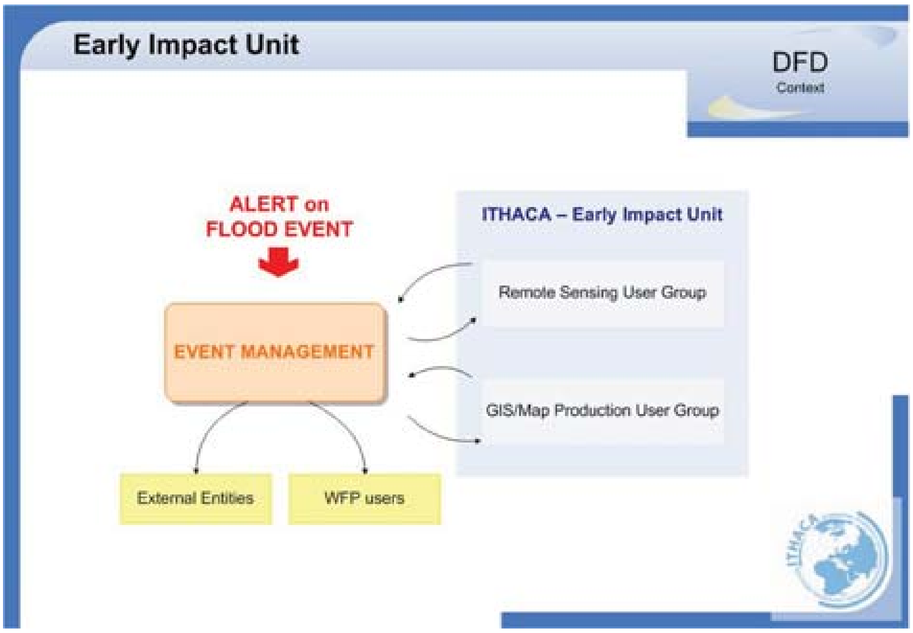


Figure 4.6 – Early Impact Unit, Data Flow Diagram: the context diagram.

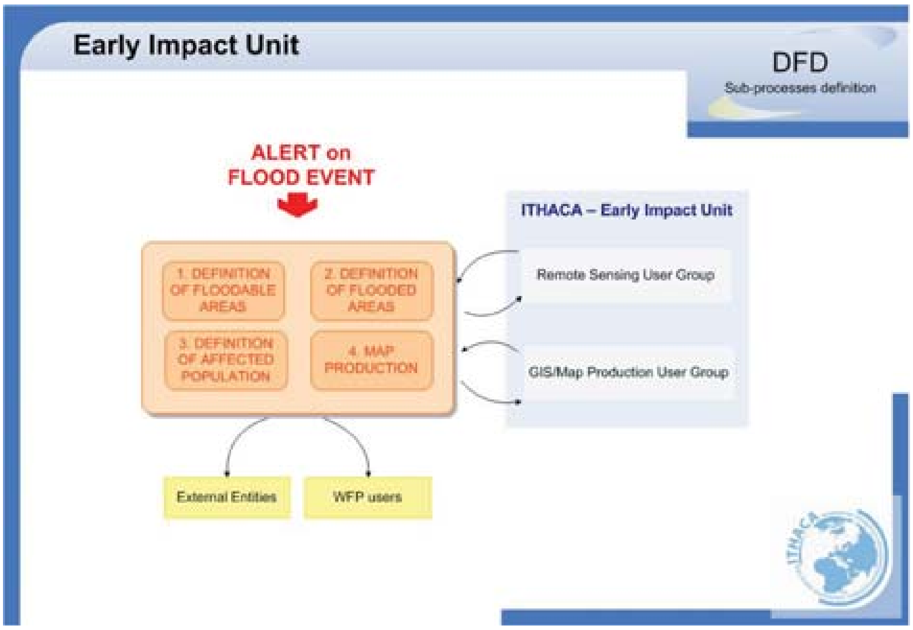


Figure 4.7 – Early Impact Unit, Data Flow Diagram: major sub-processes identification.

The mentioned sub-units are:

- *Remote Sensing user group*: this sub-unit performs satellite data processing and analysis in order to extract flooded areas and other additional information. To correctly accomplish its task, this user group needs some preliminary produced data (the floodable areas), reference satellite data and reference geographic vector data; all these are more easily found if stored in a central geodatabase, where then can be placed final results of analyses carried out by this group.

- *GIS/Map Production user group*: this sub-unit performs all GIS analyses useful to obtain intermediate products, shared with *Remote Sensing user group*, affected areas definition with estimate of affected population and, finally, map layout arranging and map production. Evidently, the activities of this group are those which may realize the maximum profit from the structuring and the organization of the used data in a central geodatabase.

Figure 4.7 shows the results of the decomposition of the “event management” process. Four sub-processes were identified:

1. *Definition of floodable areas*
2. *Definition of flooded areas*
3. *Definition of affected population*
4. *Map production*

1. Definition of floodable areas

Subsequently to either the diffusion of meteorological alerts showing next flood emergencies or the diffusion of field requests (usually by the WFP’s Country Offices, COs), ITHACA Early Impact unit puts in action. At first, maps showing an overall view, including all the areas that are potentially subject to the risk, are produced. This approach allows to produce small scale flood risk zones maps, that is, maps showing the identified *floodable areas*. These maps contribute to the hazard and vulnerability aspects of flooding, with the aim of:

- identifying areas that have greatest risk: these are considered priority areas which require deeper analyses in following processes;
- leading the collection of reference (pre-event) satellite imagery;
- supplying control data to use in following processes.

Floodable areas are obtained using a flood peak scenarios simulation approach using a specific procedure based on *geoprocessing tools* in a GIS environment. Particularly, free-share software for hydraulic modeling, coupled with the ArcView ESRI dedicated extensions are exploited.

For simulation purposes, an hydraulic model is used, coupled with historical discharge values, as input parameters. The hydraulic model is based on HEC-RAS (*Hydrologic Engineering Center’s River Analysis System*, developed by the US Army Corps).

The model analyses natural channels networks and calculates water surface profiles using one-dimensional steady and unsteady flow equations. This model requires topographic information extracted from the Digital Elevation Model, roughness coefficients (Manning’s *n,* estimated by combining land use data with tables of Manning’s *n* values) and, finally, discharge data. Discharge data are obtained from discharge measurement stations, considering historical time series; therefore, the whole floodable areas definition process is influenced by the availability of measurement stations over considered areas of interest.

Using different discharge values, it is possible to obtain different flood extension scenarios. Finally, obtained flood extent data are used in order to produce valuable flood hazard map products (Figure 4.8).

Figure 4.8 – Definition of floodable areas: output example (evaluation of flood extent in Zambesi floodplain for the extreme meteorological event occurred in Mozambique, February 2007).

Figure 4.9 shows the I Level DFD identified for this process in the needs assessment phase. As you can see, some coming and leaving data flows and corresponding data stores have been identified; they have to be included in the planned geodatabase.

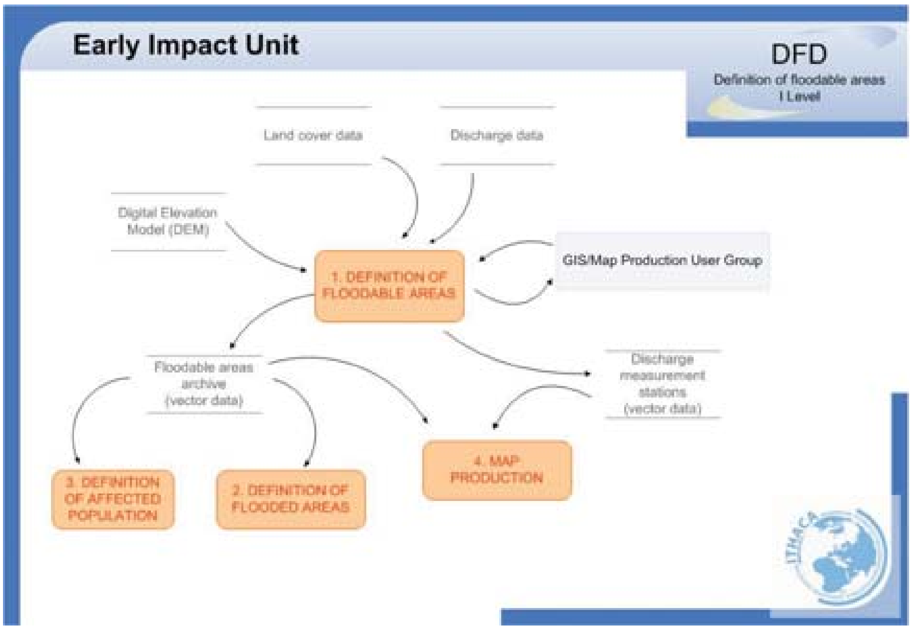


Figure 4.9 – Early Impact Unit, Data Flow Diagram: definition of floodable areas process, I Level.

<u>2. Definition of flooded areas</u>

The response category can also be called "*relief*", and refers to actions taken during and immediately following a disaster. As a matter of fact, in the response phase, the main goals of the Early Impact unit are flooded areas mapping and flood extent monitoring. Flooded areas data constitute the necessary input of the following process dedicated to the affected population estimate.

In this response phase, satellite remote sensing data suitable for effective flooded areas identification are used. A great variety of remote sensing instruments is mounted on satellite platforms, employing various measurement technologies and techniques and acquiring data in a wide range of the electromagnetic spectrum and with various spatial resolutions. Usually, the factors that determine the effectiveness of a remote sensing system for emergencies map production activities are:

- *sensor features:* spatial and spectral resolution of the sensor and area of coverage;
- *system features:* temporal resolution (or revisit time) of the system, data cost and data availability (this is the data suppliers capability of planning a new satellite acquisition to respond to an emergency need and of providing fast and reliable delivery services).

In order to extract the necessary information, various satellites are used by the Early Impact unit, in a "multi-scale" approach. Low spatial resolution multispectral images, with a wide swath, are employed to provide regional small scale flooded areas definition and to set limits of the most critical regions (priority areas or areas of interest). After identifying these zones, medium and high spatial resolution multispectral images can be utilized in order to perform accurate flooded areas extraction activities. The use of radar imagery is necessary when optical image acquisitions are ineffective because of a massive presence of clouds or heavy haze. Finally, very high spatial resolution images (panchromatic and/or multispectral) are useful to perform damage assessment activities.

As aforementioned, optical high resolution and radar data are available only if the *International Charter "Space and Major Disaster"* system is activated. Otherwise, only free satellite data are used allowing to perform only small scale analyses. Therefore, two different analysis types are carried out in order to extract required information, according to available satellite data:

- **Definition of flooded areas through MODIS data classification**

  MODIS (*Moderate Resolution Imaging Spectroradiometer*) is a sensor currently flying onboard the NASA's Terra and Aqua satellites. This instrument employs a conventional imaging-radiometer concept, consisting of a cross-track scan mirror and collecting optics, and a set of linear detector arrays with spectral interference filters located in four focal planes. The optical arrangement provides imagery in 36 discrete spectral bands from 0.4 to 14.5 µm, selected for diagnostic significance in Earth science. The spectral bands have spatial resolutions of 250 m, 500 m, or 1 km at nadir.

  Each of the MODIS instruments, operating continuously, grant a worldwide coverage providing daily images and derived products. All acquired MODIS data are available in near-real time and completely free of charge by searching and ordering from the NASA *Earth Observing System Data Gateway* (EDG) and from several ftp sites.

  The wide area seen by MODIS sensor and the near-daily availability of its data allow a regional view of the observed phenomena and flood monitoring over the entire globe. Therefore, the use of MODIS data permits:

- to observe the overall situation in a large area immediately after the event in order to identify priority areas (useful, for instance, to trigger requests to Charter for medium and high resolution satellite data acquisitions);
- to extract flooded areas and to carry out multitemporal small scale analysis of the flood event evolution in the areas of interest.

On the other hand, MODIS data, like other optical satellite data, can be ineffective because of presence of persistent cloud cover over the areas interested by the event. In these cases radar data must be used.

The detection of water bodies and flooded areas is the result of a classifying procedure of MODIS primary reflectance data (MOD02 - *MODIS Level 1B Calibrated, Geolocated Radiances*).

The MOD02 product is distributed by NASA to the users with a low processing level (*Swath* format), which performs only a first geometric correction step and which contains some residual geometric distortions due to the system peculiar scanning geometry. Therefore, a preliminary geometric data correction step is always necessary.

Then, in order to correctly compare multitemporal MODIS data, the images are also preventively radiometrically pre-processed and geometrically co-registered. The atmospheric correction of the reflectance data is performed through a simplified Dark Subtraction approach.

Moreover, since MODIS data over areas of interest are available daily, mosaics of subsequent images are used in order to reduce cloud cover presence.

After performing pre-processing operations, water bodies are extracted from images using conventional classification techniques (Figure 4.10). To this end, supervised classification algorithms and/or a threshold approach based on specific radiometric indexes are used in order to detect areas covered by water and, if necessary, wet areas, separating them from land surface. The clearest depiction of water versus land surfaces is seen in bands where land surfaces are highly reflective and water surfaces are highly absorbing. The MODIS bands which are more often used in order to accomplish classification operations are Red (620-670 nm) and Near-Infrared (841-876 nm) bands (both with spatial resolution of 250 m). The used indexes are mostly based on particular band rapports. Indexes commonly used are the NDVI (Normalized Difference Vegetation Index) and other ones defined by authors according to the specific areas of interest.

Then, using reliable water bodies reference data it is possible to isolate only flooded areas. It is important to have a reference data in order to avoid mistakes caused by inclusion in flooded areas of areas that are normally under water, such as lakes, reservoirs or wetlands. Two types of reference data are considered:

- because MODIS near-daily observations of every area on Earth are stored by NASA and made available to user community, it is possible to construct baseline maps of interested areas when they are not flooded. Imagery acquired immediately before the event can be used too;
- global water bodies and river network vector data.

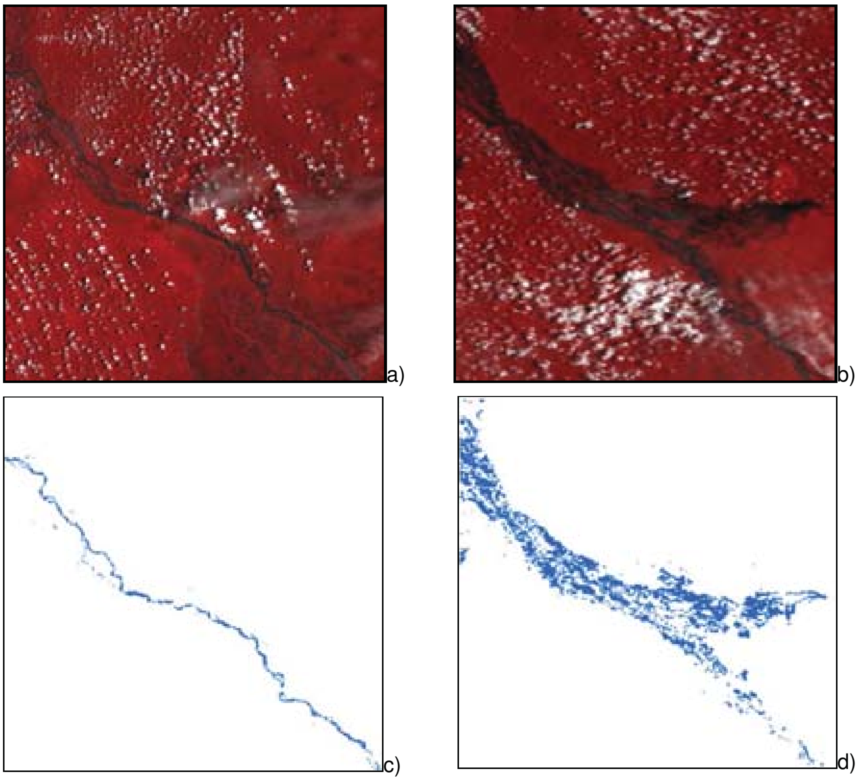


Figure 4.10 - Mozambique flood event, February 2007. Above: original MODIS data (false colour), pre-event reference data (mosaic) (a), post-event mosaic (b); below: corresponding identified water bodies and flooded areas (c) and (d).

- **Definition of flooded areas through radar images**

  Radar images (geometric resolution varying from ~1m to ~150m) enable to easily identify water bodies on the scene, also with persistent cloud cover presence. On the other side, they are affected by geometric distortions (layover, foreshortening and radar shadows) hard to be modelled, especially in mountain regions.

  In detail, radar images acquired by satellite platforms (often Envisat and Ers) before and after the event are used for the definition of flooded area (the request for an archive radar image can be skipped if reliable and updated water bodies data are available). Both images are georeferenced through attitude and position information (generally supplied as metadata) exploiting the sensor depending procedures provided by commercial softwares. In order to improve the mapping accuracy, it is possible to perform a further image-to-image georeferencing, through a simple polynomial transformation, of the radar image preceding the flood with respect to the following one.

  From a radiometric point of view it is generally advisable to preliminary apply a despeckle filter on the images, to reduce the noise that affect radar images. Then, the areas presenting water can be spotted on both images, exploiting their reflexive behaviour towards the electromagnetic radiation emitted by the radar sensor, that can be assimilated, roughly, to the one of a specular surface. It turned out that water can be easily identified (Figure 4.11b), being characterized by low radiometric values (Figure 4.11a).

a) b) c)

Figure 4.11 - Post-event SAR image (a), extraction or water bodies (b), identification of flood-affected areas (c)

By using change detection techniques it is therefore possible to isolate only the flooded areas, distinguishing them from water bodies, as shown in Figure 4.11c. Classified images often suffer from a lack of spatial coherency (speckle or holes in classified areas). In order to improve the topology of the classified flooded areas, it is possible to further process the results applying filtering algorithms aimed to remove isolated water pixels by using blob grouping and to clump adjacent similar water areas together through the use of morphological operators.

Finally, the flooded areas identified by the *Remote Sensing group*, are converted in a vector format and masked using the floodable areas layers that will allow to remove residual classification errors. The final data become then the input of *GIS/Map Production group*'s activities.
Figures 4.12 and 4.13 show the I and II Level DFDs which have been identified for this process in the needs assessment phase (the type of used satellite data wasn't considered).

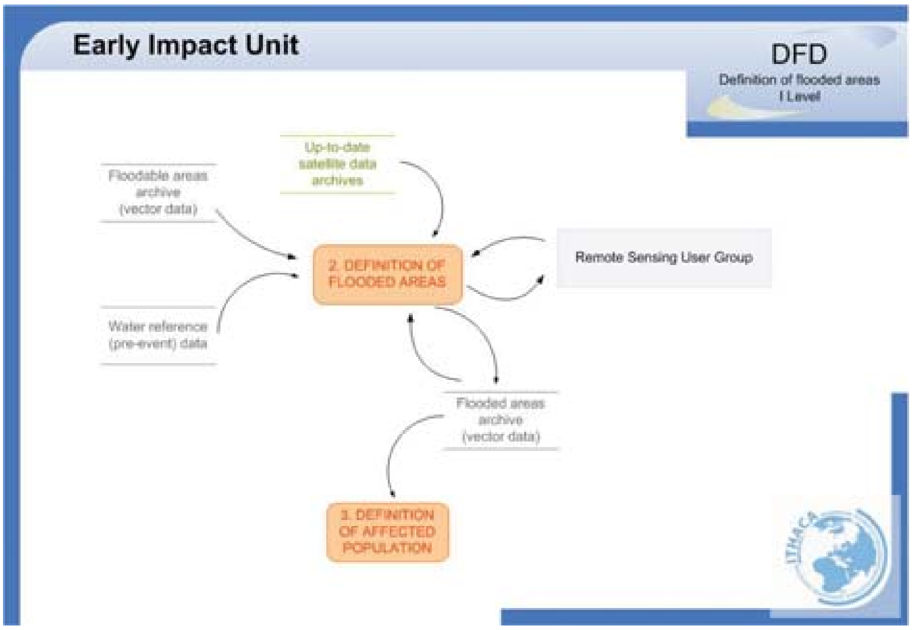


Figure 4.12 – Early Impact Unit, Data Flow Diagram: definition of flooded areas process, I Level.

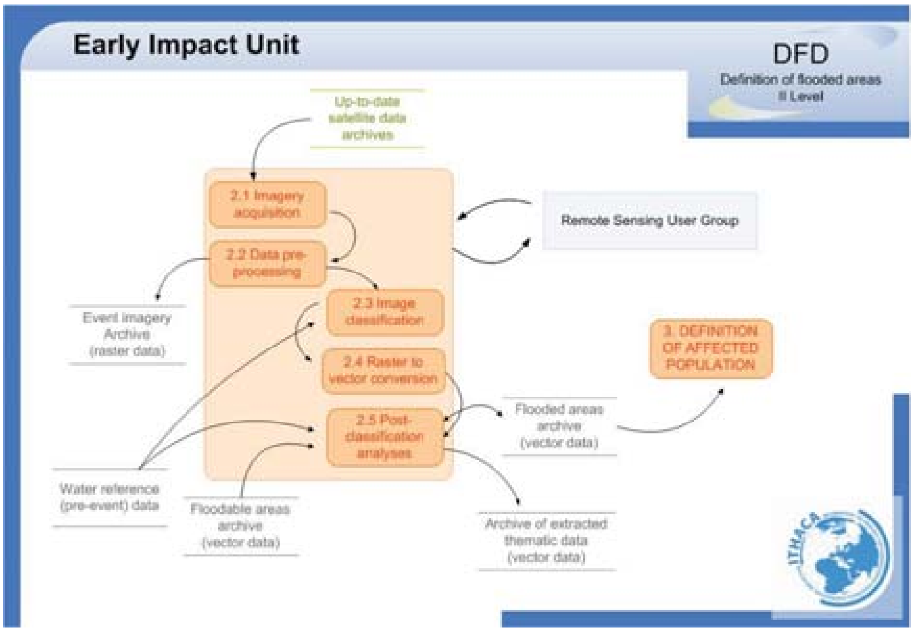


Figure 4.13 – Early Impact Unit, Data Flow Diagram: definition of flooded areas process, II Level.

As you can see in Figure 4.13, the process of definition of flooded areas has been decomposed in five basic operations requiring some data flows. In order to correctly manage identified data flows, the creation, in the planned geodatabase, of the following data stores or archives has been proposed:

- Imagery acquisition: this sub-process refers to the acquisition of base satellite data useful to extract desired information. Request for imagery is directed to *external up-to-date satellite data archives*, such as the MODIS one.
- Data pre-processing: this sub-process refers to the performing of all the preliminary processing operations that made base data ready for subsequent activities. Geometric pre-processing allows to correctly overlay data coming from different sources. The accuracy of the georeferencing operation should be suitable with the required map scale. Pre-processed data are stored in a *Event Imagery archive*.
- Image classification: this sub-process refers to the extraction of water bodies from images;
- Post-classification analyses: these are required in order to obtain correct flooded areas data from image classification results. In this phase data coming from different archives are used:
    - *Water reference (pre-event) data*: raster satellite data acquired before the event (not stored data) and/or global surface hydrology vector data. It is noticed that hydrology vector reference data should have scale consistent with satellite data resolution and, anyway, errors should be consistent with the required final map scales;
    - *Flooded areas archive*: flooded areas vector data extracted during other flood events analyses (data produced by ITHACA, also for historical flood events);
    - *Floodable areas archive*: floodable data produced in the prior process.
- Raster to vector conversion: final results are converted in vector format and stored and shared with the *GIS/Map Production group* in order to perform subsequent processes (definition of affected

population and map production). Data stored in this phase are mainly: flooded areas, areas covered by clouds, wet areas, and burnt areas (*Flooded areas archive*, *Archive of extracted thematic data*).

Another activity performed by the *Remote Sensing group* aims to identify and map the damages to the infrastructures due to the flood event. To this end, the use of high spatial resolution satellite data (geometric resolution varying from 0.6 m to 2.5 m) allows to create specific large scale products showing the flood impact on the infrastructures, with particular emphasis on road network. This phase aims to provide information on accessibility, in the context of the emergency response operations undertaken by WFP. As a matter of fact, in the food supplying process after a flood event, WFP often has to face troubles related to transportation and food aid distribution; it is therefore necessary to quickly know the condition of roads, tracks and trails where food trucks should run through.

In order to perform an assessment of damages caused by flood events, common photointerpretation and change detection techniques are used. Moreover, the use of image fusion procedures, as the pan-sharpening techniques, can improve the object identification activities. In this phase, data acquired in the field (such as georeferenced pictures acquired from helicopters or aerial platforms), provided by WFP Country Offices, often support the satellite image interpretation.

3. Definition of affected population

Flooded area extracted in the prior process are then used in order to estimate the population affected by the event. This is a very important information, useful for WFP's users to measure the amount of foodstuff that are necessary and the magnitude of the event.

In order to correctly estimate the population affected by the event it is necessary to define also the overall area that can be indirectly affected by the flood, the so called "affected areas". The floodable areas are zones where, even if they are not covered by flood water, effects due to the ongoing crisis may be detectable (for instance, in these areas interruption of road networks can be noticeable).

It is therefore necessary to process the flooded areas in order to extract the required affected areas. The goal is to have homogeneous large polygons that frame all the flood polygons, taking into consideration only the floodable areas, as illustrated in Figure 4.14.

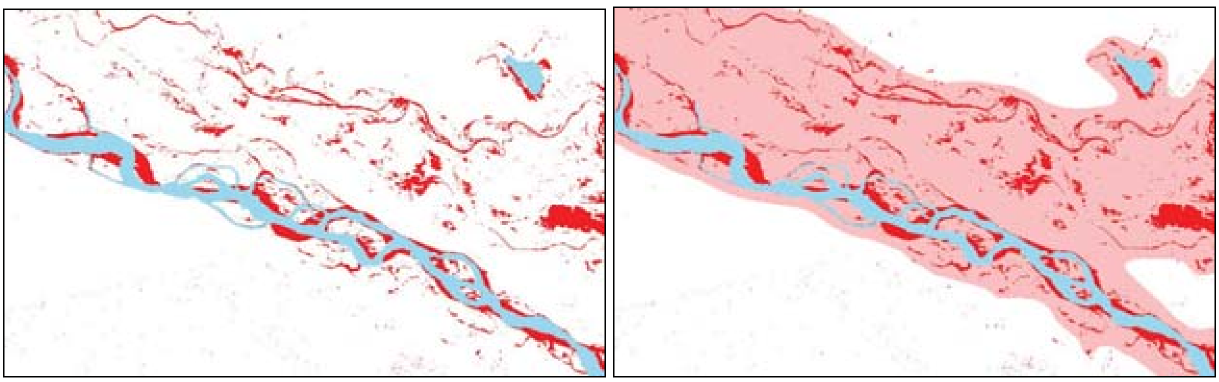


Figure 4.14 - Reference water (sky-blue) and flooded areas (red) on the left image. Affected areas (light red) on the right.

The methodology defined to perform this task is based on *mathematical morphology filtering*. First of all, it is performed a dilation of the flood polygons followed by subsequent erosion using the same structural element (“closing” filtering). This procedure allows to smooth the contours, fuse narrow breaks and long thin gulfs, eliminate small holes, and fill gaps in the contours of an image. Finally a masking is performed to remove any areas that cannot be flooded, according to the already performed GIS flood simulation.

Either polygons of flooded areas and affected areas are then used in order to calculate affected population. The final map products, delivered to WFP’s COs, carry a table (Figure 4.15) which contains the following figures, subdivided according to the considered administrative boundaries:

- *number of people living in flooded areas*: people who are living in flooded locations;
- *number of people living in affected areas*: people who are living in areas where the effects of the flood are felt. They may or may not be living in flooded locations.

Population

District Name	N. of people living in flooded areas		N. of people living in affected areas	
	Min	Max	Min	Max
Chinde	1264	1,544	7,281	8,899
Marromeu	2234	2,730	5,906	7,218
Mopeia	2767	3,381	20,445	24,989
Morrumbala	155	189	533	651
TOTAL	**6,419**	**7,845**	**34,165**	**41,757**

Figure 4.15 – Example of population figures included in ITHACA’s map products.

A rapid estimate of this parameters is obtained crossing flooded and flood affected areas with globally consistent population distribution data. These kind of data are commonly available in raster format, with about 1 to 5 km of spatial resolution.

Using GIS *Zonal Statistics* function is a way to calculate statistics on values of a raster within the zones of another dataset: the flooded or affected polygons in that case. Nevertheless, that approach gets some mistakes in the population estimates due to the presence of border raster cells which intersect only partially the borders of considered polygons. In these cases, population values of the border cells are used in order to calculate a final incorrect people sum. Therefore, using ESRI libraries, a Visual Basic procedure has been developed by the *GIS/Map Production group* with the aim of reducing errors which affect these population figures. The developed procedure, using an approach based on population density (people/unit of surface), estimates correct population values for border cells considering only the actual portion of these cells that belong to considered polygons.

Finally, the obtained affected population data constitute an input for the subsequent map production process. Figures 4.16 and 4.17 show the I and II Level DFD which have been identified for this process in the needs assessment phase.

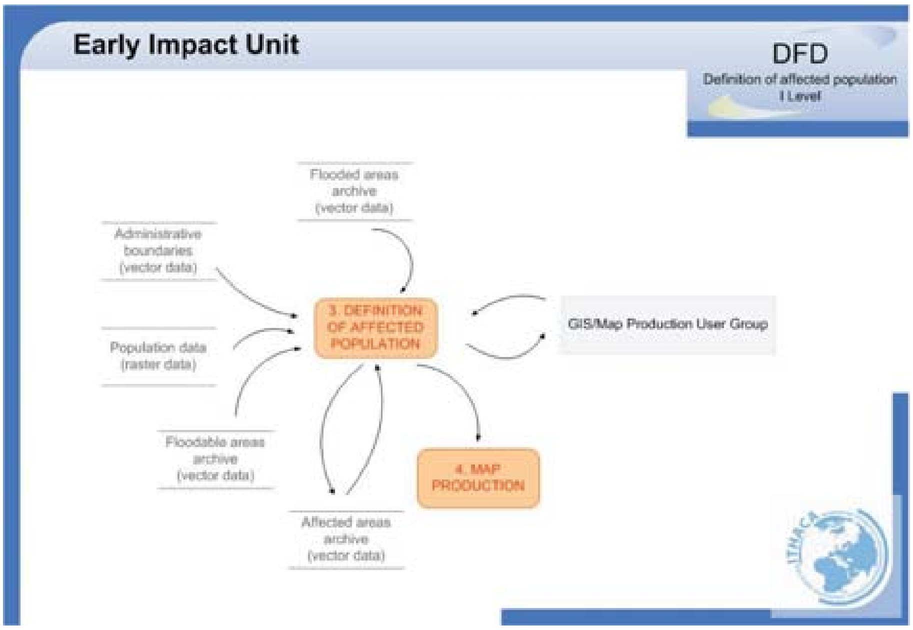


Figure 4.16 – Early Impact Unit, Data Flow Diagram: definition of affected population process, I Level.

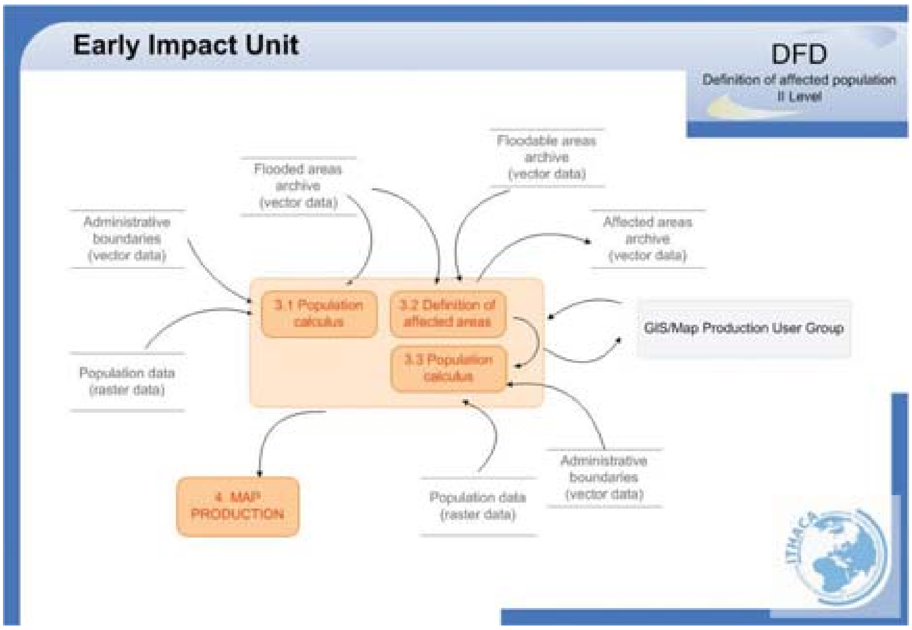


Figure 4.17 – Early Impact Unit, Data Flow Diagram: definition of affected population process, II Level.

As you can see in the DFDs, the process of definition of affected population has been decomposed in three basic operations which require some data flows. Some data stores have been considered necessary in order to retain within the geodatabase all input data needed in these operations:

- *flooded areas archive* and *affected areas archive*: these data stores contain the major input data useful to affected population calculation. For the affected areas archive, the identified data flow is coming and leaving, indicating that data are either produced or utilized during this process;

- *administrative boundaries* and *population data*: these are thematic data needed in this process;
- *floodable data archive*: these data, as aforementioned, are used in order to perform geometric control on produced affected areas polygons.

4.Map production

This is the last sub-process of the “event management” general process. In this phase the *GIS/Map Production user group* implements a GIS project concerning the flood affected areas and performs all the map layout arranging operations (in compliance with the above mentioned UNGIWG *Map Production Guidelines*) necessary to produce desired final cartographic products.

The implemented GIS project allows to collect and integrate all geographic data which must be included in planned map products.

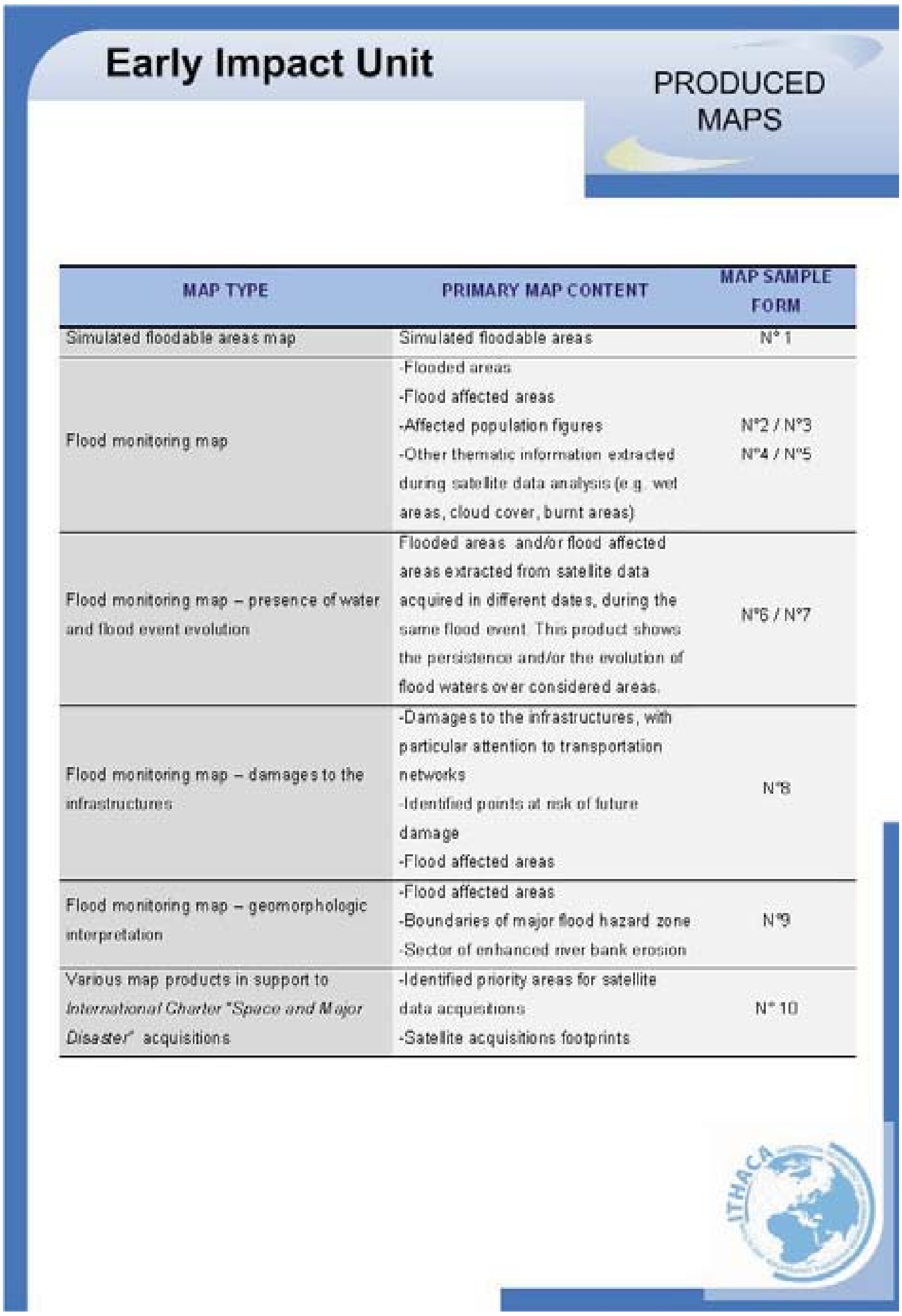
Many kinds of cartographic products are produced by ITHACA Early Impact unit. The different types of products and their features are defined by this unit according to large suggestions proposed by WFP’s map users.

During the needs assessment phase, examples of different produced maps have been collected and standard map product samples have been defined and analysed with the aim of identify all spatial data and archives which are necessary in the considered map production process and, as a consequence, which must be included in the planned geodatabase. For this purpose, *Map Product Sample Forms* have been developed.

Moreover, these forms also describe ITHACA map products in a standard manner, useful to share this kind of information with the final map users for future discussions.

The different types of maps identified are presented in the *Maps List* in Figure 4.18. Some examples of *Map Sample Forms,* used to describe considered cartographic products and their data content, are presented in Figures 4.19 – 4.28.

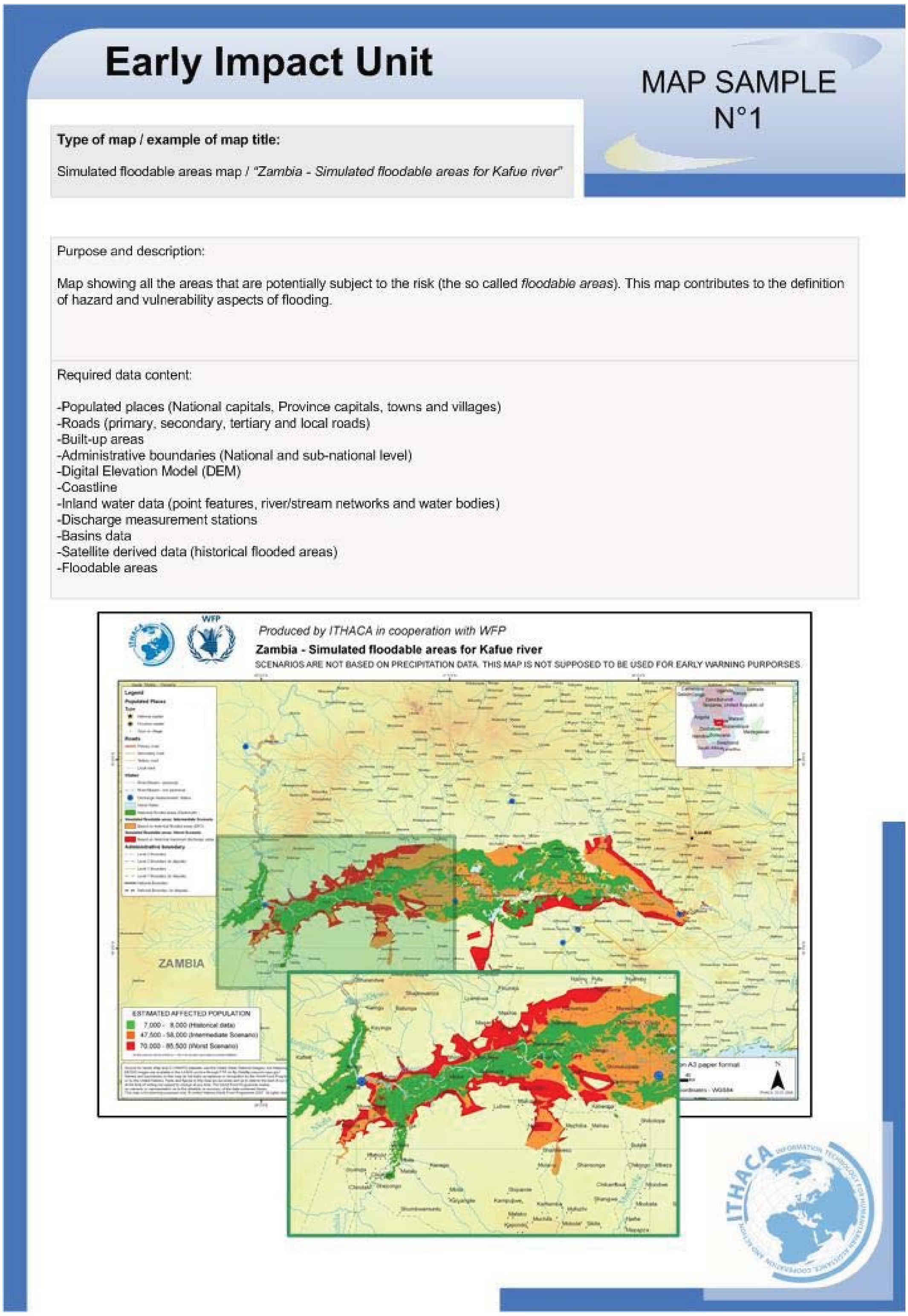
One point that can be noted here is that all maps produced by ITHACA need to be stored in order to allow WFP’s users to perform historical maps searching operations. Therefore, the planned geodatabase will contain also all ITHACA map products in raster format.



MAP TYPE	PRIMARY MAP CONTENT	MAP SAMPLE FORM
Simulated floodable areas map	Simulated floodable areas	N° 1
Flood monitoring map	-Flooded areas -Flood affected areas -Affected population figures -Other thematic information extracted during satellite data analysis (e.g. wet areas, cloud cover, burnt areas)	N°2 / N°3 N°4 / N°5
Flood monitoring map – presence of water and flood event evolution	Flooded areas and/or flood affected areas extracted from satellite data acquired in different dates, during the same flood event. This product shows the persistence and/or the evolution of flood waters over considered areas.	N°6 / N°7
Flood monitoring map – damages to the infrastructures	-Damages to the infrastructures, with particular attention to transportation networks -Identified points at risk of future damage -Flood affected areas	N°8
Flood monitoring map – geomorphologic interpretation	-Flood affected areas -Boundaries of major flood hazard zone -Sector of enhanced river bank erosion	N°9
Various map products in support to *International Charter "Space and Major Disaster"* acquisitions	-Identified priority areas for satellite data acquisitions -Satellite acquisitions footprints	N° 10

Figure 4.18 – Early Impact Unit: produced maps.

Finally, table in Figure 4.29 sums up all required entities collections identified: these collections are necessary in order to correctly produce analysed Early Impact cartographic products. Therefore, they must be included in the planned geodatabase. It is noticed that some proposed themes, as populated places or roads, are considered necessary in the map production process according to suggestions of final map users and/or according to UNGIWG *Map Production Guidelines*.



Early Impact Unit

MAP SAMPLE N°1

Type of map / example of map title:

Simulated floodable areas map / *"Zambia - Simulated floodable areas for Kafue river"*

Purpose and description:

Map showing all the areas that are potentially subject to the risk (the so called *floodable areas*). This map contributes to the definition of hazard and vulnerability aspects of flooding.

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Digital Elevation Model (DEM)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Discharge measurement stations
-Basins data
-Satellite derived data (historical flooded areas)
-Floodable areas

Figure 4.19 – Early Impact Unit: Map Sample Form N°1.

Early Impact Unit

MAP SAMPLE N°2

Type of map / example of map title:

Flood monitoring map - Flooded areas/ *"Uganda flood affected areas as detected from satellite 16 September, 2007"*

Purpose and description:

Map showing:
-flooded areas and affected areas as detected from used satellite imagery (MODIS data or medium/high spatial resolution radar data, as specified)
-calculated affected population figures concerning flooded areas and/or affected areas.

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Digital Elevation Model (DEM)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Satellite derived data (flooded data, affected areas, historical flooded areas)
-If applicable, raster satellite background (data used for analyses or archival data)
-Affected population figures
-If applicable, other thematic information extracted during satellite data analysis (e.g. wet areas, cloud cover, burnt areas).

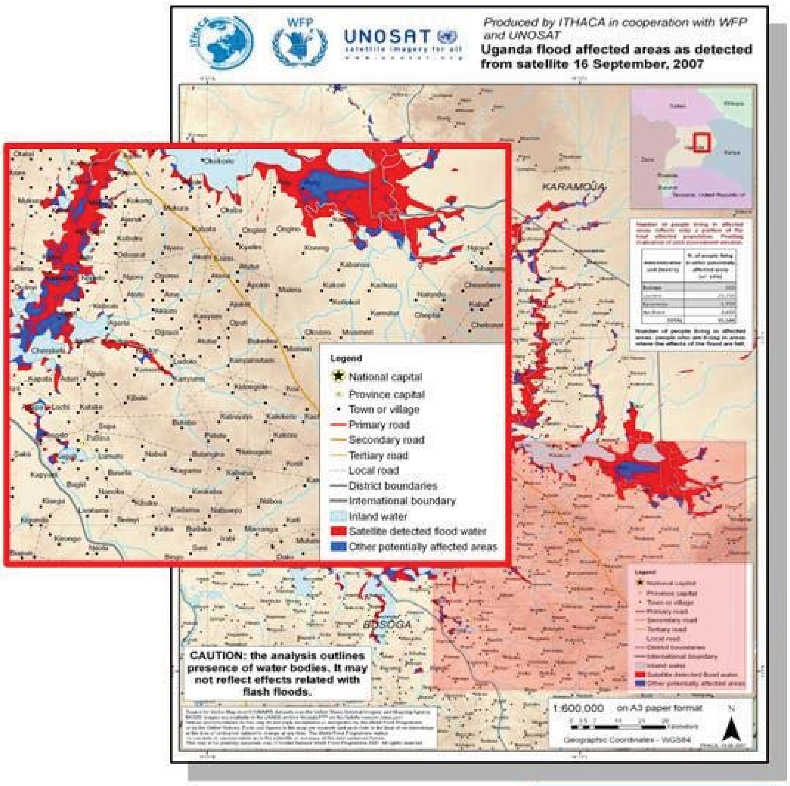


Figure 4.20 – Early Impact Unit: Map Sample Form N°2.

Early Impact Unit

MAP SAMPLE N°3

Type of map / example of map title:

Flood monitoring map - Flooded areas/ *"Ghana, Togo and Burkina Faso flood affected areas as detected from satellite 1 October, 2007"*

Purpose and description:

Map showing:
-flooded areas and affected areas as detected from used satellite imagery (MODIS data or medium/high spatial resolution radar data, as specified)
-other thematic information extracted from satellite data (wet areas, cloud cover, burnt areas, as specified)

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Digital Elevation Model (DEM)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Satellite derived data (flooded data, affected areas, historical flooded areas)
-If applicable, raster satellite background (data used for analyses or archival data)
-If applicable, other thematic information extracted during satellite data analysis (e.g. wet areas, cloud cover, burnt areas).

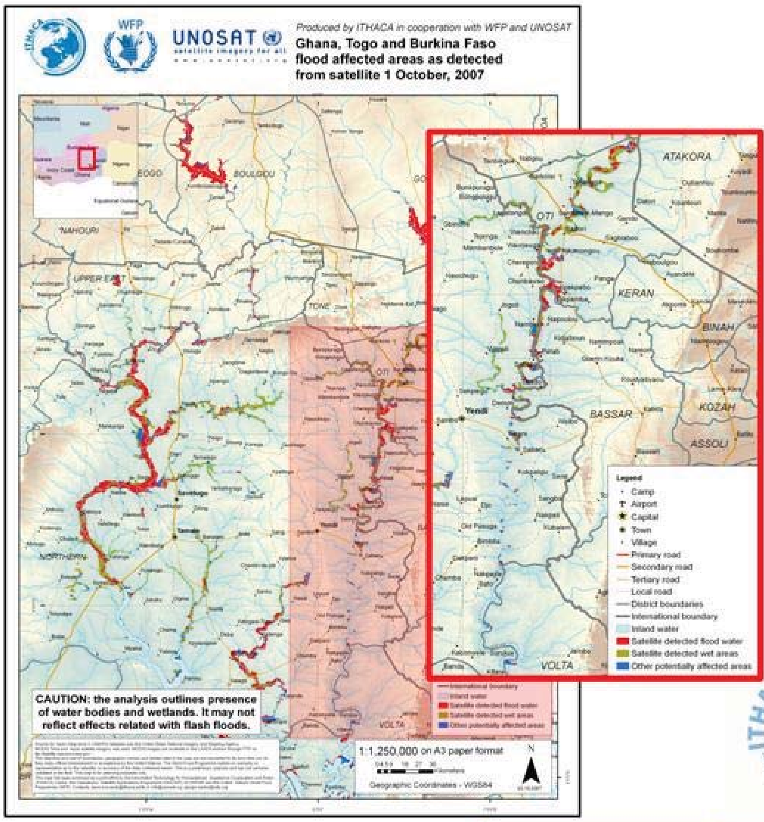


Figure 4.21 – Early Impact Unit: Map Sample Form N°3.

Early Impact Unit

MAP SAMPLE N°4

Type of map / example of map title:

Flood monitoring map - Flooded areas/ *"Pakistan Flood – Preliminary analysis of cloud-free areas based on July 2nd MODIS satellite images"*

Purpose and description:

Map showing:
-flooded areas and affected areas as detected from used satellite imagery (MODIS data or medium/high spatial resolution radar data, as specified)
-calculated affected population figures concerning flooded areas and/or affected areas
-other thematic information extracted from satellite data (wet areas, cloud cover, burnt areas, as specified)

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Digital Elevation Model (DEM)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Satellite derived data (flooded data, affected areas, historical flooded areas)
-Land cover data
-If applicable, raster satellite background (data used for analyses or archival data)
-If applicable, other thematic information extracted during satellite data analysis (e.g. wet areas, cloud cover, burnt areas).

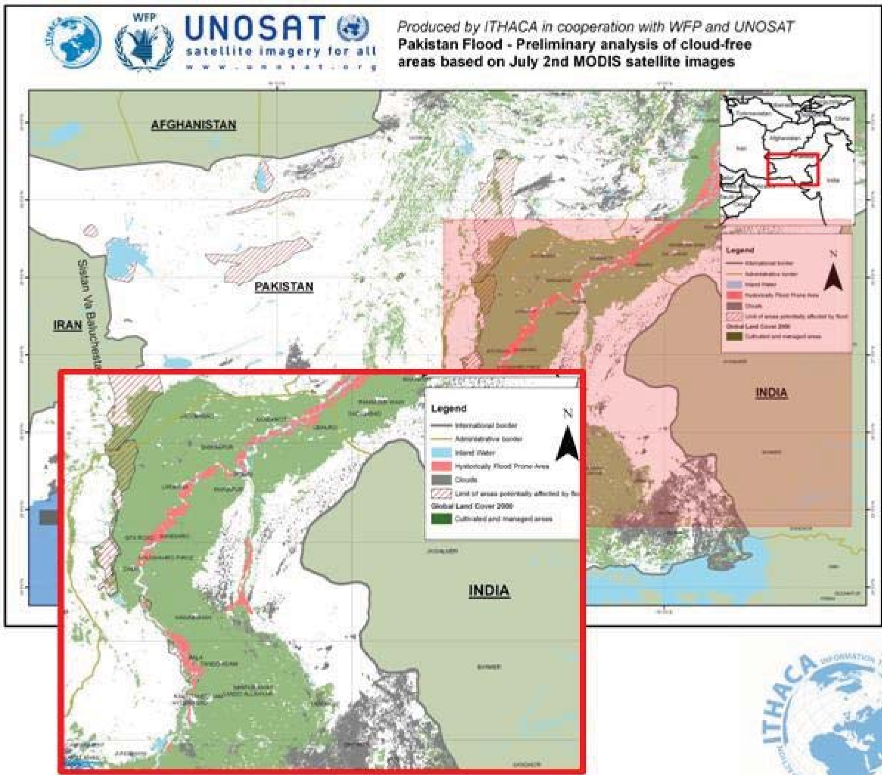


Figure 4.22 – Early Impact Unit: Map Sample Form N°4.

Early Impact Unit

MAP SAMPLE N°5

Type of map / example of map title:

Flood monitoring map - *"Mozambique, Zambezi flood monitoring, February 14, 2007 updating"*

Purpose and description:

Map showing:
-flooded areas and affected areas as detected from used satellite imagery (MODIS data or medium/high spatial resolution radar data, as specified)
-calculated affected population figures concerning flooded areas and/or affected areas

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Digital Elevation Model (DEM)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Satellite derived data (flooded data, affected areas, historical flooded areas)
-If applicable, raster satellite background (data used for analyses or archival data)
-If applicable, other thematic information extracted during satellite data analysis (e.g. wet areas, cloud cover, burnt areas).

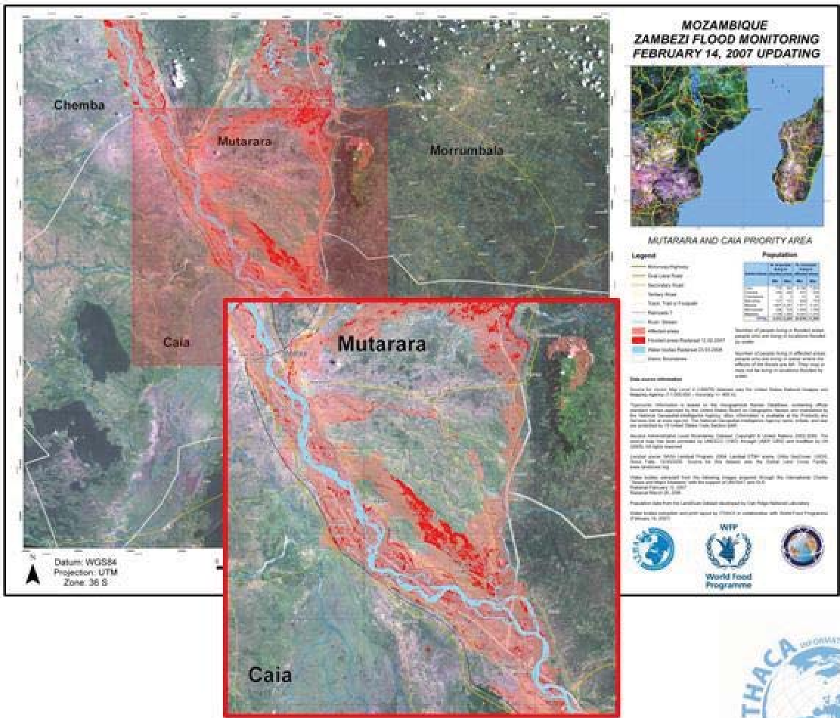


Figure 4.23 – Early Impact Unit: Map Sample Form N°5.

Early Impact Unit

MAP SAMPLE
N°6

Type of map / example of map title:

Flood monitoring map – Presence of water and flood event evolution / *"Mozambique, February - March 2007 floods, affected areas"*

Purpose and description:

Map showing:
-flooded areas and/or affected areas as detected from satellite images (MODIS data or medium/high spatial resolution radar data, as specified) acquired in different dates
-persistence of flood waters over considered areas

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Satellite derived data (flooded data, affected areas, historical flooded areas)
-If applicable, raster satellite background (data used for analyses or archival data).

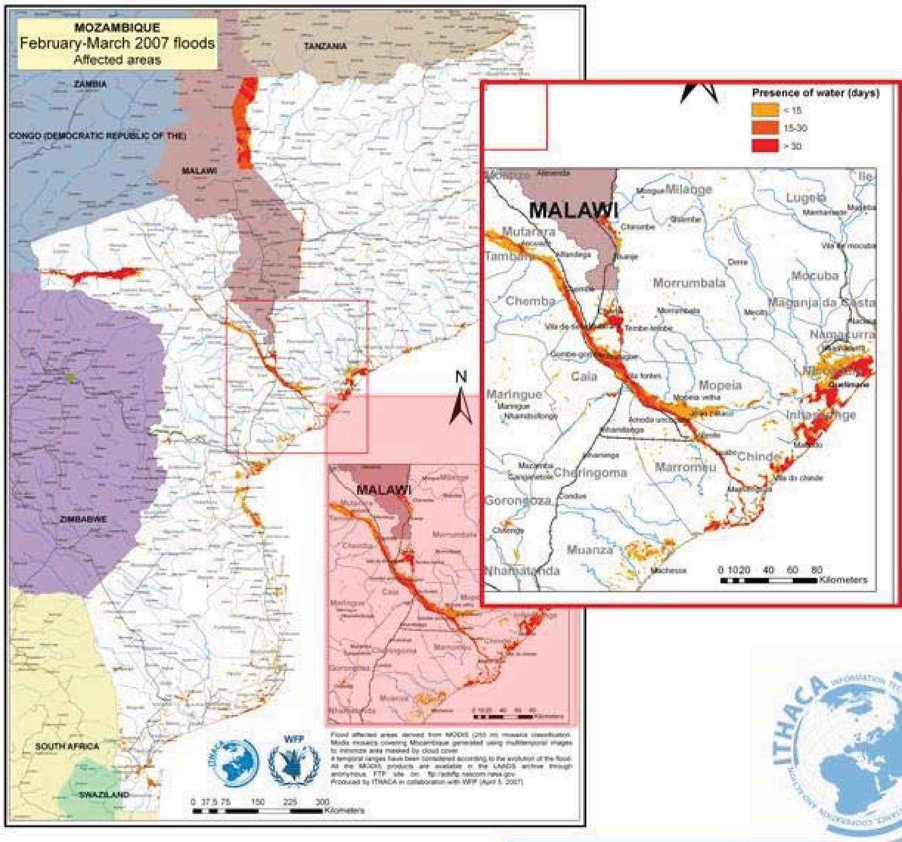


Figure 4.24 – Early Impact Unit: Map Sample Form N°6.

Early Impact Unit

MAP SAMPLE
N°7

Type of map / example of map title:
Flood monitoring map – presence of water and flood event evolution/ *"Mozambique, Zambeze flood monitoring. Zambezia e Sofala provinces. Modis satellite imagery analysis – 13 and 20 January 2008"*

Purpose and description:

Map showing:
-flooded areas and/or affected areas as detected from used satellite images (MODIS data or medium/high spatial resolution radar data, as specified) acquired in different dates.

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Administrative boudaries (National and sub-national level)
-Coastline
-Inland water data (point features, river/stream networks, and water bodies)
-If applicable, satellite derived data (flooded data and/or affected areas)
-Raster satellite background (data used for analyses).

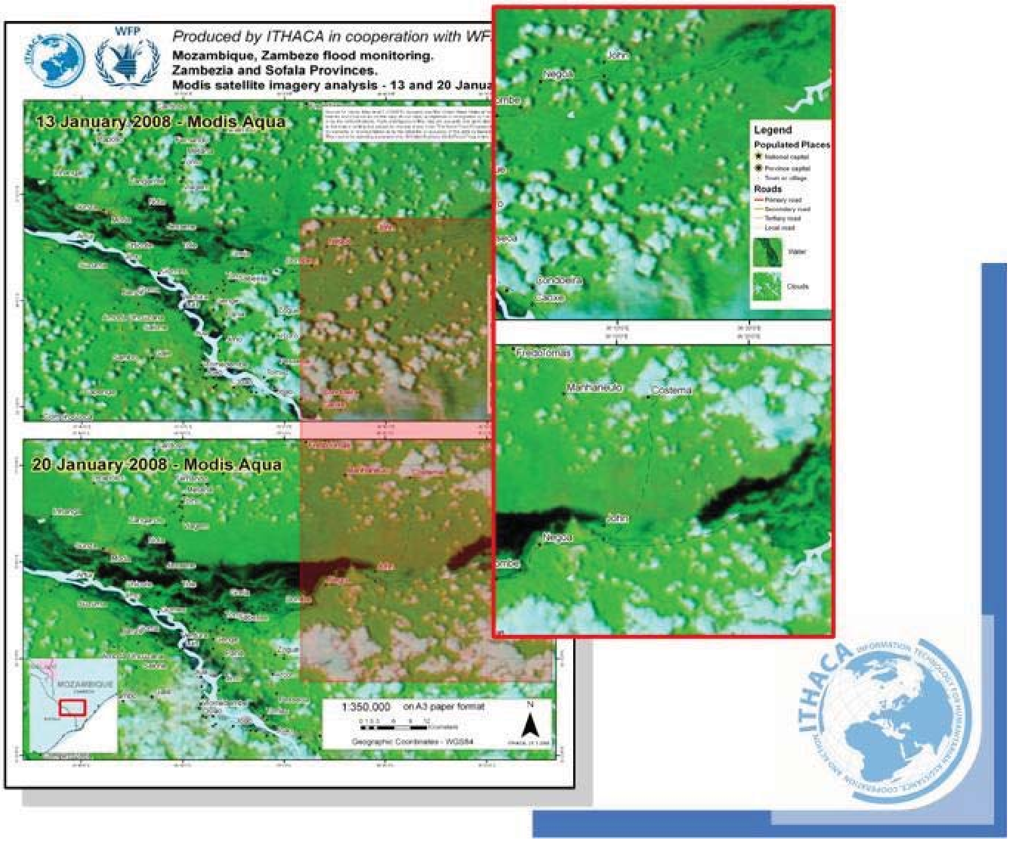


Figure 4.25 – Early Impact Unit: Map Sample Form N°7.

Early Impact Unit

MAP SAMPLE N°8

Type of map / example of map title:

Flood monitoring map – damages to infrastructures / *"Mozambique, Zambezi flood monitoring, Mutarara analysis "*

Purpose and description:

Map showing:
-damages to infrastructures, with particular attention to transportation networks
-identified points at risk of future damage
-flooded areas and affected areas as detected from used satellite imagery

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Satellite derived data (flooded data, affected areas, historical flooded areas)
-Damages assessment results (point, line or area features)
-Raster satellite background (data used for analyses)

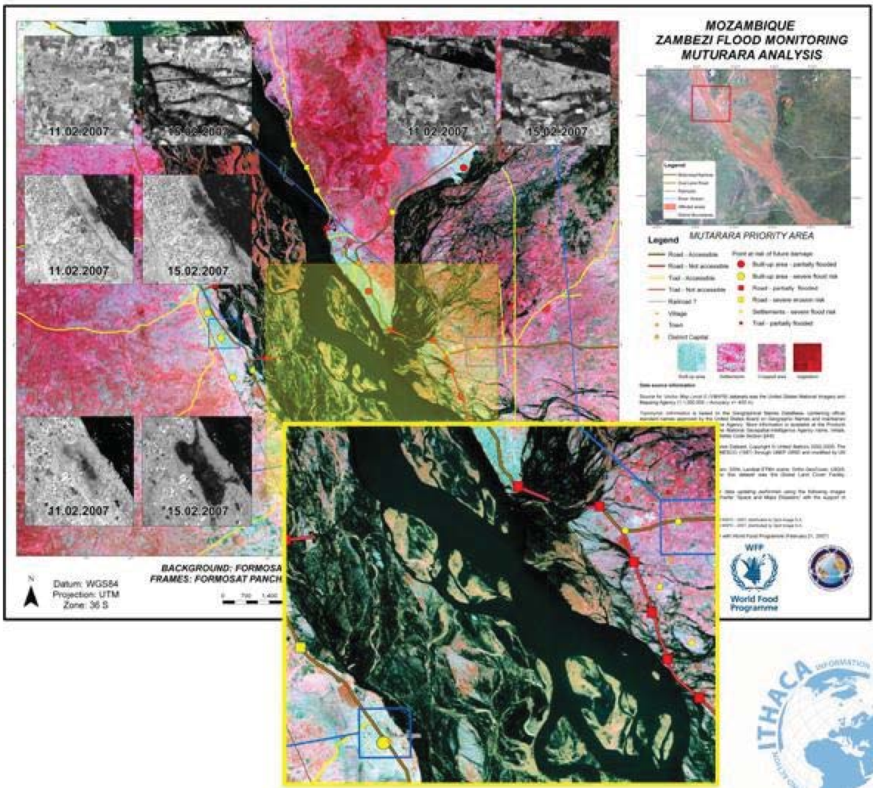


Figure 4.26 – Early Impact Unit: Map Sample Form N°8.

Early Impact Unit

MAP SAMPLE
N°9

Type of map / example of map title:

Flood monitoring map – geomorphologic interpretation / *"Mozambique, Zambezi flood monitoring, geomorphologic interpretation "*

Purpose and description:

Map showing:
-boundaries of major flood hazard zone
-sector of enhanced river bank erosion
-flooded areas and affected areas as detected from used satellite imagery

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-Digital Elevation Model
-Satellite derived data (flooded data, affected areas, historical flooded areas)

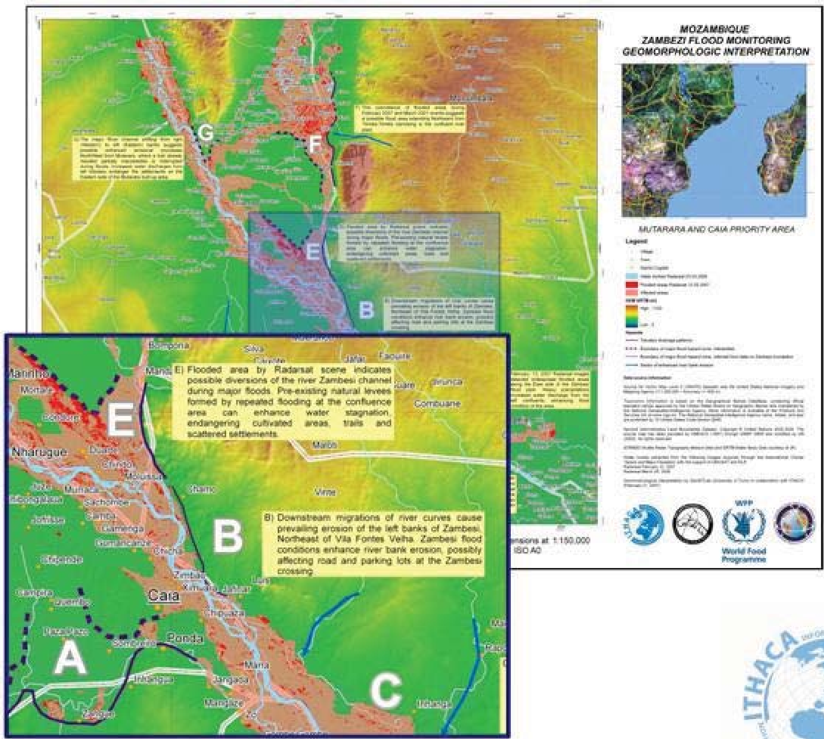


Figure 4.27 – Early Impact Unit: Map Sample Form N°9.

Early Impact Unit

MAP SAMPLE N°10

Type of map / example of map title:

Map product in support to *International Charter "Space and Major Disaster"* acquisitions/ *"Bangladesh Tropical Cyclone Sidr, November 2007 – Satellite acquisitions foootprints"*

Purpose and description:

Map showing:
-satellite acquisitions footprints
-if applicable, identified priority areas for satellite data acquisitions

Required data content:

-Populated places (National capitals, Province capitals, towns and villages)
-Roads (primary, secondary, tertiary and local roads)
-Built-up areas
-Administrative boundaries (National and sub-national level)
-Digital Elevation Model (DEM)
-Coastline
-Inland water data (point features, river/stream networks and water bodies)
-satellite acquisitions footprints

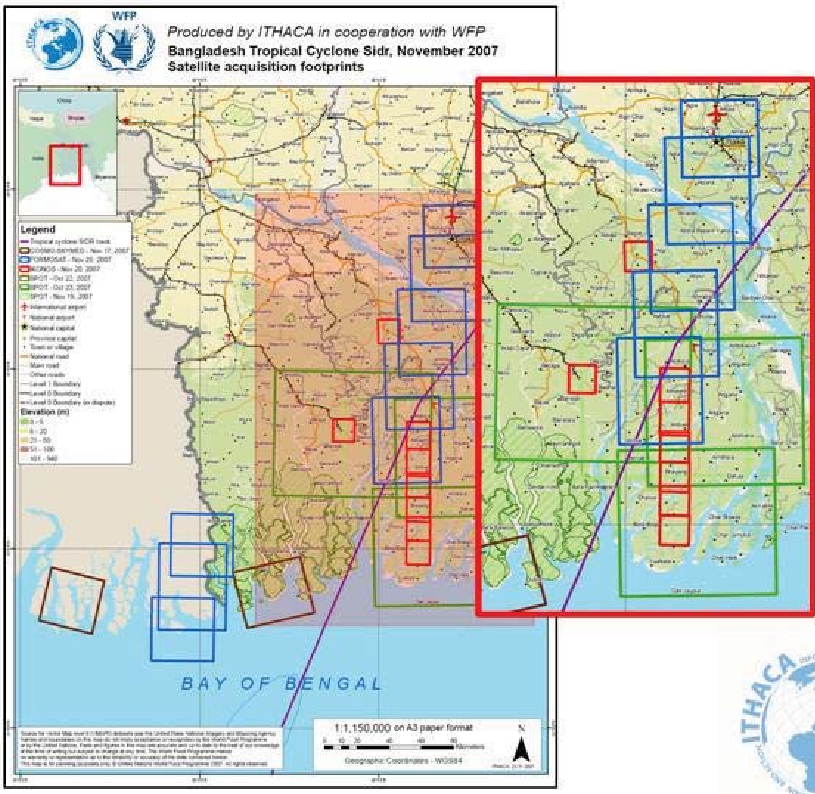


Figure 4.28 – Early Impact Unit: Map Sample Form N°10.

Early Impact Unit

DATA LIST
Map production sub-process

THEME	REQUIRED ENTITY \ MAP PRODUCT	Simulated floodable areas map	Flood monitoring map	Flood monitoring map – presence of water and flood event evolution	Flood monitoring map – damages to the infrastructures	Flood monitoring map – geomorphologic interpretation	Map products in support to *International Charter* acquisitions
Populated places	National capitals	R					
	Province capitals	R					
	Towns or villages	R					
	Built-up areas	R					
Roads	Primary roads	R					
	Secondary roads	R					
	Tertiary roads	R					
	Local roads	R					
Boundaries	Coastline	R					
	Administrative boundaries – national level	R					
	Administrative boundaries – sub-national level	R					
Hydrography	River/stream networks	R	R	R	O	R	R
	Water bodies	R	R	R	O	R	R
	Water point features (e.g. dams)	R	R	-	O	R	
	Discharge measurement stations	R	-	-	-	-	-
	Basins data	R	-	-	-	-	-
Elevation	Digital Elevation Model (DEM)	R	O	-	-	R	-
Population	Population data	-	O	-	-	-	-
Satellite base imagery used for analyses	Satellite base data	O	O	O	R	O	-
Satellite base imagery not used for analyses	Satellite backgrounds	O	O	O	-	O	-
Land cover	Land cover data	-	O	-	-	-	-
Thematic data produced by ITHACA Early Impact unit	Floodable areas	R	-	-	-	-	-
	Flooded areas	-	R	R	O	R	O
	Affected areas	-	R	R	O	R	O
	Wet areas	-	O	O	O	-	-
	Burnt areas	-	O	O	-	-	-
	Cloud cover	-	O	O	-	-	-
	Damages assessment results	-	-	-	R	-	-

R *necessary* O *optional* - *not required*

Figure 4.29 – Early Impact Unit, Map production process: required entities collections, summary.

Data list and conclusions

The final product of the *processes analysis* of all the Early Impact unit activities, carried out during the ITHACA needs assessment phase, has been the identification of all entities collections needed in order to perform the considered activities. These collections would have been necessarily included in the planned geographic database and, therefore, structured and modeled in subsequent conceptual and logical modeling steps.

As a matter of fact, each data store or archive, identified during the decomposition process of work-flows performed using Data Flow Diagrams, corresponds to a needed data collection or entity. Data List in Figure 4.30 summarizes all entities identified for the Early Impact unit.

Early Impact Unit

DATA LIST

THEME	REQUIRED ENTITY	SPATIAL TYPE	NOTES
Populated places	National capitals	*point*	
	Province capitals	*point*	
	Towns or villages	*point*	
	Built-up areas	*area*	
Roads	Primary roads	*line*	
	Secondary roads	*line*	
	Tertiary roads	*line*	
	Local roads	*line*	
Boundaries	Coastline	*line*	
	Administrative boundaries – national level	*line, area*	
	Administrative boundaries – sub-national level	*line, area*	
Hydrography	River/stream networks	*line*	
	Water bodies	*area*	
	Water point features (e.g. dams)	*point*	
	Discharge measurement stations	*point*	internally produced
	Basins data	*line, area*	
Elevation	Digital Elevation Model (DEM)	*raster*	
Population	Population data	*raster*	
Land cover	Land cover data	*raster*	
Thematic data produced by ITHACA Early Impact unit	Floodable areas	*area*	internally produced
	Flooded areas	*area*	internally produced
	Affected areas	*area*	internally produced
	Wet areas	*area*	internally produced
	Burnt areas	*area*	internally produced
	Cloud cover	*area*	internally produced
	Damages assessment results	*point, line, area*	internally produced
Maps produced by ITHACA Early Impact unit	ITHACA Map	*raster*	internally produced
Satellite base imagery used for analyses	Satellite base data	*raster*	internally produced
Satellite base imagery not used for analyses	Satellite backgrounds	*raster*	internally produced

Figure 4.30 – Early Impact Unit, comprehensive Data List.

It is noticed that collection of vector entities proposed in this table, in theory, could have different map scales, according to spatial resolution of used base satellite data and, therefore, final map scales. On the contrary, the choice of scales of vector data has been directed by the actual availability of data sources (see paragraph 4.5).

4.4.2 ITHACA's Snow Cover Unit

This unit manages mainly a project aimed at providing information on road accessibility, in the context of the emergency response operations undertaken by WFP. In the food supplying process, WFP often has to face troubles dealing with transportation and food aid distribution; it is therefore necessary to know, in near real-time, the conditions of the roads, tracks and trails that food trucks should run through. Afterward, the detection of inaccessible areas through appropriate monitoring activities plays a key role in a policy of prevention. Such information should be provided as regularly updated thematic maps.
The activity of this unit refers to the automatic production of snow cover vector data derived from satellite imagery. Snow mapping is required to verify accessibility in those zones which already faced food aid transportation troubles due to the presence of snow covers on the roads.
Two services are provided by this unit:

- *snow cover daily monitoring on specific areas of interest defined by the WFP (namely Nepal, Afghanistan, in Himalayan region);*
- *snow cover data automated production on request (global coverage).*

The features of the final snow cover information produced are:

- output in vector format to allow a simple and effective (distribution) transfer into the net;
- completely automatic vector generation;
- daily monitoring frequency;
- a maximum time delay of 48 hours from the request;
- worldwide coverage;
- low geometric resolution (500 m) allowing a regional view of the phenomenon (approximate map scale of 1:1000000);
- presence of metadata providing data reliability information.

In order to provide a daily snow cover vector mapping service, this unit defined and implemented an automatic procedure oriented to the extraction of snow cover vector data from base satellite derived data. Thanks to the collaboration with the Web Applications unit, a specific Web GIS application, based on Open Source software, has been implemented in order to distribute the final snow cover products to WFP users. This application is dedicated to the monitoring of some Himalayan areas specifically defined by the WFP. Through the implemented interface, snow cover data can be searched according to the desired date, displayed and downloaded in a vector format.
Figure 4.31 shows the general defined work-flow. As you can see, besides the Himalayan area monitoring application, this unit manages also a general purposes service on request. Interested user can send via FTP a request for a specific area of interest (worldwide coverage) and date. Needed satellite data are processed, then snow cover information is extracted and finally, final vector product is supplied to the user again through

the FTP. Using this service, users can obtain desired snow cover vector data via FTP connection in a short downtime (10'-30'), depending on the size of the area of interest.

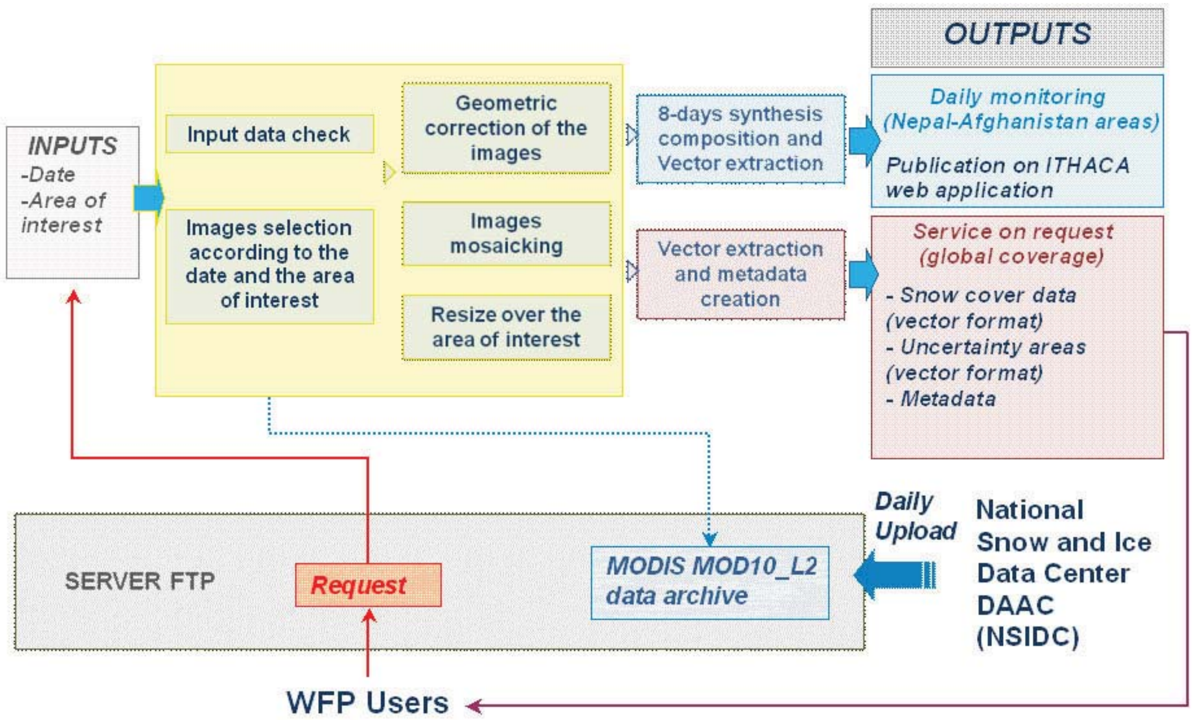


Figure 4.31 – Snow Cover Unit, snow cover vector mapping services.

In general, the proposed work-flow provides the desired output data through the following operations (Figure 4.31):

- selection of the base satellite data according to the chosen day and area;
- geometric correction of these data using a specially designed developed procedure;
- data mosaicking;
- if applicable, definition of 8-days synthesis composition, in order to avoid cloud cover problems;
- extraction of snow cover vector data (in the ESRI *shapefile* format);
- and generation of a metadata file indicating the cloud cover presence.

Each of these operations is implemented by a specific routine written in IDL (Interactive Data Language) and assembled to the others in order to create a unique and completely automatic procedure. In order to provide complete and automated services, in ITHACA a local computer hosting the above described procedure continuously updates a local archive of the base satellite data obtained via FTP from the *National Snow and Ice Data Center* (NSIDC) and periodically checks for new snow cover vector data requests.

The developed procedure requires as input satellite derived data obtained from the MODIS (*Moderate Resolution Imaging Spectroradiometer*) sensor (TERRA/MODIS *MOD10_L2 Snow Cover* product). The MODIS mission, besides having the above-mentioned features, grants a whole world coverage providing images and derived products that are completely free of charge. The *MOD10_L2* data have a geometric resolution of 500 m and a worldwide coverage. This product is distributed by NASA to users with a low

processing level (*Swath* format), which performs only a first geometric correction step. Therefore, in the *Swath* data the effects of panoramic distortion in the across-track and along-track directions (the latter distortion effect is called *bow-tie*) are not corrected. In order to use the MOD10_L2 product for the snow cover automatic map production purposes, the Snow Cover user group elaborated a specific routine for data geometric correction which performs two steps: the panoramic distortion correction and the georeferencing in the WGS84 datum.
For the Himalayan areas monitoring service, the final outputs of the elaboration, that is, the snow cover vector data, are then published in a web environment and exploited in a specific Web GIS application accessible through the ITHACA web-site (see Figures 4.32 and 4.33). The developed web application is based on MapServer and uses Ka Map and OpenLayers utilities.

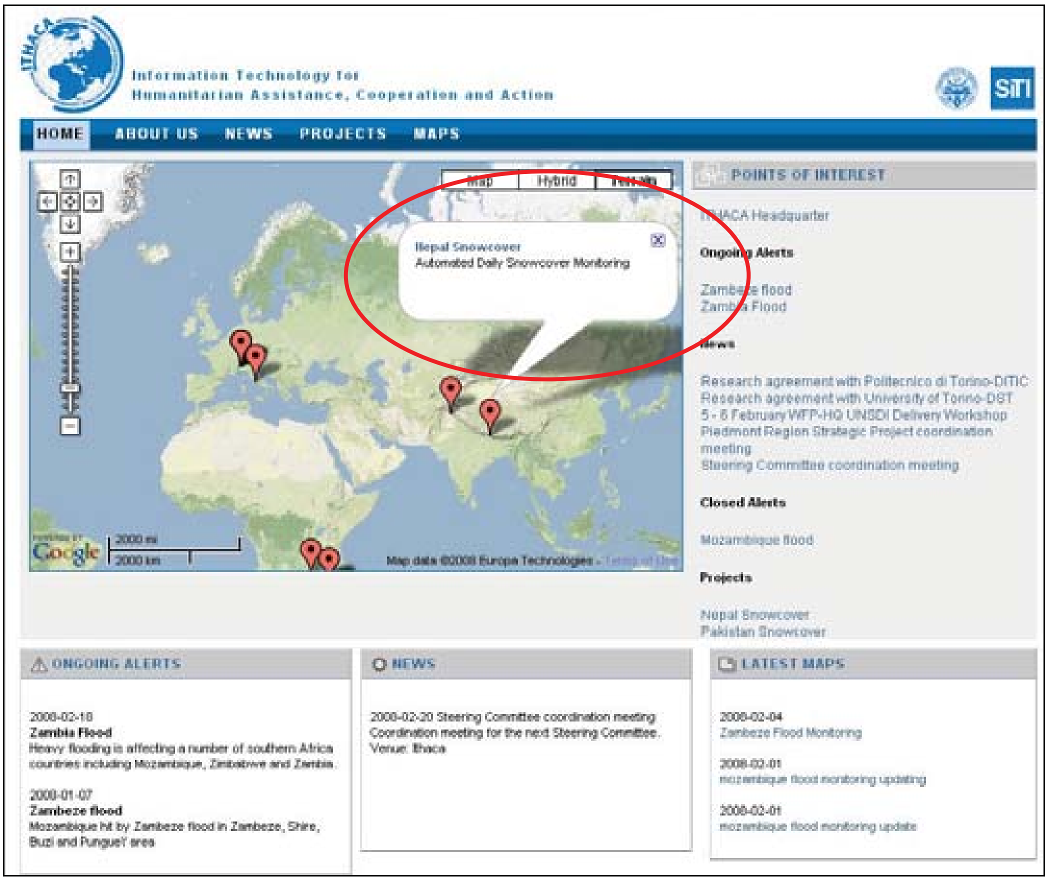


Figure 4.32 – ITHACA web-site home page, link to the snow cover monitoring application.

This application allows users to perform the following operations (Figure 4.34):

- to see the snow cover 8-days synthesis product superimposed on a general map of the area;
- to navigate the map (panning and zooming operations are available);
- to activate or deactivate base layers (map layers containing boundaries, roads, populated places, hydrography data or satellite background);
- to overlay snow covered roads (see Figure 4.34);
- to download snow cover vector data.

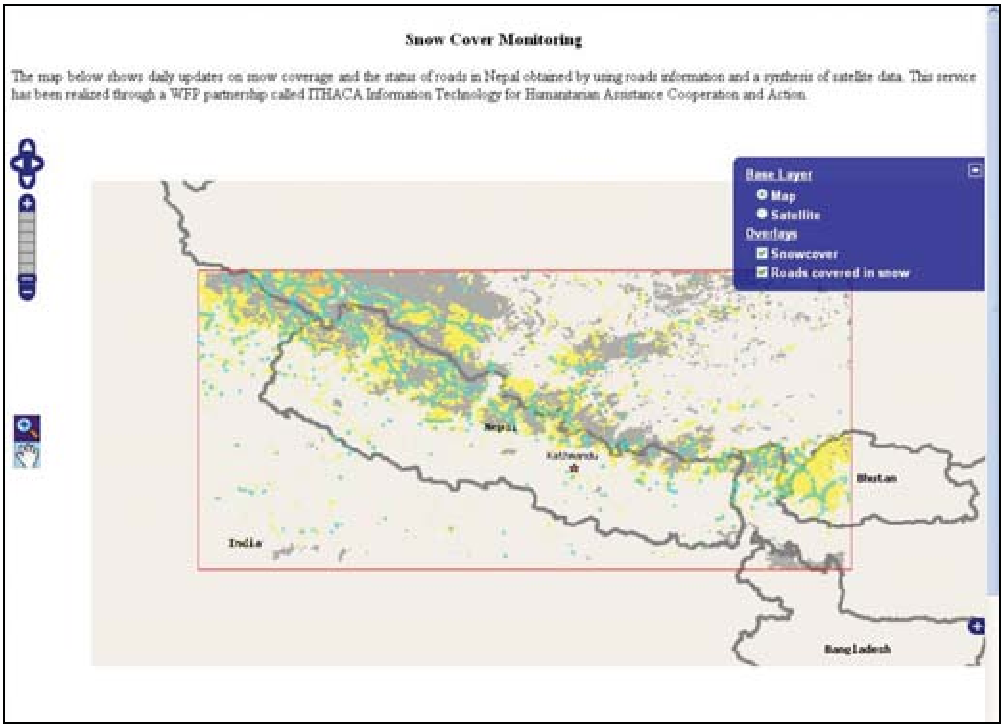


Figure 4.33 – ITHACA web-site, the snow cover monitoring web application for the Nepal area.

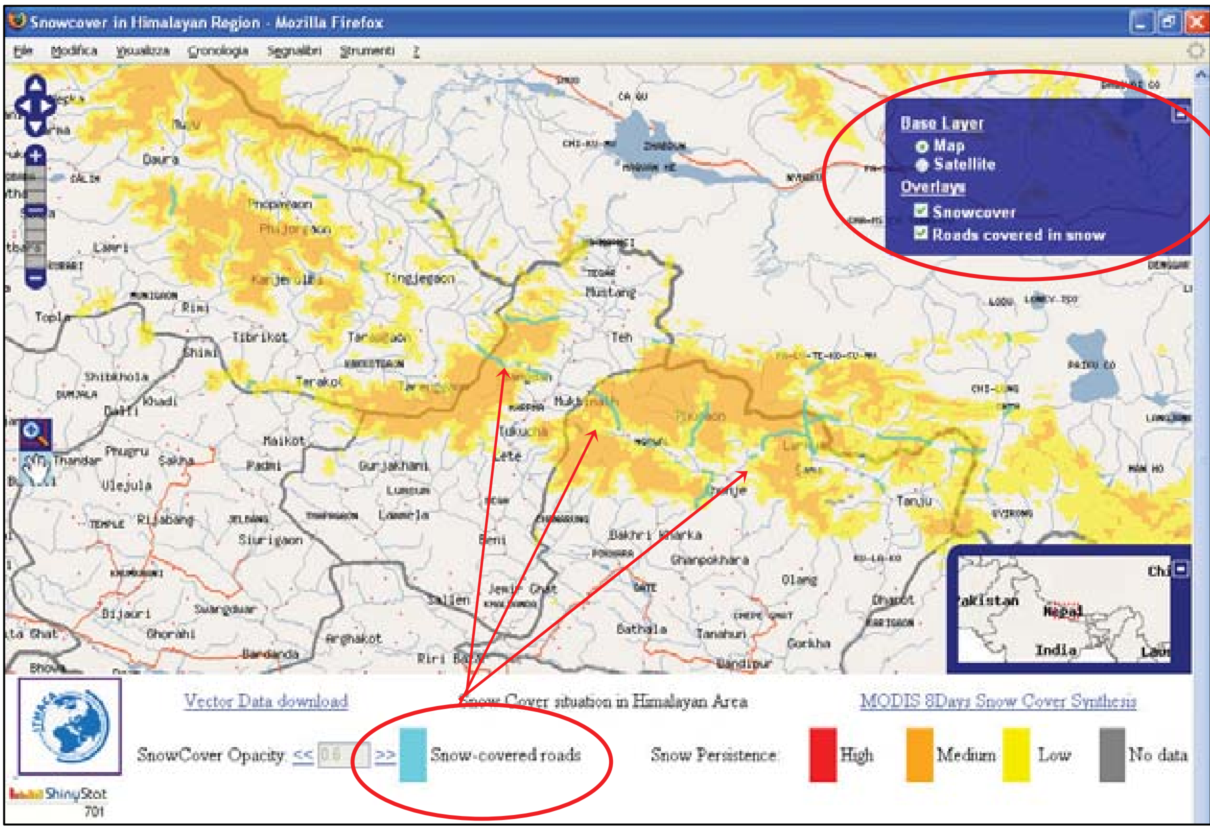


Figure 4.34 – ITHACA web-site, the snow cover monitoring web application: snow covered roads are highlighted.

Unit's needs assessment

In this phase a complete identification of all the processes carried out by this unit during its activities has been performed. As usually, DFD notations have been used.

First of all, Figure 4.35 shows the context diagram.

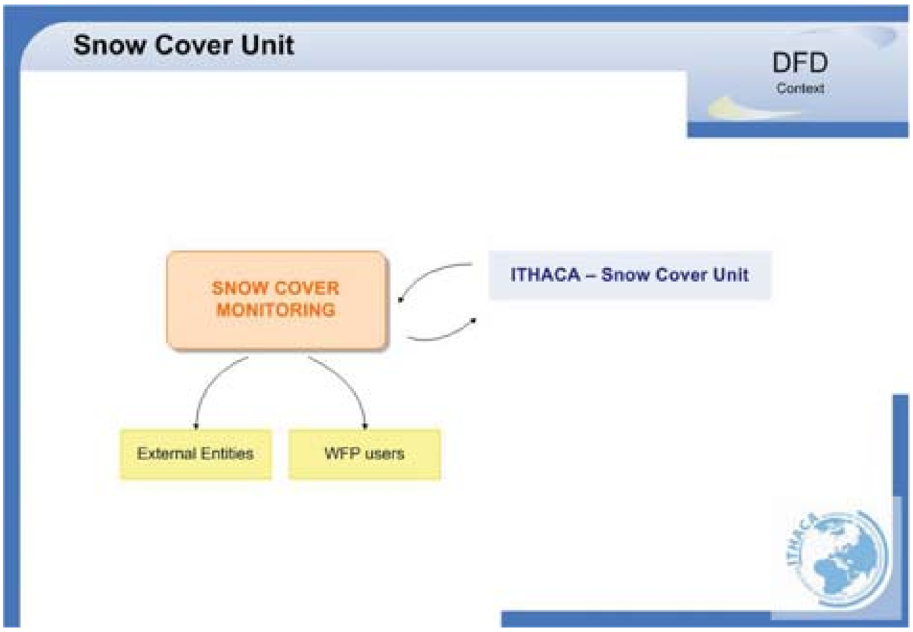


Figure 4.35 – Snow Cover Unit, Data Flow Diagram: the context diagram.

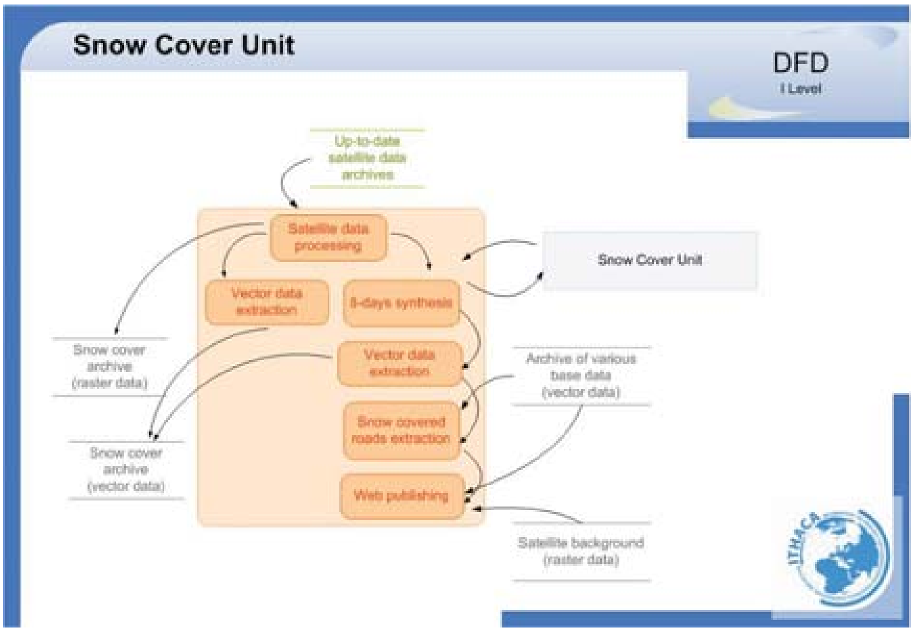


Figure 4.36 – Snow Cover Unit, Data Flow Diagram: identified operations, I Level .

The general "snow cover monitoring" process has been then decomposed in two sub-processes corresponding to the two different types of services supplied by this unit, previously described.

The sub-processes have been decomposed in some basic operations which require some data flows (Figure 4.36, I Level DFD). As you can see, some of these basic operations are common for the identified sub-processes. In order to correctly manage identified data flows, the following data stores or archives are used:

- *raster snow cover data archive*: this data store contains satellite derived snow cover data used in order to define 8-days composites. During the activities, this archive is thought as temporary. Therefore, in the planned geodatabase no permanent store of this kind of data has been considered necessary. Nevertheless, this topic definitely needs future considerations.
- *various vector base data archives*: these archives are needed for the web application development. They must contain almost administrative boundaries data (national and sub-national levels), roads data, hydrography data (river/stream networks and water bodies) and populated places. These data must be contained in the geodatabase in order to perform spatial analyses operations required by the web application .
- *satellite background*: raster background whith low spatial resolution, obtained usually from MODIS data. These satellite data must be contained in the geodatabase in order to allow the operations required by the web application.
- *vector snow cover data archive*: final outputs of the snow cover monitoring services are stored in order to allow future analyses, report generation activities, and the definition of an historical searchable archive. Nevertheless, the inclusion of this kind of data in the planned geodatabase has not been considered necessary. Of course, also this topic needs future considerations.

Data list and conclusions

The Data List proposed in Figure 4.37 summarizes all the entities collections or geographic data stores which may be included in the planned geodatabase in order to correctly support the activities of the Snow Cover Unit. Because of above-mentioned considerations, this defined Data List must be considered only as a first proposition, which requires further discussions.

Snow Cover Unit

DATA LIST

THEME	REQUIRED ENTITY	SPATIAL TYPE	NOTES
Populated places	National capitals	*point*	
	Province capitals	*point*	
	Towns or villages	*point*	
	Built-up areas	*area*	
Roads	Primary roads	*line*	
	Secondary roads	*line*	
	Tertiary roads	*line*	
	Local roads	*line*	
Boundaries	Coastline	*line*	
	Administrative boundaries – national level	*line, area*	
	Administrative boundaries – sub-national level	*line, area*	
Hydrography	River/stream networks	*line*	
	Water bodies	*area*	
	Water point features (e.g. dams)	*point*	
Satellite base imagery not used for analyses	Satellite background	*raster*	internally produced

Figure 4.37 – Snow Cover Unit, comprehensive Data List.

4.4.3 Analysis of other ITHACA's units

Early Warning Unit

When we started the needs assessment phase, no completely defined projects were established for this unit. Nevertheless, some research activities were ongoing, devoted to:

- a flood events forecast system definition. This activity aims at defining a flood events forecast system based on the identification of pluviometrical alert thresholds, defined using historical precipitation data, and on the definition of corresponding historical flood events scenarios. The operations of extraction of flooded areas hit by historical events should be carried out using satellite data archives, such as the MODIS one.
- dry weather evolution monitoring in sensitive areas and a drought forecast system definition.

In order to identify geographic data or data stores to include in the planned geodatabase needed by these activities, during the needs assessment phase several meetings and discussions with components of this user group have been conducted about future developments.

Particularly, the second project was in a very preliminary phase: research activities were in progress in order to define the effectiveness of the use of satellite derived vegetation index (NDVI, *Normalized Difference Vegetation Index*) historical series in the context of drought early warning and prevention activities.

In conclusion, the creation, in the planned geodatabase, of the following archives has been considered as certainly necessary, but further analyses are required:

- *historical flooded areas archive*: this archive contains all vector data extracted from satellite data acquired during historical flood events;
- *NDVI data archive*: this archive contains raster NDVI data, obtained from satellite images, which can be used in order to produce vegetation indices time series.

Figure 4.38 shows the preliminary Data List defined for this unit. This is only a first hypothesis, further discussions will be carried out following future researches of this user group.

Early Warning Unit

DATA LIST

THEME	REQUIRED ENTITY	SPATIAL TYPE	NOTES
Hydrography	River/stream networks	*line*	
	Water bodies	*area*	
	Water point features (e.g. dams)	*point*	
	Discharge measurement stations	*point*	internally produced
	Basins data	*line, area*	
Elevation	Digital Elevation Model (DEM)	*raster*	
Land cover	Land cover	*raster*	
Thematic data produced by ITHACA	Flooded areas	*area*	internally produced
	Affected areas	*area*	internally produced
	Wet areas	*area*	internally produced
Satellite base imagery used for analyses	Satellite base data	*raster*	internally produced
Vegetation	Satellite derived NDVI data	*raster*	internally produced

Figure 4.38 – Early Warning Unit, comprehensive Data List.

Web Applications Unit

The main aim of the activities carried out by this unit is the implementation of systems aimed at distributing and sharing georeferenced information for both early warning and early impact activities, by means of Web GIS applications, mainly based on commercial or Open Source (OS) platforms.

The research team developed some test applications. The main components under testing were OS tools such as Geoserver, Mapserver and Open Layers. The realized working solution referred to the on-line diffusion of Snow Cover project data outputs, already described. Another developed application referred to the visualization of GeoRSS (georeferenced feeds) from different sources in a single disaster map and a buffer analysis around a GeoRSS for affected population assessment. This application is accessible through ITHACA web-site, managed by this unit.

At present, an important topic faced by this unit is to contribute to the *Global Geodatabase/SDI project*, developing valuable web services and web applications oriented mainly to allow the access and the distribution of data contained in the geodatabase. In particular, a promising future application proposed by this unit is the definition of a valuable Early Impact Web GIS application, accessible through the ITHACA web-site, allowing the visualization and navigation of all data contained in the geodatabase. This application will be configured to allow users to select their desired viewing area and to select/deselect the desired datasets to be visualized. Moreover, through this application, ITHACA Early Impact activities results (vector data and raster maps products) will be published in a near real-time and easily downloaded by external interested users (mainly, WFP's users involved in humanitarian operations). Finally, historical data produced by ITHACA, contained in the geodatabase, may be searched and downloaded. Additionally, this application will provide ad hoc map production capabilities.

Other web applications and web services will be defined only after the compete implementation of the first version of the geodatabase, according to needs specifically expressed by WFP's users. They will be mainly devoted to:

- allow efficient and customized access to data derived from Early Warning and Early Impact analyses;
- supply not only geodatabase data access and consultation solutions, through a Web GIS data viewer, but also solutions for data editing, according to user's privileges.

For the purposes of the needs assessment, because of its nature, no data list has been defined for this unit. As a matter of fact, this unit works in supplying of services in order to distribute final outputs of all the other ITHACA's units. Therefore, geographic data used in its activities are the same needed by the units already analysed.

4.4.4 Final table and conclusions

According to the result of the analyses conducted in the needs assessment phase, a comprehensive Data List has been produced (Figure 4.39). This Data List contains all the entities collections, that is, all the entities which must be considered in the following database conceptual design phase in order to guarantee that the final implemented geodatabase will correctly answer to all the geographic data needs of ITHACA's activities.

Another important conclusion of this phase concerns the tools used by ITHACA's units during their activities. Like WFP's units, also ITHACA ones, in most of their operations, exploit GIS software from ESRI (ArcGIS, ArcInfo license). That confirms the need, already discussed, to develop an instrument compatible with the ESRI ArcGIS suite.

COMPREHENSIVE DATA LIST

THEME	REQUIRED ENTITY / USER GROUP	Early Impact unit	Early Warning unit	Snow Cover unit
Populated places	National capitals	R	-	R
	Province capitals	R	-	R
	Towns or villages	R	-	R
	Built-up areas	R	-	R
Roads	Primary roads	R	-	R
	Secondary roads	R	-	R
	Tertiary roads	R	-	R
	Local roads	R	-	R
Boundaries	Coastline	R	-	R
	Administrative boundaries – national level	R	-	R
	Administrative boundaries – sub-national level	R	-	R
Hydrography	River/stream networks	R	R	R
	Water bodies	R	R	R
	Water point features (e.g. dams)	R	R	-
	Discharge measurement stations	R	R	-
	Basins data	R	R	-
Elevation	Digital Elevation Model (DEM)	R	R	-
Population	Population data	R	-	-
Land cover	Land cover data	R	R	-
Thematic data produced by ITHACA Early Impact unit	Floodable areas	R	R	-
	Flooded areas	R	R	-
	Affected areas	R	R	-
	Wet areas	R	R	-
	Burnt areas	R	-	-
	Cloud cover	R	-	-
	Damages assessment results	R	-	-
Vegetation	Satellite derived NDVI data		R	
Maps produced by ITHACA Early Impact unit	ITHACA Maps	R	-	-
Satellite base imagery used for analyses	Satellite base data	R	R	-
Satellite base imagery not used for analyses	Satellite backgrounds	R	-	R

Figure 4.39 – ITHACA activities, comprehensive Data List.

4.5 Survey of available sources of data

Another important activity conducted in the needs assessment phase has been the identification of all potentially exploitable sources for geographic data to include in the planned geodatabase.
The comprehensive *Inventory Report* on core thematic datasets, prepared by UNGIWG TG 1 (see 4.3.1), has been considered as a starting point for this activity. This report presents an inventory of suitable global data sources for the *Core Data* layers proposed in the UNSDI project context. Tables 4.5 and 4.6 summarize the main findings of this inventory. The summary matrix in Table 4.5 ranks the major data sources identified for each CGDB core data layers. Short list of data sources is presented in Table 4.6.

Table 4.5 - Tabular Rating of Most Suitable Globally Identified Databases by Topical CGDB Core Data Reference. (source: UNGIWG documentation, 2005)

CGDB Data Layer	Highest Rated	Secondary Database	Tertiary Database	Current Recommended Database and Selection Criteria	Comments and Notes
Boundaries: Coastal, Administrative, and Areas of Special Interest					
Coastline and Maritime Boundaries	WVS+	UNCS/ VMap0	RWdB2	Use VMap0 – difficult to acquire and convert WVS+	As all three layers identified are from framework data libraries, it is more cost-efficient to process source library as a whole rather than individual layers
Country, Political, and Area of Dispute Boundaries	UNCS	WVS+	VMap0	Use VMap0 – difficult to acquire and convert WVS+/UNCS	See above note in reference to secondary and tertiary databases
Country and Subnational Boundary Data Layers	SALB	Vmap0	RWdB2	SALB if available (VMAP0 if not) – SALB does not yet have global coverage	See above notes
Areas of Conflict, and Landmine Dispersal	UNCS	SALB	VMap0	Use SALB – UNCS data cannot be fund	Other than the demarcation of areas in dispute, no robust area of conflict or landmine dispersal areas were identified
Parks, Conservancies, and Protected Areas	WDPA/ WCMC				Only globally consistent data source identified
Human Health: Boundaries and Facilities					
Human Health Infrastructure and Statistical Databases	EIP			Under development	No finalized sources of data were identified to support this topical CGDB reference category. However, WHO EIP data should be considered
Human Population: Population Centres and Distribution					
Population Centres and Census Databases	ALLM & Europa	CIESIN	VMap0	Use CIESIN. ALLM and Europa are commercial sources	
Population Density 2015	CIESIN/ FAO/CIAT			Best/only source	
Population Census and Distribution Databases	LANDSCAN	CIESIN	Europa	Use both LANDSCAN and CIESIN	Preferred database for either CIESIN or LANDSCAN must be based on purposed use of outputs
Rural Population density	FAO poverty				This entry is for info only as this

	mapping project				database still needs to be considered as UNGIWG CGDB
Transportation: Roads, Railways, Airports, Harbours, and Navigation/Routes					
Roads Databases	ADC	VMap0	RWdB2	Use VMap0 and VMap1 where available	VMAP1 is best but hasn't been completely released by governement into public domain. ADC is expensive.
Railway Line, Station, and Marshalling Yard Databases	VMap0	ALLM & ADC	RWdB2	Use VMap0	Any processing must be conducted in conjunction with relevant VMap population and roads data layers
Aiport Databases	ALLM	GNS, VMap0, RWdB2	Europa	Use GNS or VMap0	
Harbour Databases	Europa	ALLM	GNS, VMap0, RWdB2		
Navigation/Routes	n/e	n/e	n/e		No sources of digital data were identified to support this data layer
Bathymetry and Terrestrial Elevation					
Bathymetric Databases	GEBCO/ ETopo2	GNS	VMap0/ WVS+	Use Etopo2	Integration with terrestrial reference layers and creation of composite image backgrounds
Terrestrial Elevation	SRTM 3-arc second (90m)	SRTM 3-arc second/ GTopo30 (1km)	VMap0	Use VMap0	
Geo-Physical: Geology, Geo-Morphology, Seismic, and Hydro-Geology					
Geology or Minerals Databases	GlobalGIS	GNS	VMap0/ DCW	Use GlobalGIS	Only limited data identified
Geo-Morphology and Physiographic Databases	DCW/ VMap0	GNS	GlobalGIS	Use DCW	
Earthquake, Tsunamigenic and Volcanic Databases	Earthquake USGS-NEIC UN-GSHAP	Tsunamigenic NOAA-NGDC UNESCO/ IOC/ITSU	Volcanic Smithstonian GlobalGIS	Various. Sources identified need to be evaluated	
Hydro-Geological/Aquifer Databases	n/e	n/e	n/e		No sources of data were identified to support this data layer
Surface Hydrology: Drainage, Surface Waterbodies, Watersheds, and Water Points					
Drainage and Flow Routing Databases	VMap0 with reference to DCW	RWdB2	H1k	Use VMap0	
Surface Waterbody (SWB) Databases	VMap0 and potentially DCW	NASA-ESAD LANDSAT derivative contracted to EARTHSAT	RWdB2	Use VMap0	UNGIWG should pursue release into public domain of NASA-LANDSAT based data produced by EARTHSAT
Watershed and River Basin Databases	New SRTM based Effort	H1k	GIWA		
Water Points and Limnological Databases	WorldLakes & DCW/ VMap0	GNS	ICold Dams	Use GNS – GNS is large and commonly used	

Table 4.6 – Short List of Data Sources Identified (source: UNGIWG documentation, 2005)

Short List of Data Sources Identified	
Framework or Primary Database	**Acronym**
GEOnet Name Server Gazetteer ~1:250k	GNS
World Vector Shoreline Plus 3rd Edition	WVS+
VMap1 Data Library 1:250k (non-global)	VMap1
Digital Chart of the World 1:1m	DCW
VMap0 5th Edition 1:1m	VMap0
Relational World Databank II v 1.1	RWdB2
UNCS 1:10m & 1:5m Quick Impact Data	QID
UNCS-FAO Political & WHO-SALB Subnational Boundaries	SALB
WHO EIP/CSR Health Facilities	EIP
UNEP WDPA Parks and Protected Areas	WDPA
ALLM Gazetteer	ALLM
ADC WorldMap	ADC
EuropaTech Discovery	Europa
CIESIN GPW v. 3 and GRUMP	CIESIN
ORNL LandScan	ORNL
USGS GlobalGIS	G-GIS
ETOPO2 and/or GEBCO	ETopo2
GTopo30 and/or Globe	GTopo30
SRTM-GTopo30 & SRTM-3 arc second	SRTM
HYDRO 1 Kilometre Database	H1k
AVHRR - IGBP SPOT-GLC2000	GLC
NASA - LandSat Orthorectified Library	NASA-OLIL
EarthSat GeoCover ~1:250k	GeoCover
FAO World Lakes and Rivers Database	FAO-WorldLake
Global International Water Assessment watershed delineation	GIWA
NASA Earth Science Applications Directorate buy-in related to EARTHSAT GeoCover	NASA-ESAD
National Earthquake Information Center & Global Seismic Hazard Assessment Program	USGS-NEIC UN-GSHAP
NOAA - NGDC Tsunami Database UNESCO/IOC/ITSU Historical Tsunami Database	NOAA-NGDC UNESCO/IOC/ITSU

Important conclusions derived from analysis of the TG 1 report, and considered in this phase, are:

- publicly available global vector data libraries containing multiple and differentiated layers of information representing spatial features such as roads, rivers and populated places, were and are, at present, (according to the additionally researches we developed) available at only four broad scales: 1:1 million, 1:3 million, 1:5 million and 1:10 million;

- the Vector Smart Map Level 1 (VMap1) data library was recognized as the best baseline for vector data poverty mapping and emergency response. On the other hand, the U.S. National GeoSpatial-Intelligence Agency (NGA) Vector Smart Map Level 1 (VMap1) 1:250 000 source data library is not listed in the matrix in Table 4.5 due to the inconsistent global coverage available for this data source library. At present, this data source has not been completely released yet in the public domain. Anyway, it should be noted that for regional and country based areas of interest where the VMap1 provides complete coverage, this data library currently provides the highest resolution source of vector data available within the public domain.
- among the different core data layers at the global level, special priority must be given to:
  - international boundaries at 1:1, 1:5, 1:10 and 1:25 million;
  - *Second Administrative Level Boundaries* (*SALB*), including historic changes.

  In this field, the *SALB* initiative is indicated as a very important one. *SALB* is a UN project which has been launched in the context of the activities of UNGIWG and in the continuity of different efforts that took place in the middle of the 90's where the delimitation of the administrative boundaries was needed for the creation of population distribution grids.

  The *Second Administrative Level Boundaries* dataset aims to provide the UN community with information about administrative boundaries down to the second sub-national level. It is based on validated information received from the countries concerned and forms part of the UN geographic database. The developed *SALB* dataset is a global digital dataset consisting of digital maps and codes that can be downloaded on a country by country basis. This dataset is downloadable at no cost from the *SALB* website, but it is not recommended for use scales below 1:1,000,000.

Evidently, the conclusions proposed by the TG 1 report constituted an important point of reference for facing the design of a global geodatabase containing the Core Data needed for the WFP SDI project, conducted by ITHACA *Geodatabase/SDI Unit*. Nevertheless, the results of this report date back to the 2005, so it has been necessary to carry out further searches in order to define, for the chosen final data collections or spatial entities identified during the needs assessment phase, an updated list of data sources consistent with features described in the Preliminary Analysis.
Moreover, it has been necessary not only to consider what are the recommended databases proposed within the UNSDI project, but also to know the specific data sources commonly used by the WFP. Table 4.7 summarizes the data sources of the specific geographic data used in the different WFP's units involved, with their features. It can be noticed that only commonly used data or data globally consistent have been considered in this table.

Table 4.7 – Commonly used data sources identified for the WFP's ODAP and ODAV units. (source: ITHACA documentation, 2007)

Dataset	Description	Data type	Distribution
ODAP and ODAV-VAM units			
ADC Roads	Global database for roads, updated to year 2002. Commercial database produced and distributed by the American Digital Cartography Inc.	Spatial data type: vector (line) Global coverage: Y Map scale or resolution: 1:1,000,000 Features: Primary Roads, Secondary Roads, Tertiary Roads, Trails, Tracks	Source type: commercial data Access: restricted Distribution features: for internal usage only

GD 03	Global Discovery database produced by Europa Technology. This database refers to year 2003. The database includes the world coverage of the following layers: roads, railways, populated places, airports, ports, national boundaries, administrative boundaries first level, administrative boundaries second level, administrative boundaries third level, sea-depth contours, sprawls, linear drainage, polygon drainage, coast line, time zones, transport structure.	Spatial data type: vector (point, line, area) Global coverage: Y Map scale or resolution: 1:1,000,000 Features: various	Source type: commercial data Access: restricted Distribution features: for internal usage only
GD 03	HITOPO Digital Terrain Model raster tiles	Spatial data type: raster Global coverage: Y	Source type: commercial data Access: restricted Distribution features: restricted
GD 03	LOTOPO Digital Terrain Model raster	Spatial data type: raster Global coverage: Y	Source type: commercial data Access: restricted Distribution features: restricted
Landscan05	Global population density raster dataset updated to year 2005, produced and released by the Oak Ridge National Laboratory.	Spatial data type: raster Global coverage: Y Map scale or resolution: 1 km	Source type: Access: restricted Distribution features: restricted
CGIAR/ NASA Shuttle Radar Topographic Mission (SRTM)	The SRTM digital elevation data provided by CGIAR has been processed to fill NASA/USGS data voids. Distributed in 5 degrees x 5 degrees tiles.	Spatial data type: raster Global coverage: Y Map scale or resolution: 90 m	Source type: public domain data Access: free Distribution features: free
Global Land Cover	European Commission Joint Research Center - Global Land Cover 2000	Spatial data type: raster Global coverage: Y Map scale or resolution: 1 km	Source type: Access: free Distribution features: restricted
World Wildlife Fund - World Ecoregions	Data set available for Africa and Central America. Distributed by USGS. The ecoregions are defined as relatively large areas of land or water that share a large majority of their species, dynamics, and environmental conditions. There are nearly 900 ecoregions in the world.	Spatial data type: vector (area) Global coverage: N Map scale or resolution: 5,000,000 Features: ecoregions	Source type: Access: restricted Distribution features: restricted
VMAP0 - Land Use		Spatial data type: vector (point, line, area) Global coverage: Y Map scale or resolution: 1:1,000,000 Features: land cover point mine, quarry, physiography and other vegetation feature datasets	Source type: public domain Access: free Distribution features: free
VMAP1	Data for particular areas (Pakistan, Sudan, Turkmenistan, Afghanistan, Niger,...)	Spatial data type: vector (point, line, area) Global coverage: Y Map scale or resolution: 1:250,000 Features: various	Source type: public domain Access: free Distribution features: free
GAUL - Global Administrative Unit Layer	Administrative boundaries at level 0 (national), 1 and 2 at global scale, also levels 3 and 4 available for a few countries.	Spatial data type: vector (area) Global coverage: Y Features:unique coding system, unit name	Source type: Access: free Distribution features: restricted
Relational World Data Bank II	Similar to DCW and much in common. The dataset contains national boundaries, roads, rails, ports, airports, rivers, water bodies, populated places, etc.	Spatial data type: vector (point, line, area) Global coverage: Y Map scale or resolution: 1:1,000,000 Features: roads, rivers and populated places classified in 5 categories	Source type: Access: free Distribution features:
Global map of Irrigated Areas	Percentage of land equipped for irrigation	Spatial data type: raster Global coverage: Y Map scale or resolution: 10 km	Source type: Access: free Distribution features: restricted

Global Farming Systems	Defines broad farming system categories at a global scale	Spatial data type: vector (area) Global coverage: Y Map scale or resolution: Features*:* simple farming systems classification	Source type: Access: free Distribution features: restricted
Normalised Difference Vegetation Index (NDVI)	Measure of vegetation cover. Only available for Africa and part of Asia.	Spatial data type: raster Global coverage: N Map scale or resolution: 1 km	Source type: commercial data Access: restricted Distribution features: restricted
RFE Rainfall Estimates	Rainfall estimates based on NOAA imagery	Spatial data type: raster Global coverage: N Map scale or resolution: 8 km	Source type: Access: free Distribution features: restricted
WFP Programme Presence	Presence of WFP School Feeding programmes, Mother & Child health and nutrition programmes, HIV/AIDS programmes at subnational unit level.	Spatial data type: vector (area) Global coverage: Y Map scale or resolution: Features*:* presence or absence of programme activities	Source type: internally produced (ODAV) Access: free Distribution features: for internal usage only
WFP Facilities	Location of WFP offices and warehouses	Spatial data type: vector (point) Global coverage: Y Map scale or resolution: 1:3,000,000 Features*:* info regarding the facility	Source type: internally produced (ODAV/ODAP) Access: free Distribution features: for internal

Afterwards, further researches about updated globally consistent spatial data sources have been conducted in order to satisfy also ITHACA's geographic data needs, according to proposed final Data List (Figure 4.39). A final and complete list of all the identified geographic data sources, classified according to the adopted themes, is proposed in Table 4.8. These proposed data sources, which are the most updated available at time of the needs assessment phase, provide all geographic data that have been included in the planned global database and which have been considered in the subsequent design phases.

As already discussed, the object of the activities of research conducted by the author, and described in this thesis, is the design of a first version of a geodatabase which guarantees the correct support of basic activities performed either by WFP's interested units and ITHACA's user groups. Therefore, we initially worked on the definition of a base version of the geodatabase, containing only geographic data which can be considered as *Core Data* in all supported activities and communities of users. Considered data are globally consistent, well-established and tested too. These data are included in Table 4.8.

At present, the implemented first version of the geodatabase is subjected to a test phase, both in ITHACA and WFP's offices. After this phase, we expect this global spatial database will be enriched with other data used or produced by single users or needed in very specific activities.

Table 4.8 – Comprehensive list of data sources identified for the global geodatabase development. (source: ITHACA documentation, 2007)

BOUNDARIES
VMAP0 – Vector Map Level 0 Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
SALB – Second Administrative Level Boundaries Data source: UNGIWG Data geometry: vector Native format: shapefile Geographic area: near global Scale: 1:1000000
POPULATION
LandScan – LandScan Global Population Data source: Oak Ridge National Laboratory Data geometry: raster Native format: ESRI Grid Geographic area: world Spatial resolution: 1 km
INDUSTRY
VMAP0 – Vector Map Level 0 Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
UTILITIES
VMAP0 – Vector Map Level 0 Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
ELEVATION
VMAP0 – Vector Map Level 0 Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
DTM from Shuttle Radar Topography Mission (SRTM) Data source: NASA - JPL Data geometry: raster Native format: binary Geographic area: world Spatial resolution: 90 m
HYDROGRAPHY
VMAP0 – Vector Map Level 0 Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
STRM Water Body Data (SWBD) Data source: NASA - JPL Data geometry: vector Native format: shapefile Geographic area: near global
Drainage Basins Level 1, 2, 3 Data source: USGS – EROS Data geometry: vector Native format: shapefile Geographic area: world Scale: 1:5000000
PHYSIOGRAPHY - VEGETATION
VMAP0 – Vector Map Level 0 Data source: NIMA Data geometry: vector Native format: VPF Geographic area: world Scale: 1:1500000-1:750000
Orthorectified Landsat Thematic Mapper Mosaics Data source: Geo Community Data geometry: raster Native format: TIFF/Geotiff Geographic area: world Spatial resolution: 30m
GLC2000- Global Landcover Classification for the year 2000 Data source: JRC- IES Data geometry: raster Native format: ESRI Grid Geographic area: world Spatial resolution: 1 km
MODIS Land Imagery Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Spatial resolution: 250 m, 500 m, 1 km
MODIS NDVI Data Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Spatial resolution: 250 m, 500 m, 1 km
MODIS Land Cover Data Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Spatial resolution: 1 km, 5 km

<u>Landsat 7 ETM+ Ortho GeoCover Imagery</u> Data source: GLFC Data geometry: raster Native format: TIFF/Geotiff Geographic area: world Spatial 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
FLOOD EVENTS DATA - ITHACA EARLY IMPACT ACTIVITIES ARCHIVE
<u>MODIS Land Imagery</u> Data source: NASA EOS Data geometry: raster Native format: HDF Geographic area: world Spatial resolution: 250 m, 500 m, 1 km <u>Satellite Land Imagery (Charter acquisitions)</u> Data source: various Data geometry: raster Native format: various Geographic area: world Spatial resolution: various <u>Thematic data produced by ITHACA</u> Data source: ITHACA Data geometry: vector Native format: shapefile Geographic area: world Scale: according to the spatial resolution of used base satellite data <u>Maps produced by ITHACA</u> Data source: ITHACA Data geometry: raster Native format: jpeg Geographic area: world Map Scale: according to the spatial resolution of used base satellite data <u>Dartmouth flood events data</u> Data source: Dartmouth Flood Observatory (DFO) Data geometry: vector Geographic area: world
TRANSPORTATION
<u>ADC Roads</u> Data source: ADC World Map Data geometry: vector Geographic area: world Scale: 1:1000000 <u>UNJLC database</u>
OTHER
<u>DCW – Digital Chart of the World</u> Data source: ESRI for USDMA Data geometry: raster Geographic area: world Scale: 1:1000000

5

DATA MODELLING

The adopted methodology for the definition of conceptual and logical schemas for the global geographic database is here presented. Results obtained in either conceptual and logical design phases are also described.

Moreover, some basic concepts about conceptual and logical design phases are presented, with particular attention to the spatial nature of considered data.

5.1 Introduction

In this chapter, first we present general methodologies used for data modelling (with particular attention to the nature of geographic data), and then we describe the development of the conceptual model for the required geographic database, from the information collected during the needs assessment phase.
According to the analyses carried out during the needs assessment phase, we decided to implement the defined geographic database model as an ESRI ArcSDE Enterprise Geodatabase based on Oracle 10g as DBMS. This choice has been made because of the extensive use of ESRI (*Environmental Systems Research Institute*) ArcGIS Desktop applications either in ITHACA's user groups and in WFP's units.
In this paragraph some basic concepts about data modelling activity are presented. A general description of the ESRI Geodatabase data model is also given.

5.1.1 Data models

A *data model* is a formal definition of the data required in a database, according to the results of the needs assessment phase. The main purpose of the data model, and the process of specifying the model, is to ensure that the data was identified and described in a completely rigorous and unambiguous manner and that all the users agree on the data definitions. Moreover, a data model is a formal description of the database structure. This structure must allow the correct storing, management and sharing of required data, and the valuable exploitation of these data by all the users' applications defined within the organization. In the considered case, ITHACA and WFP's applications are mostly based on GIS instruments. So, the data model is the formal specification for the *entities*, their *attributes* and all the *relationships* between the entities, required for the GIS development.
The data model can take one of several forms. As a matter of fact, several data models have been proposed in literature, which are classifiable according to the concepts types they use in order to describe the database structure. *High level data models,* or *conceptual*, provide concepts which are very close to the users' way of perceiving the data, while *low level data models*, or *physical*, provide concepts which describe the details about the way data are stored in the computer.
As you will see in following paragraphs, conceptual data models use concepts as entity, attributes and relationships, and abstraction concepts as generalization, specialization, and categories. Physical data models describe the way data are stored in a computer disguised as files.
Moreover, between these two types of model, there is a class of *data model which can be implemented* (that is, the data models which are available in commercial database systems). These models provide concepts that can be understood by users, but that are not very far from the way data are organized in a computer.
Data model which can be implemented, are the most used in traditional commercial DBMS. They include the *relational data model* (very common) and other models very used in the past, such as the *reticular* or *hierarchical* ones. Moreover, *object data model* (ODMG) are a new class of data models which can be implemented at a high level; as a matter of fact, they are close to conceptual data models.

5.1.2 Modelling geographic data

What distinguishes geographic objects from all the others (i.e., the classical ones, such as a "car") is their position in space (actually, the fact that their position in space matters). Geographic data are *entities* which have a location. Particularly, geographic data include the location information and other information about the entity of interest. This information will be referred to as *attributes* of the spatial entities.
Therefore, a geographic, or spatial data, differs from other data that are included in databases in how entities are defined and in the relationships between entities. In particular, entity identification (Figure 5.1) for spatial data includes the definition of a *physical* or *abstract entity* (e.g., a building) and the definition of a corresponding *spatial entity* (i.e., a polygon to represent the building footprint).
This latter, or second entity does not exist for other types of computer databases. The existence of the corresponding spatial entity is one of the major factors that distinguishes GIS from other types of systems. That is why it is very important to utilize proper planning and design techniques when building a geographic database.

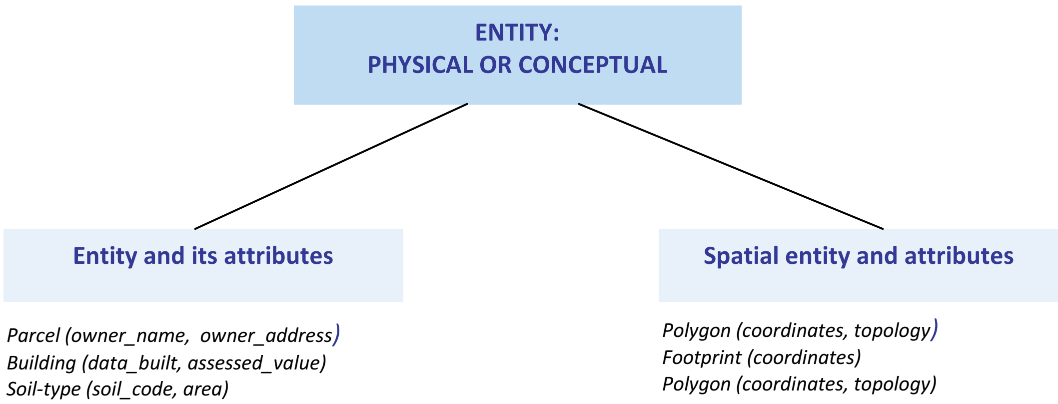


Figure 5.1 - Examples of conceptual entities and their corresponding spatial entities. (source: [UNESCO, "Training Module on the Applications of Geographic Information Systems (GIS)"])

Modelling is a standard activity in a database design process. Typical modelling methodologies, used in this field, also apply in spatial database design process, but for this they must be extended in order to correctly model geographic objects and their relationships. As a matter of fact, there are some spatial concepts of geographic applications that distinguish them from the classical ones in terms of semantics and modelling needs:

- entities in a geographic database are georeferenced. This implies that information about *shape* and *position* of an entity must be provided and managed. This information can be thought as additional attributes of the entity.
- as aforementioned, geographic data differ from the regular ones in how entities are defined. In particular, in the modeling phase, it is necessary to identify, for each considered entity, the geometric primitive suitable for its spatial representation. As previously described in Chapter 3 (paragraph 3.1.3), object/field models are used in the conceptual modeling phase, while vector/raster models can be defined at the implementation level.

- additionally, geographic objects are related to each other in space. Relationships among geographic objects are actually conditions on objects' position and are called *spatial relationships*. Spatial relationships are visually obvious when data are presented in the graphical form; however, it is difficult to define spatial relationships at the data model level. Spatial relationships can be organized into three categories (see also Chapter 3, par. 3.1.3):
    - *topological relationships*: describe the connectivity, orientation, containment and adjacency relationship among spatial objects. These relationships are invariable under topological transformations such as translation, scaling, and rotation. Conceptual modeling should lead to straightforward solutions for explicitly storing topology in the logical and physical levels, a common practice despite topology being derivable from objects' positions;
    - *direction relationships*: for instance, they include "above, below, or north of, southwest of";
    - *metric relationships*: are those relationship that are described in terms of distances and directions. An example is "within a distance of 2000 m". These descriptions depend on the absolute positions of objects relative to a given reference system. These relationships are normally computed from the database using the objects coordinates. So, in the data modelling process the focus is on the topological relationships.

An important issue related with geographic objects is the topological codification. In most GIS systems some relationships among spatial entities can be topologically coded in the geodatabase. Therefore, data within the database can be stored in a topological form using topological constraints or in non-topological form (in this case, relationships between coordinates of two entities are not formally coded within the database).

Topological consistency constraints are related to the topological/geometric elements of objects in the vector data model. As aforementioned in Chapter 3, in GIS, three types of geometric objects (nodes, arcs and polygons) are used in order to define five topological constraints: (a) every arc is bounded by two nodes; (b) for every arc, there are two polygons (left and right polygon);(c) every polygon is bounded by a closed cycle of nodes and arcs; (d) arcs intersect only in nodes; (e) every node is surrounded by a closed cycle of arcs and polygons. These constraints relate to the topological elements of objects and may be used to define a topological data model. All five conditions must hold for a space partition into area features, and for a network analysis, only the first and second rules are necessary.

In conclusion, during the geographic data modelling process, a fundamental issue is related to the use of particular data models, which are different to those used for general database modelling. In spatial data modelling, traditional conceptual and logical data model concepts are extended in order to adapt to the different types of considered spatial objects. Moreover, particular physical structures should be also considered in order to manage these types of data.

In particular, it is necessary to introduce in the conceptual model some new concepts in order to:

- define the spatial component of entities (shape and position);
- model spatial features of data;
- model objects according to object/field models;
- model objects with different spatial representations;

- define a particular organization of the data in a thematic way;
- differentiate spatial and not-spatial relationships and model topological relationships.

As we will see in the next paragraph 5.2, common used conceptual models, which extend existing data models, are the EER model (*Enhanced Entity-Relationship* or *Extended ER*) and the UML model extended with the introduction of spatial classes and types (e.g. ArcInfo UML model).
Finally, in the following section 5.1.3, the ESRI Geodatabase spatial model will be presented.

5.1.3 Introduction to the ESRI Geodatabase model

As aforementioned, according to the analyses carried out during the needs assessment phase, to implement a system compatible with the ESRI ArcInfo environment was a priority. Therefore, we decided to implement an ESRI ArcSDE Enterprise Geodatabase.
In this section a brief introduction to the ESRI ArcGIS software and to the ESRI Geodatabase model is proposed. This introduction is necessary in order to better understand the next conceptual design phase.

ArcGIS Desktop

ArcGIS Desktop is a comprehensive set of professional GIS applications of the *Environmental Systems Research Institute* (ESRI). It is a commercial GIS software which constitutes a platform for GIS users, allowing to manage complex GIS workflows and projects and to build data, maps, models, and applications. This platform involves a suite of applications including ArcCatalog, ArcMap, ArcGlobe, ArcToolbox, and ModelBuilder. Using these applications and interfaces, users can perform many GIS tasks, from simple to advanced.
ArcGIS Desktop is available at three functional levels:

- *ArcView* focuses on comprehensive data use, mapping, and analysis;
- *ArcEditor* includes advanced editing capabilities for shapefiles and geodatabases in addition to the full functionality of ArcView. ArcEditor also includes the ability to administer and use ArcSDE geodatabases;
- *ArcInfo* is a complete, professional GIS desktop containing comprehensive GIS functionality, including rich geoprocessing tools.

Additional capabilities can be added to all seats through a series of ArcGIS Desktop extension products from ESRI and other organizations. Users can also develop their own custom extensions to ArcGIS Desktop by working with *ArcObjects*, the ArcGIS software component library. Figure 5.2 shows a list of some of the key ArcInfo capabilities.

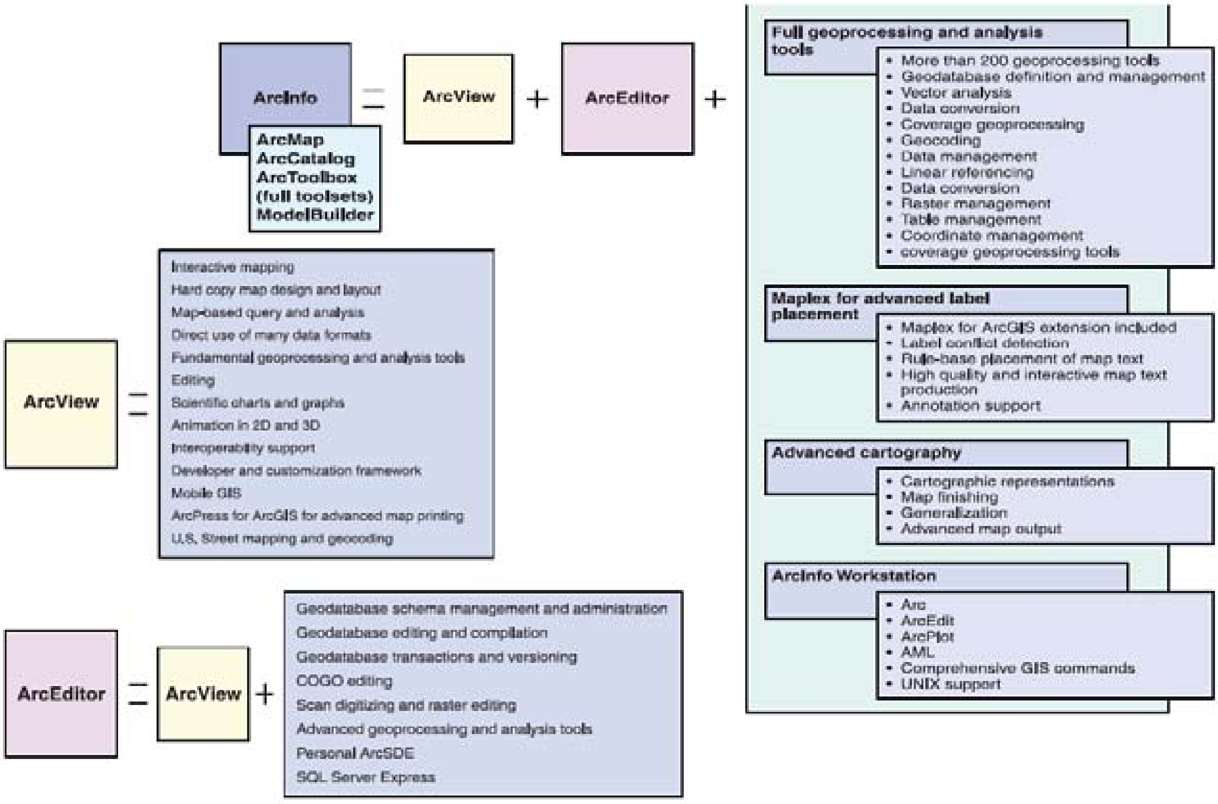


Figure 5.2 – Some of the key ArcInfo capabilities. (source: [ESRI, ArcGIS Desktop Help, 2006])

ESRI ArcGIS Geodatabase

With the recent release of its integrated ArcGIS software (ArcInfo 8), ESRI introduces a new object-oriented data model called the *geodatabase data model.* This new model is implemented as extension to the standard relational model by integrating it with object-oriented concepts in a manner that allows geographic objects to be modeled with their behaviours attached to them.

The geodatabase data model brings a physical data model closer to its logical data model. The data objects in a geodatabase are mostly the same objects defined in a logical data model, such as owners, buildings, parcels, and roads.

Further, the geodatabase data model allows to implement the majority of custom behaviour without writing code. Most behaviour is implemented through domains, validation rules, and other functions of the framework provided in ArcInfo. Writing software code is only necessary for the more specialized behaviours of features.

ArcGIS Geodatabase content

An ArcGIS Geodatabase is a collection of geographic datasets of various types held in a common file system folder, a Microsoft Access database, or a multi-user relational database (such as Oracle, Microsoft SQL Server, or IBM DB2).

The ArcGIS geodatabase can contain four types of *geographic representations* (that is, the basic ways to model data in a GIS):

- vector data for representing features (collections of points, lines, and polygons);

- attributes;
- raster data for representing imagery and gridded thematic data;
- Triangulated Irregular Networks (TINs) for representing surfaces.

The total GIS behaviour for representing and managing geographic information is based on these fundamental types. A geodatabase stores all of these representations of geographic data in a commercial relational database.

Features - Points, lines, and polygons

Geographic features are representations of things located on the surface of the earth. Geographic features can occur naturally (such as rivers and vegetation), can be constructions (such as roads, pipelines, wells, and buildings), or subdivisions of land (such as counties or political divisions). Many of the features in the world have well defined shapes. Vector data (Figure 5.3) represent the shapes of features precisely and compactly as an ordered set of coordinates with associated attributes. This representation supports geometric operations such as calculating length and area, identifying overlaps and intersections, and finding other features which are adjacent or nearby.

Vector data can be classified by dimension:

- *points* are zero-dimensional shapes that represent geographic features too small to be depicted as lines or areas. Points are stored as a single x,y coordinate with attributes.
- *lines* are one-dimensional shapes that represent geographic features too narrow to be depicted as areas. Lines are stored as a series of ordered x,y coordinates with attributes. The segments of a line can be straight, circular, elliptical, or splined.
- *areas* are two-dimensional shapes that represent broad geographic features stored as a series of segments that enclose an area. These segments form a set of closed areas.

Another type of vector data in a geodatabase is *annotation*. These are descriptive labels associated with features to display names and attributes.

Vector data in a geodatabase has a structure that directs the storage of features by their dimension and relationships.

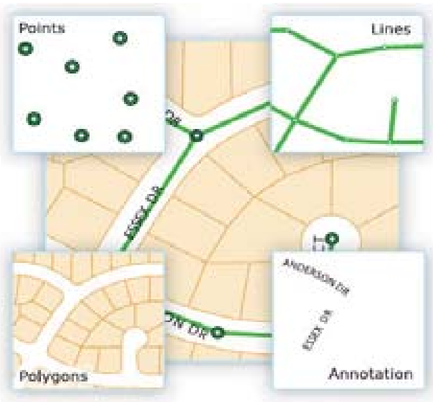


Figure 5.3 – Vector data in a geodatabase. (source: [ESRI, ArcGIS Desktop Help, 2006])

Attributes

In a GIS, descriptive attributes are managed and organized in tables, which are based on a series of simple, essential relational database concepts.

Raster data

Much of the data collected in a geodatabase is in gridded form. This is because cameras and imaging systems record data as pixel values in a two dimensional grid, or raster.

A pixel is a cell element of a raster and its values can depict a variety of data. A cell can store the reflectance of light for part of the spectrum, a thematic attribute such as vegetative type, or surface value, or elevation (Figure 5.4). In addition to the map projection, the coordinate system for a raster dataset includes its cell size and a reference coordinate (usually the upper left corner of the grid). These properties enable a raster dataset to be described by a series of cell values starting in the upper left row. Each cell location can be automatically located using the reference coordinate, the cell size, and the number of rows and columns.

Typical instances of raster data in a geodatabase are: satellite or aerial imagery, and scanned maps or pictures.

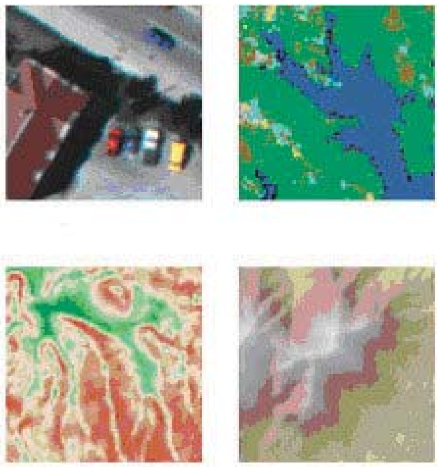


Figure 5.4 – Examples of raster data (up: scanned aerial photo and land use data; bottom: slope and elevation data). (source: [ESRI, ArcGIS Desktop Help, 2006])

Triangulated Irregular Networks (TINs)

A triangulated irregular network (TIN) is a model of a surface. A geodatabase stores TINs as an integrated set of nodes with elevations and triangles with edges through which an elevation (or z value) can be interpolated for any point within the geographic extent of a TIN.

Geographic representations are organized in the geodatabase in series of *thematic layers* or *themes*. A thematic layer is a collection of common geographic elements, such as a road network, a collection of parcel boundaries, soil types or satellite imagery for a certain date. The concept of a thematic layer was one of the early notions in GIS.

Each GIS will contain multiple themes for a common geographic area. The collection of themes acts as layers in a stack. Each theme can be managed as an information set independent of other themes. Because

the various independent themes are spatially referenced, they overlay one another and can be combined in a common map display.
GIS *datasets* are collections of representations. Like the layers in a map, GIS datasets represent collections of individual features with their geographic locations and shapes as well as with descriptive information stored as attributes. Many themes are represented by a single collection of homogeneous features, such as soil types, while other themes, such as transportation framework, are represented by multiple datasets, such as streets, intersections, bridges or highway ramps. These data collections can be organized as feature classes and raster-based data layers in the geodatabase.
More specifically, a geodatabase implements the vector data representation with *feature datasets* and *feature classes*, the raster data representation with *raster datasets*. A feature dataset is the container of spatial entities (*features*) and the *relationships* between them. Non-spatial entities (*objects*) are stored in *tables*. *Topological associations* are represented with geometric networks and planar graphs. Finally, a geodatabase also stores *validation rules* and *domains* to ensure that when features are created or updated, their attributes remain valid in the context of related features and objects.
All geodatabase elements are managed in standard DBMS tables using standard SQL data types and adhere to the ISO/OGC Simple Features specifications.
Creating a collection of these dataset types is the first step in designing and building a geodatabase. Users typically start by building a number of the fundamental dataset types. Then they add to or extend their geodatabase with more advanced capabilities (such as by adding topologies, networks, or subtypes) to model GIS behaviour, maintain data integrity, and work with an important set of spatial relationships.

Here is a brief introduction to some of the more common ArcGIS geodatabase elements used in design phase:

- ***Feature classes***. A feature class is a collection of features representing the same geographic elements. All the features in a feature class have the same spatial representation (for example, point, line, or polygon) and share a common set of descriptive attributes. Individual features in a feature class can also share spatial relationships with other features. There are two categories of feature classes:
    - *simple feature classes* contain points, lines, polygons or annotation without any topological associations among them. That is, points in one feature class may be coincident with, but distinct from, the endpoint of lines in another feature class. These features can be edited independently of each other.
    - *topological feature classes* are bound within a graph, which is an object that binds a set of feature classes that comprise an integrated topological unit. ArcInfo 8 introduce the geometric networks as a type of graph in a geodatabase. *Networks* are comprised of two fundamental components, *edges* and *junctions* (Figure 5.5). There are two types of junction and edge features: *simple* and *complex*.

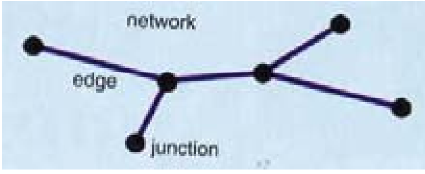


Figure 5.5 – Networks and its components. (source: [ESRI, ArcGIS Desktop Help, 2006])

In the database, a feature class is a table with a shape field containing point, line, or polygon geometries for geographic features. Each row of this table is a feature.

- ***Feature datasets***. Feature datasets are organized collections of related feature classes. Feature classes are organized in integrated feature datasets for many purposes, primarily to manage spatial relationships among related feature classes. Simple feature classes can be organized inside or outside of feature datasets, but topological feature classes must be contained within a feature dataset to ensure a common coordinate system. Moreover, feature classes in a feature dataset share a common coordinate system.

  There are additional situations in which users apply feature datasets in their geodatabases: to organize thematically related feature classes (sometimes, users will organize a collection of feature classes for a common theme into a single feature dataset); to organize data access based on database privileges (sometimes, users organize data access privileges using feature datasets; all feature classes contained within a feature dataset have the same access privileges); to organize feature classes for data sharing.

- ***Raster datasets and raster catalogs***. A raster dataset stores a single raster. Many rasters may be combined, with more recent overlapping cells replacing previous cells, into a continuous raster, often called a mosaic. A raster dataset is a single row table in the geodatabase.

  A raster catalog is a container of raster datasets, which can tile, overlap, or be of any irregular coverage. A raster catalog is organized as multirow table in the geodatabase where each row is a raster. Raster catalogs are used for massive data collections and for raster time series management.

- ***Object classes***. An object class is a ***table*** in a geodatabase. Object classes keep descriptive information about objects that are related to geographic features, but are not features on a map. An example of an object class is “owners” of land parcels. It is possible to establish a database join between a polygon feature class for land parcels and an object class for owners.

- ***Relationship classes***. A relationship class is a table that stores relationship between features or objects in two feature classes or tables. Relationships model dependencies between objects and allow to control what happens to an object when its related object is removed or changed.

  Attribute relationships are widely used in GIS, just as they are in all relational database management system applications. They define how rows in one table are associated with rows in another table. Relationships have a direction of cardinality (one-to-one, one-to-many, or many-to-many) and other properties.

- ***Domains***. Domains are constraints on attributes. These represent the list or range of valid values for attributes columns of a feature or object class. These rules control software behaviuor to maintain data integrity in certain attribute columns. With the definition of coded values domains it is possible to ensure, during the editing operations, that the attributes definitely have one of the expected values (during editing operations, invalid attribute values are highlighted).

- ***Subtypes***. Objects in an object class and features in a feature class may be further subdivided into subtypes. A subtype is a special attribute that lets to assign distinct simple behaviours for different classifications of objects or features. All subtypes of a class share the same set of attributes.

  Moreover, each subtype can have its own set of default values. Each subtype can also have its own range or coded attribute domain for a given field. Each subtype may also have different connectivity, relationship, or topology rules associated with it.

  An important geodatabase design issue arises when we must decide where it is appropriate to use subtypes and where additional feature classes are required. It is recommended to create separate subtypes for a single feature class or table when we are trying to distinguish objects by their default values, attribute domains, connectivity rules, and relationship rules. Otherwise, when we want to distinguish objects based on different behaviours, attributes, access privileges, or whether the objects are multiversioned, it is necessary to create additional feature classes.

- ***Spatial rules***. Spatial rules, such as topologies and their properties, are used to model how features share geometry with other features. This is a critical and widely used GIS mechanism to enforce certain spatial behaviuor and integrity in GIS databases.

  As previously discussed, topology can be considered a special type of relationship among features and it is used in a GIS to ensure the integrity of spatial relationships among features that share geometry. In particular, a GIS topology is a set of rules and behaviours that model how points, lines, and polygons share coincident geometry. Some examples of topology rules are: - adjacent features, such as two counties, will have a common boundary between them (that is, they "share" this edge); - the set of county polygons within each state must completely cover the state polygon and share edges with the state boundary.

  In earlier GIS systems, topology has been traditionally implemented as a data structure. The ArcInfo coverage model is an example. Advances in software development now permit a new and improved implementation of topology as a collection of feature classes with ranks and topology rules.

  There are two types of topologies in ArcGIS.

  Geodatabase topologies are stored in the geodatabase as a set of ranked feature classes and a defined set of topology rules. There are many *topology rules* which can be implemented in a geodatabase, depending on the valid spatial relationships between features that are most important to maintain for the organization and users. Some topology rules govern the relationships of features within a given feature class, while others govern the relationships between features in two different feature classes (see Figure 5.6). Topology rules can also be defined between subtypes of features in one or another feature class.

  Many topology rules can be imposed on features in a geodatabase. A well-designed geodatabase will have only those topology rules that define key spatial relationships needed by the organization. When a topology is edited, a topological graph composed of edges, nodes, and vertices is used to control geometric editing and discover feature errors.

  Map topologies are temporary, defined for the duration of an editing session, and allow quick shared-edge editing. A map topology allows to simultaneously edit features that overlap or touch each other using dedicated tools.

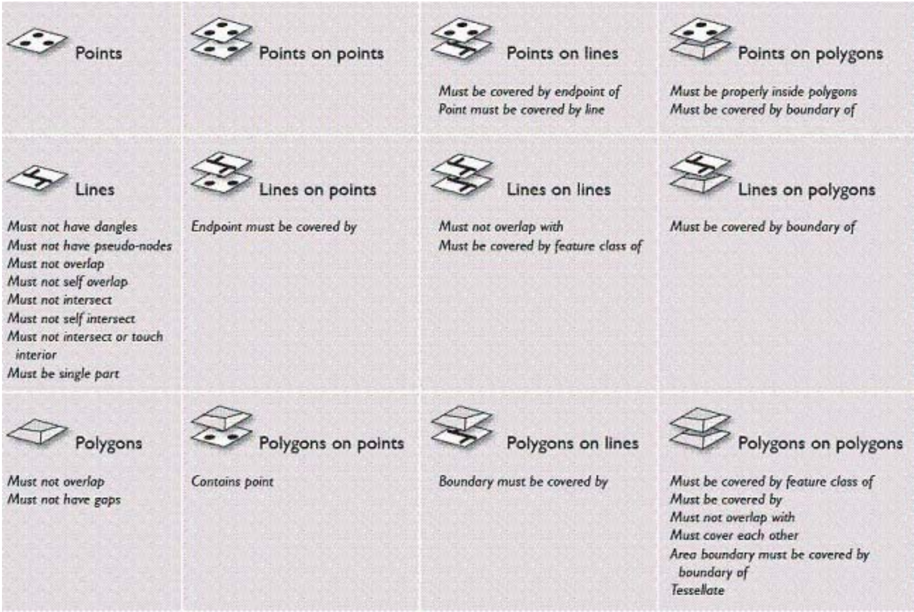


Figure 5.6 – A conceptual view of spatial relationships and integrity rules that can be managed using a topology. Each of these situations defines a potential case for using topology rules to maintain data integrity. (source:[ESRI, ArcGIS Desktop Help, 2006])

- ***Map layers***. A critical part of each dataset is the specification for how it is symbolized and rendered in maps. Map layers are typically defined as layer properties, which specify how features are assigned map symbology (color, fill patterns, line and point symbols) and labeling specifications.

Geodatabase storage in tables

All three primary datasets in the geodatabase (feature classes, attribute tables, and raster datasets) as well as other geodatabase elements are stored using tables. The spatial representations in geographic datasets are stored as either vector features or as raster. These geometries are stored and managed in attribute columns along with traditional tabular attribute fields.

As you can see in Figure 5.7, in the geodatabase, each feature class is managed in a single table. A Shape column in each row is used to hold the geometry or shape of each feature.

In the feature class table:

- individual features are held as rows;
- feature attributes are recorded in columns;
- the Shape column holds each feature's geometry (point, line, polygon, and so forth);
- the Object ID column holds the unique identifier for each feature.

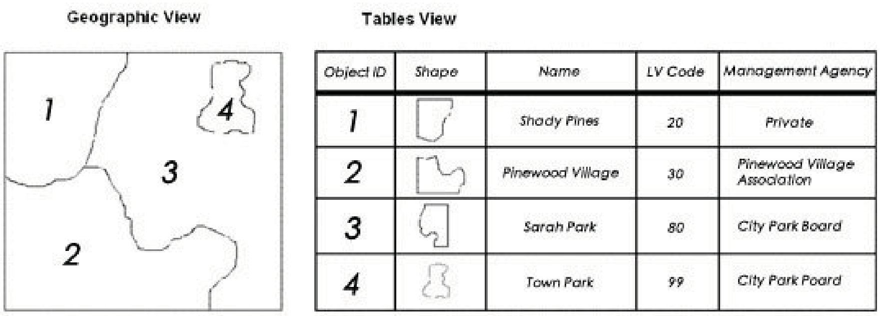


Figure 5.7 – Feature classes are stored as tables with a shape column in the geodatabase. (source: [ESRI, ArcGIS Desktop Help, 2006])

Moreover, as aforementioned, various geodatabase elements are used to extend simple tables, features, and rasters to add rich behaviour, data integrity, and data management capabilities. The geodatabase schema includes the definitions, integrity rules, and behaviour for each of these extended capabilities. These include properties for coordinate systems, coordinate resolution, topologies, networks, raster catalogs, relationships, domains, and so forth. This schema information is persisted in a collection of geodatabase meta tables in the DBMS. These tables define the integrity and behaviour of the geographic information. Further information will be provided in 5.3.2.

SQL is very strong at query and set processing of rows in tables, and the geodatabase strategy is designed to leverage these capabilities. The ArcInfo geodatabase supports SQL access to feature geometry in the following DBMSs:

- Oracle with or without Oracle Spatial;
- IBM DB2;
- IBM Informix.

The underlying SQL API (*application programming interface*) is based on the ISO SQL/MM Spatial and OGC's simple feature SQL specifications, which extend SQL with standards for vector geometry types. The API approach allows interactions with database within an application software, namely the ArcGIS system, using a database functions library.

Types of geodatabases

The geodatabase comes in two variants, namely, personal and multi-user geodatabases. The personal geodatabase is implemented on the Microsoft Jet Engine, which stores data in Microsoft Access database, and is built into the ArcInfo software. The personal geodatabase is limited to maximum storage capacity of 2 gigabytes and is therefore suitable for project-oriented GIS. Personal geodatabase supports single editing and several simultaneous viewers.

For large enterprise databases, ArcInfo provides a multi-user data access extension called the ArcSDE. ArcSDE provides the gateway between the ArcInfo software and the DBMS to share and manage the spatial data as tables. It allows remote access to spatial data and allows many concurrent editing of the same database. In ArcSDE geodatabases, there is no limit to size or numbers of users. ArcSDE supports many relational and object-relational DBMSs, including Oracle, Microsoft SQL Server, IBM DB2, and Informix.

5.2 Conceptual database modelling

5.2.1 Basic concepts

The general database design process, described in Chapter 3, requires three phases (see Figure 3.8). The first design phase, the conceptual one, refers to “what” must be represented in a database, while the other phases refer to “how” to do it.

The main output of the conceptual design phase is a *conceptual schema* which refers to a *conceptual data model*. The conceptual schema must be complete, that is, must contain all data needed to meet the system’s objectives, and must be directly translatable into the logical and physical database schema. Data identification and description include to define the objects (*entities*), the *relationships* between the objects and the *attributes* of the objects that will be represented in the database. As aforementioned, conceptual models allow to describe data organization at a high level of abstraction, without considering implementation aspects. Therefore, the conceptual schema is free from the physical structure of the database. This means it should be independent from the software and object storage techniques. This makes possible a change at the physical data level without involving any modification of the conceptual schema.

The purpose of the conceptual data modelling process is to prepare an unambiguous and rigorous description of the data to be included in the database in a form that:

- is understandable by the proposed users of the database or system;
- is sufficiently structured for a programmer or analyst to design the data files and implement data processing routines to operate on the data.

Therefore, the emphasis is on:

- communication between users and the programmer/analyst;
- review and verification of the data model and database design by both users and analyst. Description of the conceptual schema diagram can provide a useful tool which grant better communication among database users, and programmers/analysts. Because high level data models are based on easily comprehensible concepts, communication is correct and simple.

Moreover, in the design phase it is necessary to use a conceptual data model with the following features:

- *expressivity*: the data model must allow to distinguish different types of data, relationships and constraints;
- *simplicity*: the data model must be understood also by not expert users;
- *minimalism*: the data model must be based on few basic concepts, which are clear and unmistakable;
- *diagrammatical representation*: the data model must have a diagrammatic notation, useful to visualize the conceptual schema;
- *formality*: the conceptual schema, expressed in the data model, must represent an unambiguous formal specification of data.

Conceptual modelling is performed using an appropriate conceptual formalism (Figure 5.8). Therefore, the conceptual schema is represented through a formal language of description which supplies both suitable

syntax and graphical notation. Moreover, for each conceptual formalism, different languages for the description of the schema can be used, which are consistent with the formalism.
The formalism supplies concepts, elements and rules which can be used in the reality modelling process. The resulting schema is defined according to grammar rules of the adopted language of description.

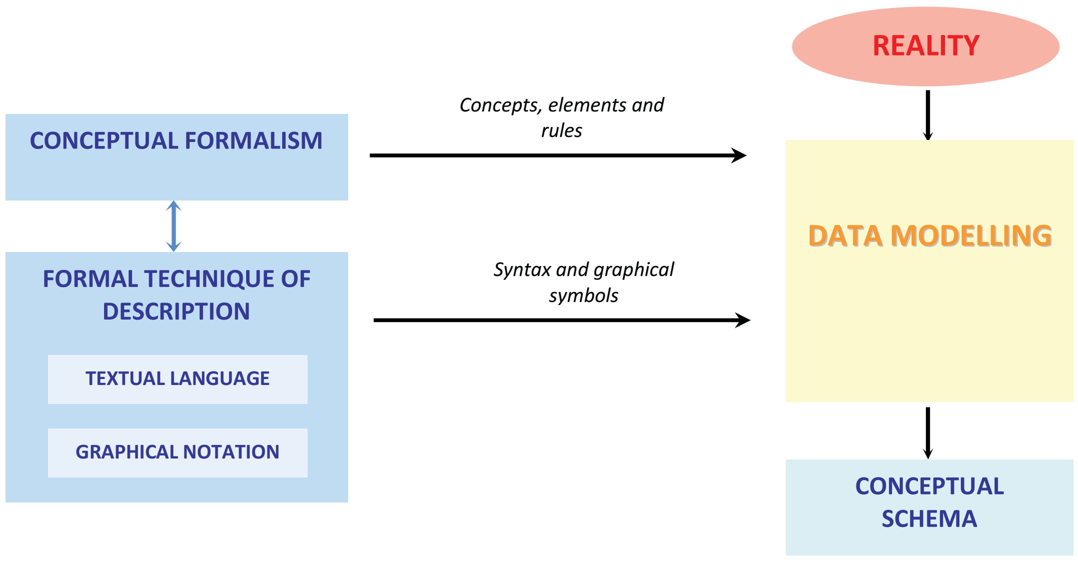


Figure 5.8 – Conceptual modelling process.

Two different approaches are commonly adopted in order to carry out the conceptual modelling phase for a geographic database:

- the use of existing data models, such as *Entity-Relationship model* (ER) and *Object-Oriented* (OO), in order to represent spatial objects;
- the extension of these models.

At the moment, there are many conceptual data models that are suitable to GIS and which are based on existent common formalisms. In addition to the aforementioned ER and OO formalism extensions, some extensions of IFO (*Is-a relationships, Functional relationships, complex Objects*) formalisms are proposed too. Finally, many specific formalisms which allow to capture the peculiarities of GIS were proposed, in order to model the spatial dimension and correctly manage geographic data, according to requirements proposed in 5.1.2. Some examples are the GeoER and the GeoIFO models. For an introduction to this issue, please see [T. Hadzilacos, N. Tryfona].
Following, in this paragraph, principal modelling concepts of the ER model and its diagram notation will be presented. Object modelling methodologies, such as UML (*Unified Modelling Language*) will be also discussed. Particular attention will be paid to the UML *Class Diagram*, that is very similar to ER diagrams and that has been used to define geodatabase conceptual model. Description of other data model formalisms is out of the scope of this document.

Entity-Relationship and Extended Entity-Relationship models

The Entity-Relationship (ER) modeling technique was developed by Chen (1976). The ER model is a conceptual model that supply many structures, also called *constructions*, which allow to describe the reality in an easy way, regardless of data organization in the computer. These constructions are used in order to define the schema which describe the data structure and organization. Moreover, each construction has its specific graphical representation. These representations are used in order to produce an ER schema through a clear diagram. Therefore, the ER technique is also a graphical method of representing objects of a database, all important relationships between the objects, and all attributes of either objects or relationships which must be captured in the database.

Main constructions of this model are:

- *Entities*. These represent classes of objects (facts, things, people,...) that have properties in common and an autonomous existence. “City”, “Department”, “Employee”, “Purchase” and “Sale” are examples of entities. An *instance* of an entity represents an object in the class represented by the entity. “Stockholm”, “Helsinki”, are examples of instances of the entity “City”.
- *Relationships.* They represent logical links between two or more entities. “Residence” is an example of a relationship that can exist between the entities “City” and “Employee”, while “Exam” is an example of a relationship that can exist between the entities “Student” and “Course”. An instance of a relationship is an n-tuple made up of instances of entities, one for each of the entities involved.
- *Attributes.* These describe the elementary properties of entities or relationships. For example, “Surname”, “Salary” and “Age” are possible attributes of the “Employee” entity, while “Date” and “Mark” are possible attributes for the relationship “Exam” between “Student” and “Course”. An attribute associates with each instance of an entity (or relationship) a value belonging to a set known as the domain of the attribute. The domain contains the admissible values for the attribute. It is sometimes convenient to group attributes of the same entity or relationship that have closely connected meanings or uses. Such groupings are called *composite attributes*.
- ***Cardinalities.*** These are specified for each entity participating in a relationship and describe the maximum and minimum number of relationship occurrences in which an entity occurrence can participate. Cardinalities state how many times an entity instance can participate in instances of a given relationship.

  In principle, a cardinality is any pair of non-negative integers (n,m) such as n≤m or a pair of the form (n, N) where N means “any number”. If minimum cardinality is 0, we say that entity participation in a relationship is *optional*. If minimum cardinality is 1, we say that entity participation in a relationship is *mandatory*. If maximum cardinality is 1, each instance of the entity is associated at most with a single instance of the relationship; if maximum cardinality is N, then each instance of the entity is associated with an arbitrary number of instances of the relationship.
- ***Identifiers.*** Identifiers (or *keys*) consist of one or more attributes which identify uniquely instances of an entity. In many cases, an identifier is formed by one or more attributes of the entity itself: in this case we talk about an *internal* identifier. Sometimes, however, the attributes of an entity are not sufficient to identify its instances unambiguously and other entities are involved in the identification. Identifiers of this type are called *external* identifiers. An identifier for a relationship consists of identifiers for all the entities it relates.

These basic ER modelling concepts are enough to represent database schemas for many traditional applications. Otherwise, in other applications, such as GIS, it is necessary to define more complex database schema, in order to comprehend different data features and useful constraints. Therefore, in order to correctly model GIS databases it is necessary to incorporate some other modeling concepts in usual conceptual models, such as the ER one. Features of the *Extended-ER* (*EER*) are here briefly presented.

EER model contains all the ER modeling concepts and, additionally, the *subclass* and *superclass* concepts, with related *specialization* and *generalization* concepts. Another considered concept is the *union type* one, or *category*, which is used in order to represent an objects collection defined by the union of objects that belong to different entities types. The *inheritance* mechanism is closely related.

<u>*Generalization*</u> represents a logical link between an entity E, known as *parent entity* (also *superclass*), and one or more entities $E_1,..., E_n$ called *child entities* (also *subclasses*), of which E is more general, in the sense that they are a particular case. In this situation we say that E is a *generalization* of $E_1,..., E_n$ and that the entities $E_1,..., E_n$ are *specializations* of E. Every instance of a child entity is also an instance of the parent entity. Every property of the parent entity (attribute, identifier, relationship or other generalization) is also a property of a child entity. This property of generalizations is known as *inheritance*.

A generalization is *total* if every instance of the parent entity is also an instance of one of its children, otherwise it is *partial*. A generalization is *exclusive* if every instance of the parent entity is at most an instance of one of the children, otherwise it is *overlapping*. The generalization “Person” of “Man” and “Woman” is total (the sets of men and women constitute ‘all’ the people) and exclusive (a person is either a man or a woman). The generalization “Vehicle” of “Automobile” and “Bicycle” is partial and exclusive, because there are other types of vehicle (for example, motor bike) that are neither cars nor bicycle. The generalization “Person” of “Student” and “Employee” is partial and overlapping, because there are students who are also employed.

In Table 5.1 ER and EER constructions are presented: as you can see each proposed construction has its specific graphical representation.

In particular (see Figure 5.9), total generalization (i.e., every instance of the superclass is an instance of some subclass) is represented by a solid arrow. One arrow with multiple subclasses (e.g., arrow from Woman/Man to People) means that subclasses are mutually exclusive.

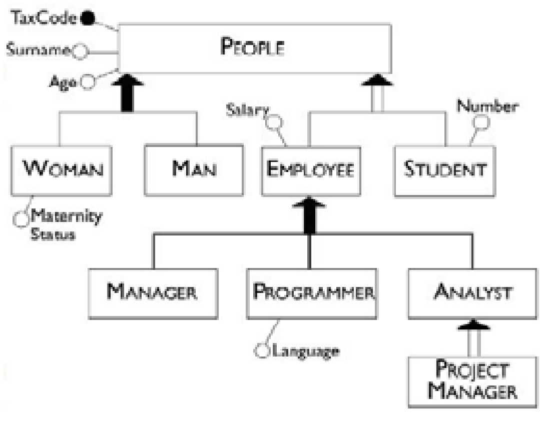
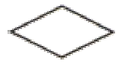
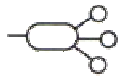
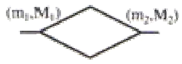
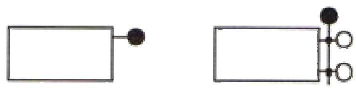
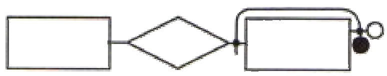


Figure 5.9 –Generalizations examples. (source: [R. A. Elmasri, S. B. Navathe, 2007])

Table 5.1 – ER and EER model constructions with their graphical representation. (source: [R. A. Elmasri, S. B. Navathe, 2007])

Constructions	Graphical representation
Entity	
Relationship	
Simple attribute	
Composite attribute	
Relationship cardinality	
Attribute cardinality	
Internal identifier	
External identifier	
Generalization	
Subclass	

5.2.2 The Unified Modelling Language (UML)

In this work, the geodatabase conceptual model has been developed in the *Unified Modelling Language* (UML). The UML has been chosen because of its solid semantics, notation definitions and its wide acceptance, and especially because ESRI software directly supports it.

The UML has become the de-facto standard for building *Object-Oriented* software. This language is an industry standard for modelling, design and construction of software systems as well as more generalized business and scientific processes. At present, databases designers usually use UML in many database design phases.

An important point to note here is that UML is a “language” for specifying and not a method or procedure. The UML is used to define a system, to detail the artifacts in the systems, to document and construct. The UML may be used in a variety of ways to support a development methodology, but in itself does not specify a methodology or process.

UML defines various basic diagram types, divided into two general sets: *structural modeling diagrams* and *behavioural modeling diagrams*. Structure diagrams define the static architecture of a model. They are used to model the "things" that make up a model, that is the classes, objects, interfaces and physical components. Moreover, they are used to model the relationships and dependencies between elements too. Relevant elements of UML *Class Diagrams*, the UML structure diagrams which have been used in our work, are briefly described in the following. UML Class Diagrams are very similar to ER diagrams. Nevertheless, they allow to specify, besides the structure of database schema, also operations accomplished on objects, according to the object-orientation principles.
The major UML advantage is that, even though its concepts are based on object-oriented techniques, resulting structure models can be used in order to design relational, object-oriented and relational-object databases.

UML Class Diagrams

UML Class Diagrams, which are very similar to the ER and EER ones, allow to define a structural specification of database schema according to an object-oriented approach defining, for each class, its name, attributes and operations. Class diagrams depict a static view of the model and are most useful in illustrating relationships between classes and interfaces. Generalizations, aggregations, and associations concepts are all supported. They are valuable in reflecting inheritance, composition, and connections respectively.
A Class Diagram shows the building blocks of any object-oriented system. Nevertheless, they are commonly used in order to describe sets of data objects and their relationships, in accordance with database conceptual design purposes.

In a Class Diagram *entities types* are modelled as *class*. An *entity* in a ER model is the equivalent of an *object* in UML. Fundamental elements of the Class Diagram are here presented.

Objects

Formally, an object can be defined as an identifiable entity, real or abstract, that has a precise role for an application domain. To constitute an entity, something must be identifiable (have identity), relevant (be of interest to the application domain) and describable, i.e. have characteristics. The object in the model should have properties that describe the characteristics of the corresponding entity in the universe of discourse. In object-oriented systems, each object is defined by three things:

- its *identity,* which distinguishes it from other objects. In an object-oriented system, each object is unique. This uniqueness of an object is achieved by means of the object identity. Object identity is that property of an object that uniquely distinguishes it from all other objects.
- its *state*. The state of an object is described by the values of its *attributes* (or properties) at a specific moment. An object attribute is actually a named property of the object *class* that describes a value held by each object of that class. Objects can have a single state throughout their lifetime, or they can go through many state transitions.
- its *behaviour,* described by the operations performed by the object. Behaviour describes object dynamics and interactions with other objects. In object-oriented systems, the behaviour of an object

is encapsulated in its *methods* or *operations* (also called *services or responsibilities*) performed by the object or performed on the object by other objects.

In the UML notation, an object is shown as a rectangle with two compartments. The top compartment shows the name of the object and its class, all underlined. The second compartment shows the attributes for the object and their values as a list (see Table 5.2).

Classes

Similar objects are grouped in *classes*. An object is an instance or occurrence of a class. Three concepts are associated with the use of classes. First, a class provides the *structural* definition for the instances of that class, that is, the names and types of their attributes and methods. Objects being instances of the class inherit the attributes and methods specified for the class. Second, a class serves as a template for creating new objects. New objects are created as instances of the class; hence they have access to all the attributes and methods of the class to which they belong. This is called *inheritance.* Third, each class is associated with a group of objects, which comprise its instances, and the class has to administer them.

Some classes are never directly instantiated, that is, they cannot have instances themselves. These classes are called *abstract classes*. Abstract classes define attributes and operations, but they don't contain objects. They are used in order to specify set of attributes and operations which can be inherited. The descendants of abstract classes, however, can have instances. Abstract classes are useful for reducing redundancy of class specification: general specifications can be defined in an abstract class that can be shared by other inheriting classes.

Classes are represented in the UML by a solid-outline rectangle with three compartments separated by horizontal lines, as shown in Table 5.2. The top compartment holds the *class name* and other general properties of the class (e.g. stereotypes); the middle compartment holds a list of *attributes* of the objects that belong to this class; the bottom compartment holds a list of *operations* which can be executed over the objects. In ER diagrams operations cannot be specified.

Optional listing of attribute data types, visibility and initial values may also be provided. In addition, it is possible to define the domain of attributes.

Interface

A class interface is a *specifier* for the externally visible operations of a class without specification of internal structure. The interface of a class specifies the set of operations that the class presents to other classes. An interface is represented in the UML as a rectangle with two compartments, the top showing the interface name with the keyword *«interface»*, and the bottom part showing the list of operations. The relationship between a class and its interface is shown by means of a *refinement* or *realization* symbol (see Table 5.2).

Relationships among objects

Objects and classes do not exist in isolation. In UML there are three general relationships between objects or classes, namely, *association*, *generalization* and *aggregation*. All these are abstraction mechanisms and aim at simplifying the modelling process:

- *Association.* Association is a physical or conceptual link between objects, and denotes some semantic dependency between the objects. An association in UML is the general relationship type between elements and it is the equivalent of a relationship in the ER mode. An example is the marriage relationship between a man and a woman. Objects in an association usually play a *role*. For example, the marriage association captures the fact that the man plays a “husband” role to the woman, while the woman plays a “wife” role to the man. Association has another property called *cardinality*, which describes the number of objects from one class that relate to a single object in the associated class. There are three general kinds of cardinality across an association, namely: one-to-one, one-to-many, and many-to-many. The marriage association, for example, is one-to-one relationship.

Association is represented in the UML as a line connecting two classes with the association name just above the line. This kind of association is called *binary association* in the UML. The association connector may include named roles at each end, cardinality, direction and constraints. Association *role* is shown at both ends of the line next to the class. Association cardinality or *multiplicity* is shown just above the association line near the appropriate class. Multiplicity is defined in the *min…max* form. An asterisk (*) indicates that there is not a participation maximum limit. In UML, a single asterisk signify a 0…* multiplicity, and a single 1 an 1…1 multiplicity. An association may have attributes and operations just like a class. In such case, it is called an *association class*. This is represented in the UML like an ordinary class with a dotted line connecting it to the association line (Table 5.2).

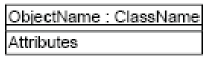
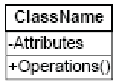
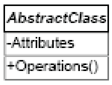
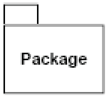
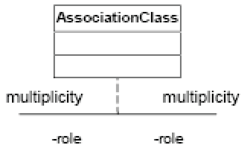
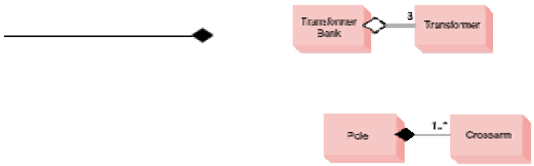
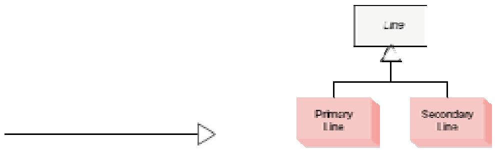
- *Generalization.* A generalization is used to indicate inheritance. Generalization is a relationship between a class (superclass) and one or more variation of the class (subclasses). The superclass holds common attributes and methods and the subclasses inherit them, adding their own attributes and methods. *Specialization* has the same meaning as generalization except that it takes the top-down perspective, starting with the superclass and splitting out variation. Generalization is also known as *‘is-a’ relationship*.

  In UML, generalization is represented by a line that connects the subclass to the superclass, with an open triangle on the end of the line that point to the parent class, as shown in Table 5.2.

- *Aggregation.* Aggregation is an abstraction mechanism, which allows a more complex object to be composed of one or more basic objects. Aggregation is a kind of association between the whole (complex object) and its parts (basic objects). Thus the relationship formed by aggregation is called a *‘part-of’ relationship*. This means that the aggregated instances are parts of the aggregate object. Aggregation relationships are shown by a white diamond-shaped arrowhead pointing toward the target or parent class.

  Composition is a strong form of aggregation in which the composite object controls the lifetime of the parts. Thus if a composite object is destroyed, it must destroy all of its parts. It is possible to have several levels of aggregation hierarchy among objects or classes. For example, a country is made of provinces, which are composed of municipalities. Composition in UML may be shown by a solid filled diamond. The multiplicity of the composite end may not exceed one (i.e., it is unshared).

Table 5.2 – UML static structure elements with their graphical representation.

Constructions	Graphical representation
Object	
Class	
Abstract class	
Interface	
Package	
Association	
Association class	
Composition	
Generalization	
Refinement	

The conceptual data model has been developed in UML, using the *Class Diagram* formalisms. In order to define the conceptual model using Class Diagrams, we used the *Microsoft Office Visio Professional 2003* software in order to carry out the visual modelling phase. This software is included in the general category of CASE tools (*computer-aided software engineering*), which are devoted to the database design activities.
This software tool has been chosen because of its compatibility with ESRI GIS environment, particularly with the ArcGIS 9.x platform, which is the one adopted by ITHACA's user groups and WFP's units. Finally, some ArcInfo UML model templates are available. These templates contain modeling objects which are necessary

in order to design with Visio a complete ESRI Geodatabase conceptual data model. Features of these templates are described in the next section.

5.2.3 The ArcInfo UML model

ArcInfo UML model templates extend the general UML model previously described allowing the representation of spatial entities in the conceptual model, that is, allowing the inclusion in the model of the ESRI objects types, such as feature classes, feature datasets, networks, and so on. These objects types are those defined in the ESRI Geodatabase logical model. Therefore, the conceptual UML model defined using the ArcInfo template can be easily converted into logical/physical schemas of the ESRI Geodatabase. This topic will be faced in the paragraph 5.3. We present in this paragraph the used ArcInfo UML model, the final proposed conceptual model, and the adopted modelling methodology.

ESRI Geodatabase data access objects

The ESRI geodatabase model is implemented as extension to the standard relational model by integrating it with object-oriented concepts. Entities in the geodatabase are represented as objects with attributes, relationships and behaviour. The ArcInfo software manages both the attributes and behaviour of objects, and database developers work in an object-component environment that is abstracted from the underlying physical relational database model and augmented by a programming framework of interfaces and methods. ArcInfo automatically provides the object-relational mapping and manages the integrity of the data in the underlying database.

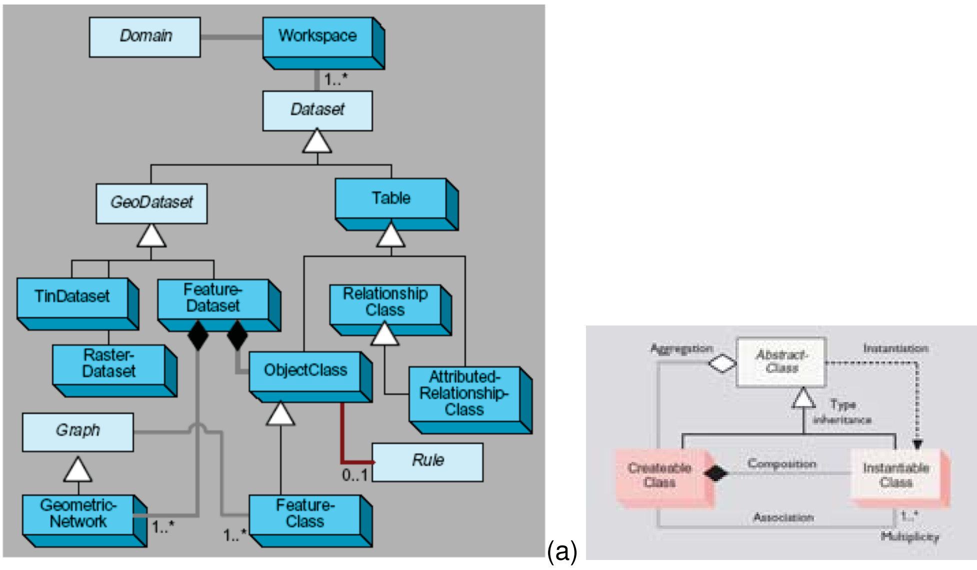


Figure 5.10 – (a) Simplified UML diagram of a portion of the geodatabase data access objects; (b) key for the object model diagram: an abstract class cannot be used to create new objects, but it is a specification for subclasses; a createable class represents objects that the user can directly create, and an instantiable class cannot directly create new objects, but objects of this class can be created as a property of another class or created by functions from another class. (source: [M. Zeiler, 1999])

What differentiates ESRI Geodatabase model from other GIS spatial data model, is that object-oriented methodology is applied to geographic data modelling. A developer can interact with data objects through a framework of object-oriented software classes called the *geodatabase data access objects*. Figure 5.10 shows a simplified UML diagram of a portion of these geodatabase data access objects in *ArcObjects*[1]. These objects have already been presented in section 5.1.3.

Particularly, ArcInfo developers interact with geographic data through a set of data objects, such as datasets, tables, feature classes, rows, objects, and features. Moreover, a data modeler can use these standard feature types to implement a rich data model. As a matter of fact, for advanced applications, a developer can extend the standard feature types and create custom features using the object-oriented technique of *type inheritance*. Inheritance means that an object class can be defined to include the behaviour of another object class and have additional behaviours. Therefore, users can create custom feature types in ArcInfo and inherit the behaviour of standard features.

ArcInfo UML model template and basic modelling techniques

The ArcInfo model template provides in UML a hierarchy of object classes ready for the use in MS Visio. These are *object, feature, simple junction feature, complex junction feature, simple edge feature,* and *complex edge feature*. These object classes are shown in Figure 5.11.

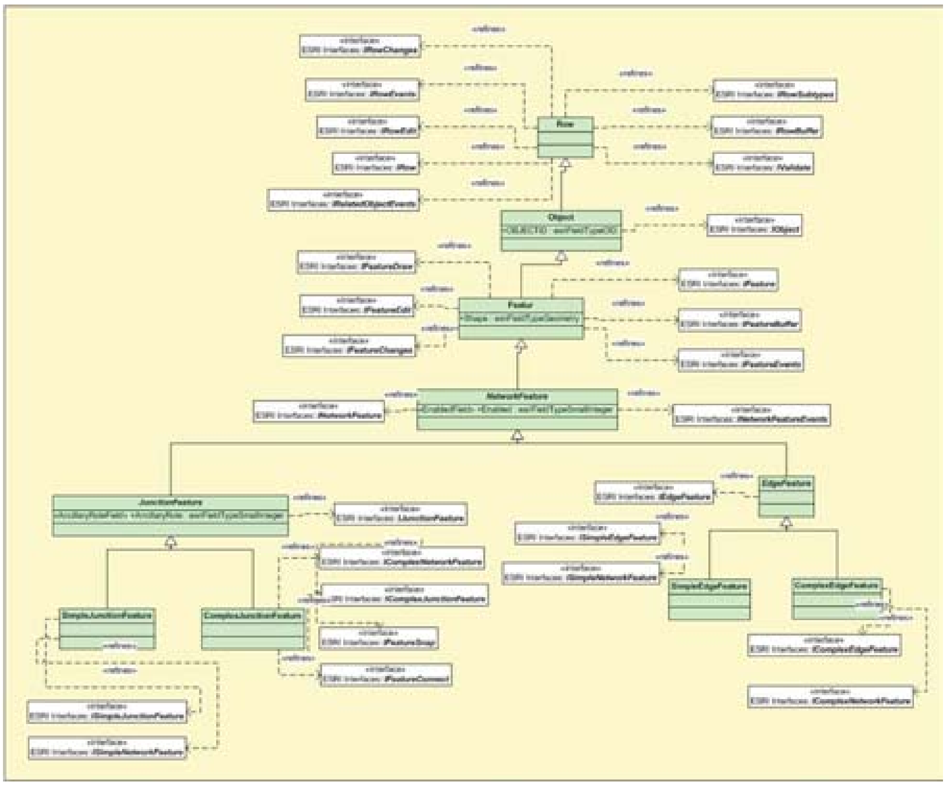


Figure 5.11 – Standard ArcInfo object classes in the ESRI UML template. These classes are part of the geodatabase data access objects and are supplied with the template. Many of these classes have already been presented in section 5.1.3. A feature is a type of row that stores geometry in a table column. (source: [M. Zeiler, 1999])

[1] ArcGIS (hence the geodatabase) is built on a technology framework known as ArcObjects. ArcObjects comprises an integrated library of software components with GIS functionality and programmable interfaces. ArcObjects framework is based on the Microsoft Component Object Model (COM) protocol and is compliant with the OpenGIS simple feature specifications.

The ArcInfo object classes include a number of predefined fields such as object identifiers and geometries. These fields define the required properties of the objects in these classes. Then, the data modeler can add additional custom fields.

Moreover, each of these objects classes implements a set of interfaces (see Figure 5.11). Each interface contains a set of related methods for actions such as storing, editing, drawing, querying, and validating objects. An ArcInfo object class provides default implementation for each of its interfaces.

The user can customize these object classes by extending the ArcInfo object classes using the MS Visio tool functionalities. It is possible to create new types of features that have the attributes and behaviour of these classes, but that add new attributes and behaviours. This is called *type inheritance*.

Fundamental operations, which are necessary in order to define a data model in UML, are here presented.

Figure 5.12 shows the main MS Visio interface. In the *Drawing Page* it is possible to develop a own model diagram using the UML Static Structure elements contained in the *Stencils* area. Moreover, the *Model Explorer* allows to see the whole developed data structure.

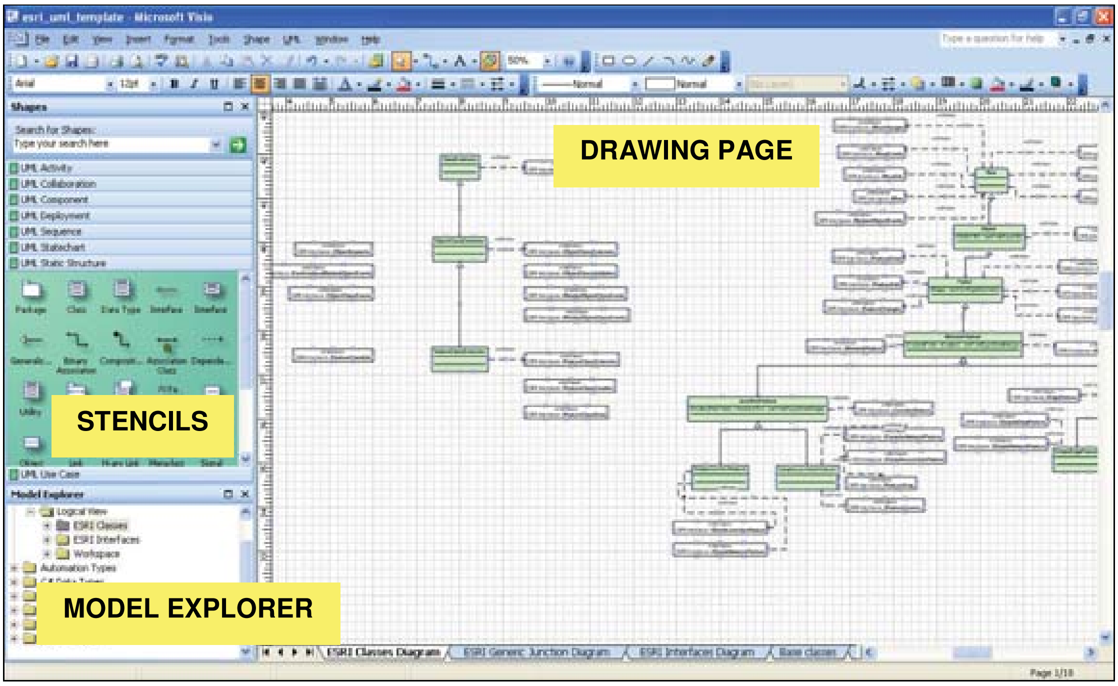


Figure 5.12 – Principal MS Visio interface.

The ArcInfo UML Model contains four different *packages*: *Logical View*, *ESRI Classes*, *ESRI Interfaces* and the *Workspace*. They can be shown in the Model Explorer (see Figure 5.13). The *Logical View package* contains the other ones. The *Workspace package* is the equivalent of the ESRI ArcInfo Geodatabase. Many other *packages* can be created (using the *package* shape in the *Stencils* area); they are equivalent to Geodatabase *feature datasets* and they must be placed in the *Workspace*. For each new *package* a *static structure* is automatically created. It can be then edited in the corresponding *Drawing Page*.

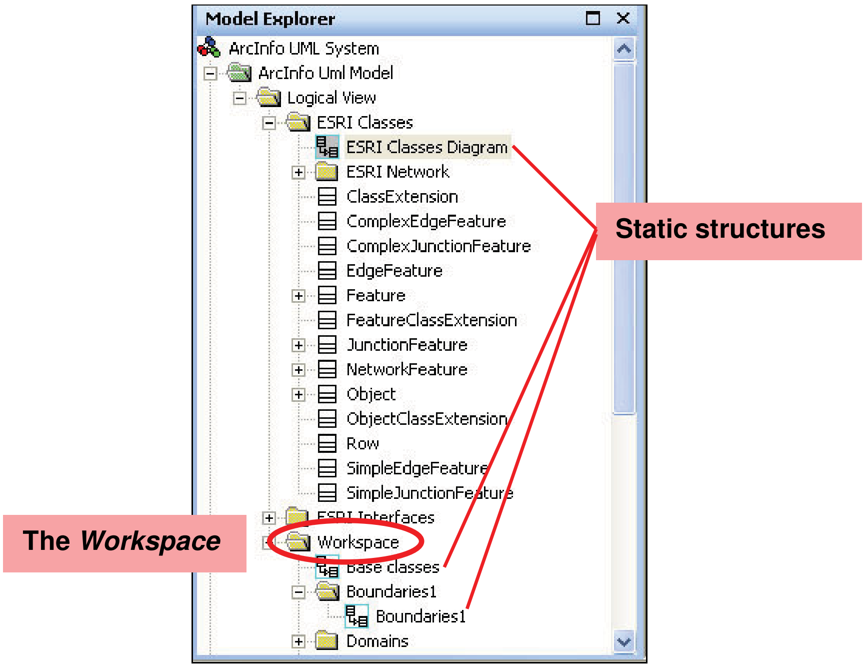


Figure 5.13 – The ArcInfo UML model in MS Visio Model Explorer

Geodatabase *feature classes* are defined in the UML data structure model using *object* and *feature* object classes supplied within the ESRI template. These object classes can be reached in the *ESRI Classes package* in the *Model Explorer* (see Figure 5.13). A new *class*, extracted from the *Stencils* area, can be placed in the *Drawing Page*. Then, using the *generalization* construction, the new created *class* can inherit OBJECTID and SHAPE attributes, which are fundamental for GIS objects (Figure 5.14). As a matter of fact, in the ESRI Geodatabase simple feature classes have two predefined fields: a feature ID and a geometry field; these predefined fields are used for uniquely identifying objects and storing feature shapes in the feature class table and they are managed by ArcInfo.

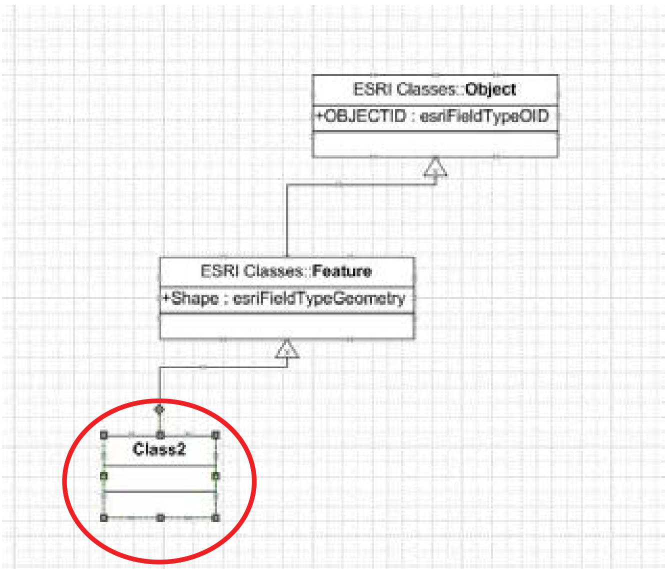


Figure 5.14 – Creation of a new feature class within the Geodatabase static structure in UML.

Afterwards, properties of the created feature class, such as its name, attributes, and geometry type, can be defined using the UML Class Properties interface (see Figures 5.15, 5.16, 5.17).

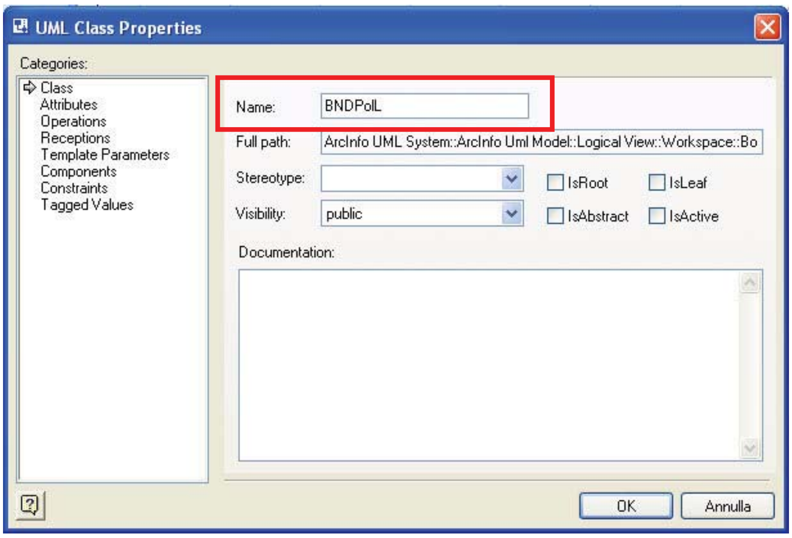


Figure 5.15 – Definition of feature class properties: the name.

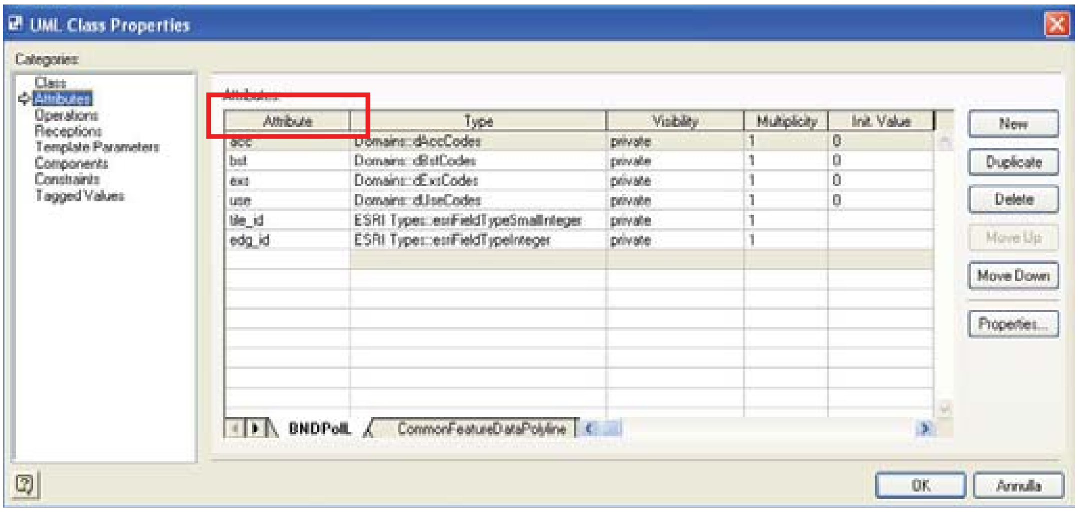


Figure 5.16 – Definition of feature class properties: attributes.

As you can see in figure 5.16, it is possible to define some custom fields in a feature class table in order to implement the various types of attributes needed to realize the properties of custom features. Moreover, in this phase it is possible to define the attribute field types using standard ESRI types (*float, double, short integer, long integer, text, date, object ID, binary large object-BLOB* and *raster,* see Figure 5.16).

The feature geometry type can be defined using *tagged values* (see Figure 5.17). Available geometry types are: *EsriGeometryPolygon*, *EsriGeometryPolyline*, *EsriGeometryPoint*.

Other ArcGIS object classes, such as tables, can be defined in UML in a similar way.

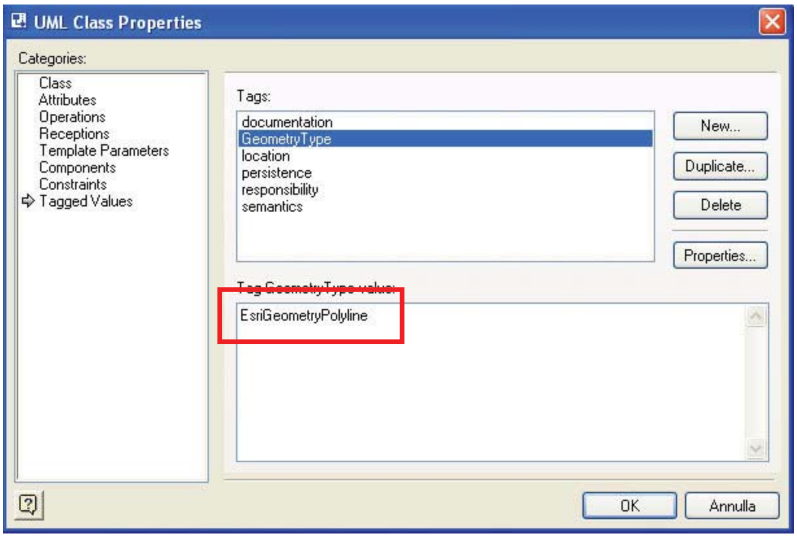


Figure 5.17 – Definition of feature class properties: the geometry type.

Relationships (see Figure 5.18) are defined in the UML data structure model using *association* objects extracted from the *Stencils* area.

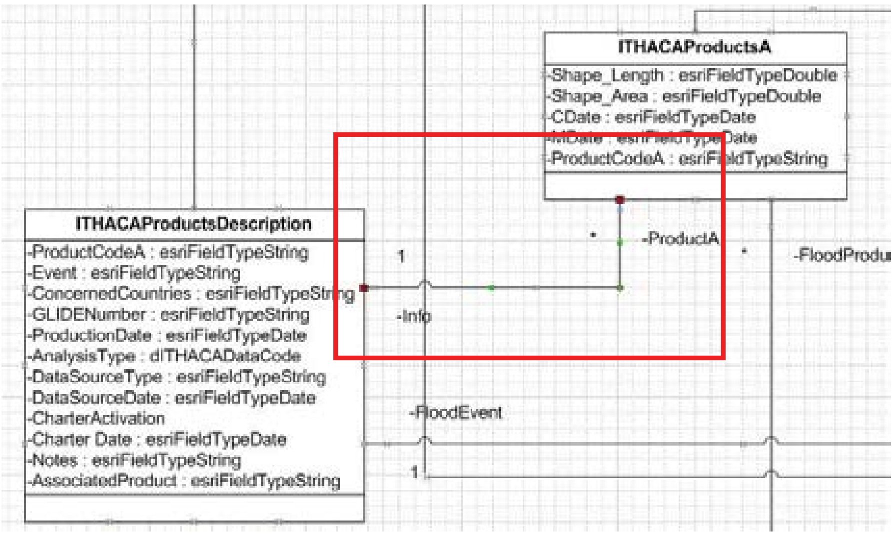


Figure 5.18 – Creation of a relationships between two feature classes within the static structure in UML.

A relationship in ArcGIS has several properties that define how objects in the origin relate to objects in the destination. In UML it is possible to specify these properties during the creation of the association. These properties are:

- *Origin* and *destination classes*;
- *Primary* and *foreign keys* (see Figure 5.22), defined as *tagged values*;
- *Cardinality* (see Figure 5.21). A relationship cardinality specifies the number of objects in the origin class that can relate to a number of objects in the destination class. A relationship can have one of three cardinalities (Figure 5.19).

One-to-one: one origin object can relate to only one destination object. For example, a parcel can have only one legal description. In ArcGIS, this cardinality also covers many-to-one. An example of a many-to-one relationship is many parcels relating to the same legal description.

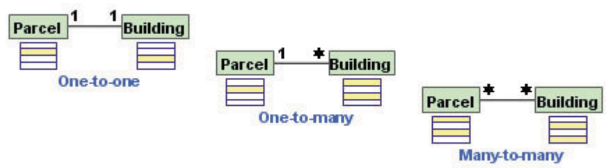


Figure 5.19 – Cardinalities of relationships.

One-to-many: one origin object can relate to multiple destination objects. For example, a parcel may have many buildings. In a one-to-many relationship, the "one" side must be the origin class and the "many" side must be the destination class.

Many-to-many: one origin object can relate to multiple destination objects and, conversely, one destination object can relate to multiple origin objects. For example, a given property may have many owners, and a given owner may own many properties (Figure 5.20).

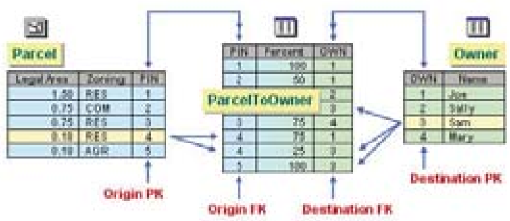


Figure 5.20 – Many-to-many relationships.

In one-to-one and one-to-many relationships, values in the primary key of the origin class directly relate to values in the foreign key of the destination class. Many-to-many relationships, on the other hand, require the use of an intermediate table to map the associations. The intermediate table maps primary key values from the origin to foreign key values from the destination. Each row associates one origin object with one destination object.

- *Attributes* for the association;
- *Name* of the association (see Figure 5.21);
- *Forward* and *backward labels* that display when you navigate related records in ArcMap (see Figure 5.21).

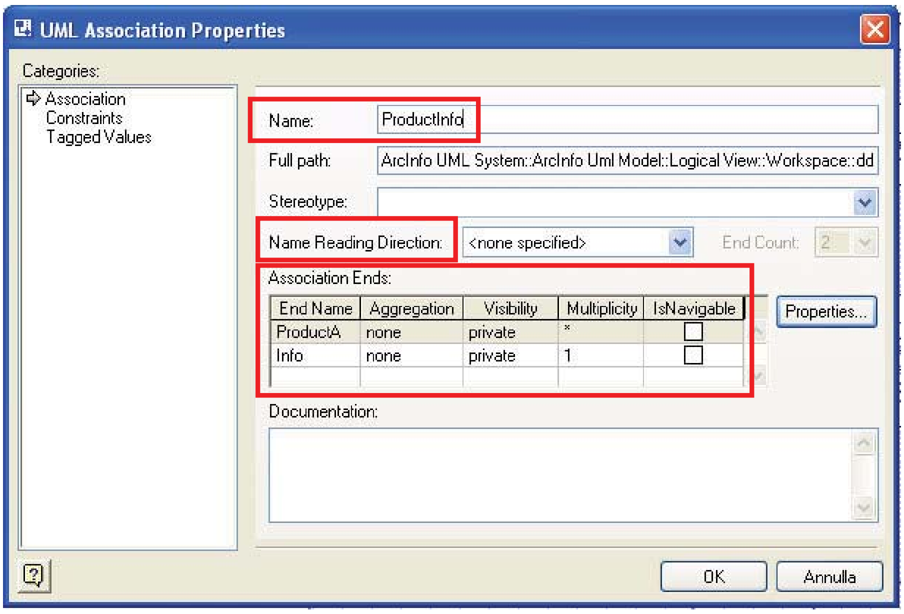


Figure 5.21 – Definition of association properties.

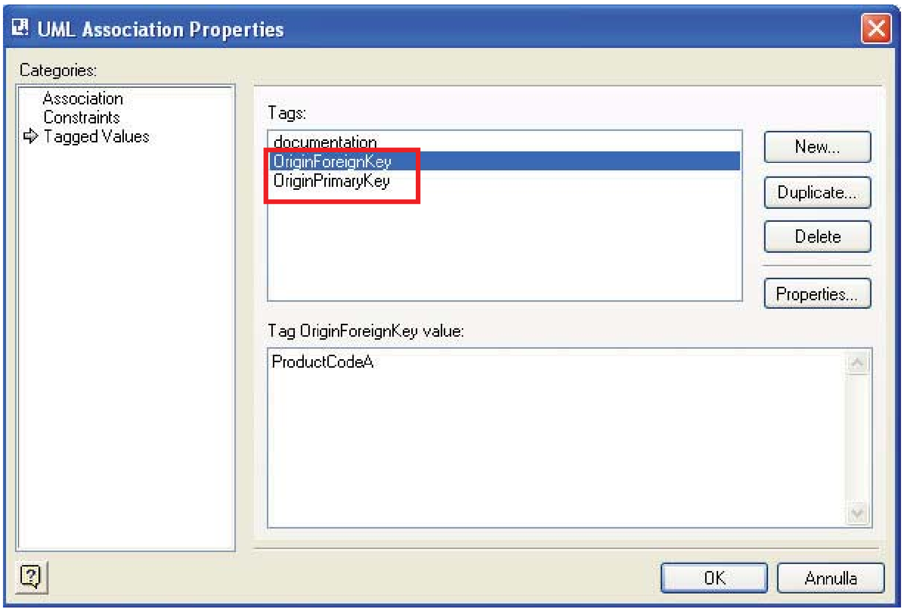


Figure 5.22 – Definition of association primary and foreign keys.

In the UML static structure model also Domains can be defined using the *TemplateCodedValueDomain* object supplied in the ArcInfo template. Attribute domain is a constraint on attribute value in feature and object classes. This constraint can be a range of numeric values or a list of valid values. In the *Properties* interface of the created domain, the specific attributes values can be defined (see Figure 5.23). Then, in the involved Feature Class, we must choose the defined domain as the related attribute data type.

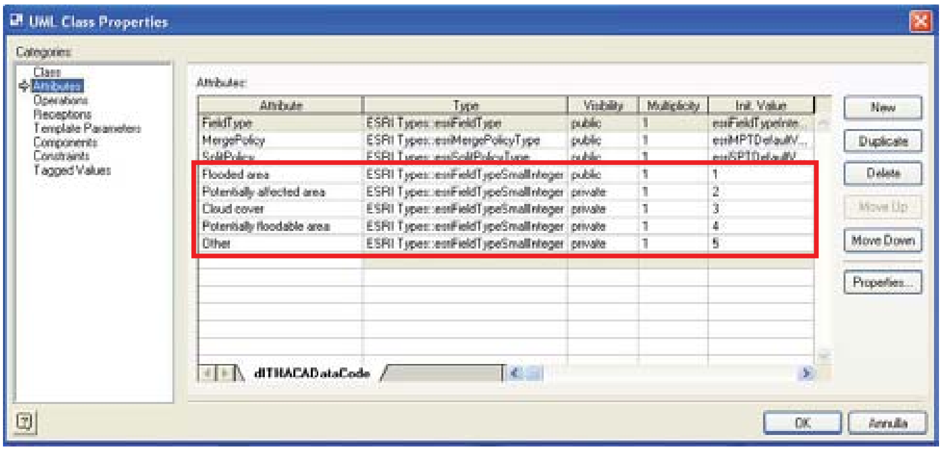


Figure 5.23 – Definition of domain coded values.

Finally, UML supplies simple tools in order to define other ArcInfo Geodatabase class objects in the data model, such as Subtypes or Networks. The description of necessary operations for creation of these objects is out of the scopes of this document, and therefore it has been avoided.

5.2.4 Geodatabase conceptual modelling: design steps

The conducted geodatabase design process has been performed as a series of steps. The considered steps are:

a. organization of the needs assessment phase results;
b. definition of entities and their relationships;
c. identification of the valuable representations of entities;
d. selection of necessary ArcInfo Geodatabase data model objects. In this phase geometry type of features has been defined, relationships and attribute types have been implemented, and data have been organized into geographic datasets.

The development of the UML model of the whole geodatabase data structure has been conducted side by side with these steps, allowing subsequent refinements of the final model.
The first three steps develop the effective conceptual model, classifying features based on an understanding of the data required to support the ITHACA and WFP's functions, and deciding their spatial representation (point, line, area, raster, or non geographic). The last two steps already refers to the logical data model design phase, matching the conceptual model to ArcInfo datasets.

a. Organization of the needs assessment phase results

In the needs assessment phase:

- functions that support involved organizations goals and objectives have been identified;
- the data required to support these functions have been identified;
- the providers and the consumers of geographic data have been identified and the key data flows have been described;
- the data have been organized into logical groupings;
- finally, all data sources have been defined.

The results of this phase have been described in Chapter 4, Figure 4.39 and Table 4.8 can be thought as the starting points for the whole conceptual design phase. Therefore, in this preliminary step, results of the needs assessment have been analysed and shared with all the users interested by the geodatabase implementation (ITHACA's user groups and WFP's units). The main result of this step has been the involvement of all the users in order to meet a joint and coherent vision of common geographic data.
According to the results of the preliminary step, we decided to implement a data model where, for any geodatabase objects, its modelled data structure is matching with the one of the corresponding identified source. This structure, which is not optimized, allows, on the other hand, to preserve and maintain the original information completeness. So, the defined data structure modelled in this way constitutes the starting point for future discussions aimed at defining a more optimized data model solution. Moreover, this preliminary solution allows: to rapidly and easily load source data in the implemented geodatabase, with no or few needs of import procedures development, to supply to the users a solution containing very familiar data structures and field definitions, and, finally, to reuse already developed users' procedures for data selection, symbolization, export, and so on.

Another important point to note is that we decided to organize all the needed geographic data into two different geodatabases:

- the *base data geodatabase* which include basic spatial and alphanumeric information, in order to produce basic geographic outputs and analysis;
- the *transportation geodatabase* (according to the conceptual schema developed by UNJLC).

As aforementioned in Chapter 4, the conceptual model for the *transportation* theme has been developed by the UNJLC unit and supplied to ITHACA for the translation in the UML *Class Diagram* formalism. Therefore, the data belonging to the *transportation* theme did not require the activities performed in steps presented later on. For this reason, these data will not be considered in the next part of this chapter.

b. Definition of entities and their relationships

In this step we examined the data classification produced during the prior analyses and we identified all the necessary distinguishable objects, or *entities*, that had a common set of properties.

Tables 5.3 and 5.4 show the result of this activity, that is, all the identified entities with their description. As you can see, 11 themes have been identified and proposed (Table 5.3). For each of these themes, the needed entities, or geodatabase objects defined, have been summarized in Table 5.4, with their proposed names and a brief description. It can be noticed that in Table 5.4 *transportation* entities are not included, as aforementioned, as well as *vegetation* entities, which were not identified at the moment of the geodatabase design.

Table 5.3 - Thematic data groupings proposed for the geodatabase.

DATASETS	
Name	**Acronym**
Boundaries	BND
Elevation	ELV
Hydrography	HYD
Physiography	PHY
Population	POP
Transportation	TRN
Industry	IND
Utilities	UTI
Vegetation	VEG
Flood Analyses	FLD
Names	NMS

Table 5.4 – Entities implemented in the base data geodatabase.

DATASETS	CLASSES	Description
Boundaries	BNDCoastline	*Coastlines*
	BNDOceanSea	*Oceans and Seas*
	BNDPolA	*Political Boundaries (areas)*
	BNDPolL	*Political Boundaries (outlines)*
Elevation	ELVCntline	*Contour lines on land*
	ELVDepthline	*Depth lines*

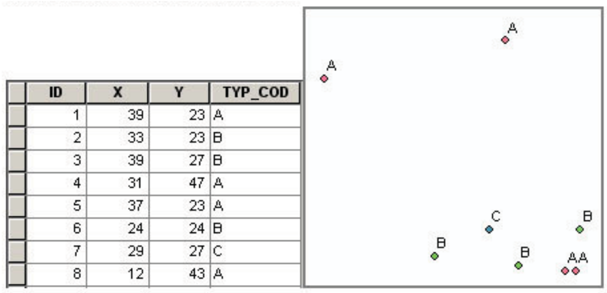
	ELVDtm	*Digital Terrain Model*
	ELVElevPoint	*Elevation points*
Hydrography	HYDBasin	*Drainage basins*
	HYDCanal	*Inland water canals*
	HYDInWaterA	*Inland water bodies*
	HYDInWaterL	*Inland water paths*
	HYDMiscWaterL	*Miscellaneous water elements*
	HYDMiscWaterP	*Miscellaneous water elements*
	HYDSwbd	*SRTM Water Body Areas*
Physiography	PHYGroundCover	*Ground cover broad classes*
	PHYLandCover	*Land cover (1km resolution)*
	PHYLandsatMos	*Orthorectified Landsat TM Mosaics (year 2000)*
	PHYModisLC	*Modis Land Cover data*
	PHYModisLI	*Modis Land Imagery*
	PHYModisNDVI	*Modis NDVI data*
Population	POPAnthrFeatP	*Anthropic features*
	POPBuiltUpA	*Built-Up areas*
	POPPopDensity	*Population density*
Industry	INDIndA	*Industries (extraction/fish)*
	INDIndP	*Industries (extraction/fish)*
Utilities	UTITransmLines	*Transmission lines (power, pipelines, etc.)*
	UTITransmNodes	*Transmission nodes (plants, pumping, etc.)*
Flood Analyses	ITHACAProductsA	*Results of flood events analyses performed by ITHACA during Early Impact activities*
	ITHACAProductsDescription	*Description of products and flood events*
	DartmouthData	*Historical areas hit by flood. Data produced by the Dartmouth Flood Observatory (DFO)*
	DartmouthTable	*Description of DFO historical data*
	ITHACAMaps	*Maps produced by ITHACA during flood events*
	EventImagery	*Base satellite imagery used for the flood analyses performed by ITHACA*
Names	NMSAdmReg	*Names of administrative region features*
	NMSHydType	*Names of hydrographic type features*
	NMSHypType	*Names of hypsographic type features*
	NMSLocType	*Names of locality or area type features*
	NMSPopPlace	*Names of populated place features*
	NMSSptType	*Names of spot type features*
	NMSStrType	*Names of street, highway, road or railroad type features*
	NMSUndType	*Names of undersea type features*
	NMSVegTyp	*Names of vegetation type features*

c. Identification of entities representations

In this step we classified previously chosen entities by their type of representation. Some entities have a geometric representation with corresponding attributes; these are classified by their geometric characteristics. Other entities are represented by alphanumeric information only, while other by images. In addition, generally features will have different representations at different map scales.

The following terms are provided for assigning considered types:

- point: illustrates the location of a feature whose shape is too small to be defined as an area on a map of a given scale;
- line: illustrates the location of a feature whose shape is too narrow to be defined as an area on a map of a given scale;
- area: illustrates the location and polygonal shape of a feature on a map of a given scale;
- table: a table contains a X and a Y coordinate fields. A simple table of coordinates and other data can be converted to point events (Figure 5.24). The point events created from the table behave just like a point feature class, and can be symbolized and labelled using attributes in the table;



ID	X	Y	TYP_COD
1	39	23	A
2	33	23	B
3	39	27	B
4	31	47	A
5	37	23	A
6	24	24	B
7	29	27	C
8	12	43	A

Figure 5.24 – XY events from a table. (source: [ESRI, ArcGIS Desktop Help, 2006])

- object: identifies a feature for which no point, line or area is required, and for which there is no geometric or graphic representation;
- raster: represents an area using rectangular cells (satellite image, aerial photograph, continuous data);
- image, photo, drawing: each of them represents a digital picture that cannot be used for analysis.

One point that must be noted here is that the different representations for the proposed datasets have been forced by the selected source for each considered entity. As a matter of fact, for this application not many data sources, characterized by scales and consistent resolutions with used map scales, have been identified because of restrictive data search parameters defined in the needs assessment phase.
Table 5.5 shows entities with their chosen spatial types.

Table 5.5 – Entities implemented in the base data geodatabase and their spatial representation.

DATASETS	CLASSES	DataType
Boundaries	BNDCoastline	line
	BNDOceanSea	area
	BNDPolA	area
	BNDPolL	line
Elevation	ELVCntline	line
	ELVDepthline	line
	ELVDtm	raster
	ELVElevPoint	point
Hydrography	HYDBasin	area
	HYDCanal	line

	HYDInWaterA	area
	HYDInWaterL	line
	HYDMiscWaterL	line
	HYDMiscWaterP	point
	HYDSwbd	area
Physiography	PHYGroundCover	area
	PHYLandCover	raster
	PHYLandsatMos	raster
	PHYModisLC	raster
	PHYModisLI	raster
	PHYModisNDVI	raster
Population	POPAnthrFeatP	point
	POPBuiltUpA	area
	POPPopDensity	raster
Industry	INDIndA	area
	INDIndP	point
Utilities	UTITransmLines	line
	UTITransmNodes	point
Flood Analyses	ITHACAProductsA	area
	ITHACAProductsDescription	object
	DartmouthData	area
	DartmouthTable	object
	ITHACAMaps	object
	EventImagery	raster
Names	NMSAdmReg	table
	NMSHydType	table
	NMSHypType	table
	NMSLocType	table
	NMSPopPlace	table
	NMSSptType	table
	NMSStrType	table
	NMSUndType	table
	NMSVegTyp	table

d. Selection of necessary ArcInfo Geodatabase data model objects

The objective of this phase is to determine how data are to be represented in ArcInfo. For each of the spatial types identified in the previous step, we assigned a corresponding ArcInfo geodatabase representation. Moreover, in this phase attribute types of features and relationships have been implemented and data have been organized into geographic feature datasets.

In the ArcInfo geodatabase discrete entities can be stored as simple features, complex or connected features and objects. Points can be unconnected (point feature) or connected (simple or complex junction feature) and lines can be unconnected (line feature) or connected such as road network (simple or complex edge feature). An area can be stand-alone (polygon feature) or space-filling such as vegetation cover (polygon feature later assigned to a planar topology).

Table 5.6 shows the ArcInfo types assigned to geodatabase entities.

Table 5.6 – Entities implemented in the base data geodatabase and their ArcInfo geodatabase type.

DATASETS	CLASSES	DataType	ArcInfo type
Boundaries	BNDCoastline	line	Polyline feature
	BNDOceanSea	area	Polygon feature
	BNDPolA	area	Polygon feature
	BNDPolL	line	Polyline feature
Elevation	ELVCntline	line	Polyline feature
	ELVDepthline	line	Polyline feature
	ELVDtm	raster	Raster catalog
	ELVElevPoint	point	Point feature
Hydrography	HYDBasin	area	Polygon feature
	HYDCanal	line	Polyline feature
	HYDInWaterA	area	Polygon feature
	HYDInWaterL	line	Polyline feature
	HYDMiscWaterL	line	Polyline feature
	HYDMiscWaterP	point	Point feature
	HYDSwbd	area	Polygon feature
Physiography	PHYGroundCover	area	Polygon feature
	PHYLandCover	raster	Raster catalog
	PHYLandsatMos	raster	Raster catalog
	PHYModisLC	raster	Raster catalog
	PHYModisLI	raster	Raster catalog
	PHYModisNDVI	raster	Raster catalog
Population	POPAnthrFeatP	point	Point feature
	POPBuiltUpA	area	Polygon feature
	POPPopDensity	raster	Raster catalog
Industry	INDIndA	area	Polygon feature
	INDIndP	point	Point feature
Utilities	UTITransmLines	line	Polyline feature
	UTITransmNodes	point	Point feature
Flood Analyses	ITHACAProductsA	area	Polygon feature
	ITHACAProductsDescription	object	Object class
	DartmouthData	area	Polygon feature
	DartmouthTable	object	Object class
	ITHACAMaps	object	Object class
	EventImagery	raster	Raster catalog
Names	NMSAdmReg	table	Table
	NMSHydType	table	Table
	NMSHypType	table	Table
	NMSLocType	table	Table
	NMSPopPlace	table	Table
	NMSSptType	table	Table
	NMSStrType	table	Table
	NMSUndType	table	Table
	NMSVegTyp	table	Table

In this phase we defined relationships among entities. There are some benefits in implementing relationship classes in a geodatabase:

- relationship classes help enforce referential integrity. A relationship class can be set up so that, when an object is modified, related objects update automatically. This can involve physically moving related features, deleting related objects, or updating an attribute. Moreover, by setting rules, a relationship class can restrict the type of relations that are valid.
- relationship classes facilitate editing in *ArcMap*. By providing automatic updates to related objects, a relationship class can save the user from performing additional edit operations. Moreover, relationship classes help accessing objects during editing activities. User can select an object and then find all related objects. Because relationship classes are stored in the geodatabase, they can be managed with versions. Versions allow multiple users to edit the features or records in a relationship at the same time.
- relationship classes allow to query related features classes and records.

Then, we defined the structure of feature classes with subtypes, if necessary, and whether they stand as a separate feature classes or are contained within a feature dataset. Feature datasets are used to group feature classes for which topologies must be defined or that must be edited simultaneously. Moreover, a feature dataset is a container for feature classes that share the same spatial reference and can be used simply to group feature classes thematically.

In particular, as it will be presented in the following part, within the developed geodatabase all the defined feature classes are contained in a unique feature dataset.

Finally, in this phase needed topological rules and domains can be defined too.

The implemented UML model

According to the activities conducted in the previously described phases, we implemented in UML the data structure model of the ArcInfo Geodatabase. Features of the defined model are here described.

Figure 5.25 shows a survey of the whole defined structure, as presented in the Visio Model Explorer.

As you can see, within the *Workspace package* (namely, the geodatabase) we defined a unique feature dataset (the "*dd_wgs84*" one) containing all the needed feature classes, which are separated in the proposed thematic groupings or themes ("*Boundaries*", "*Elevation*", "*FloodAnalyses*", "*Hydrography*", "*Industry*", "*Physiography*", "*Population*", "*Transportation*", "*Utilities*", and "*Vegetation*").

Within the *Workspace* there are also all the defined domains, the tables contained in the "*Names*" and "*FloodAnalyses*" themes, and the *Base classes* static structure (see Figure 5.26). Within this static structure base classes (object data, point data, line data and polygon data) have been defined, that is, the classes which contain the OBJECTID and SHAPE attributes and some common custom fields (namely the *Creation Date* or *Cdate*, the *Modification Date* or *Mdate,* and the *Feature Code* defined in the original VMap data or *f_code*). Using generalization property, all other classes, subsequently created, can inherit automatically these fundamental attributes.

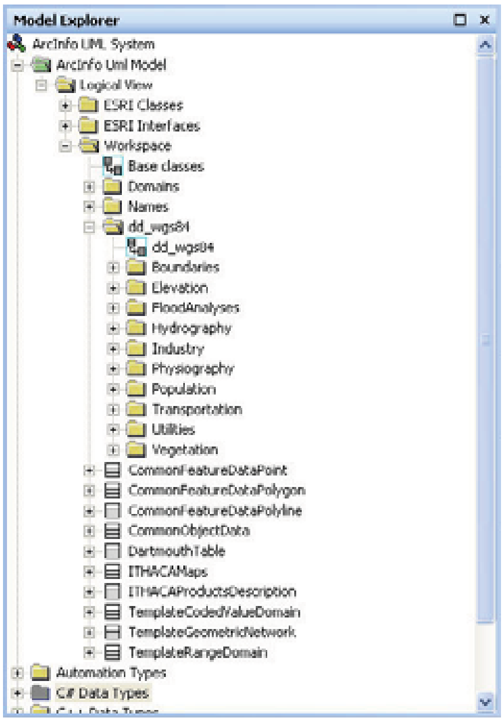


Figure 5.25 – Survey of the whole implemented UML geodatabase model.

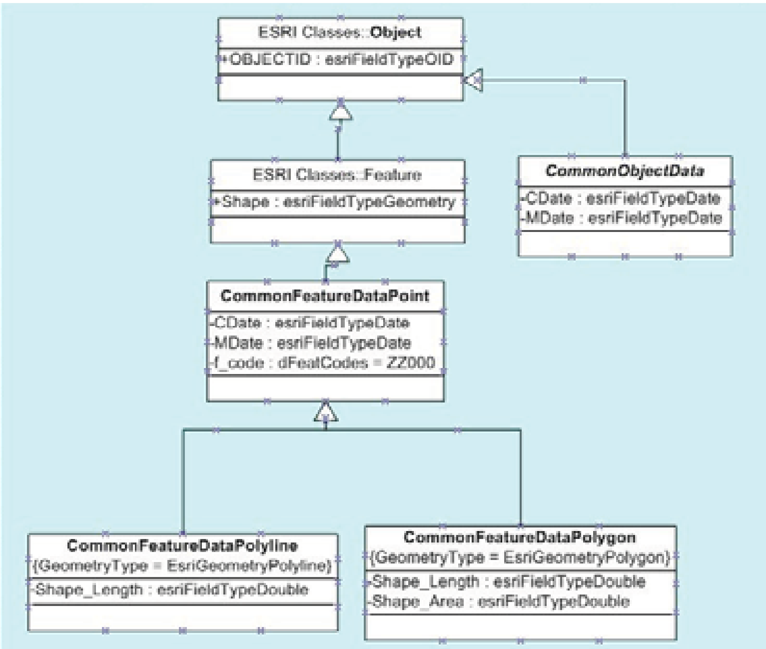


Figure 5.26 – UML base classes static structure.

In the following figures the UML static structures implemented for the different themes of the geodatabase are shown.

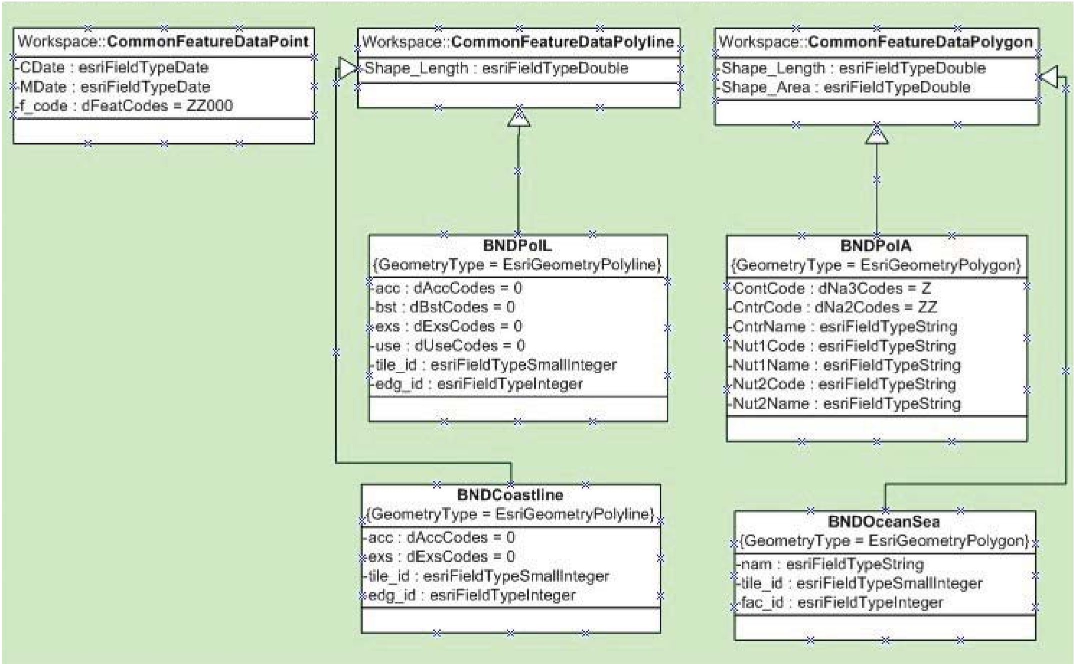


Figure 5.27 – UML Boundaries static structure.

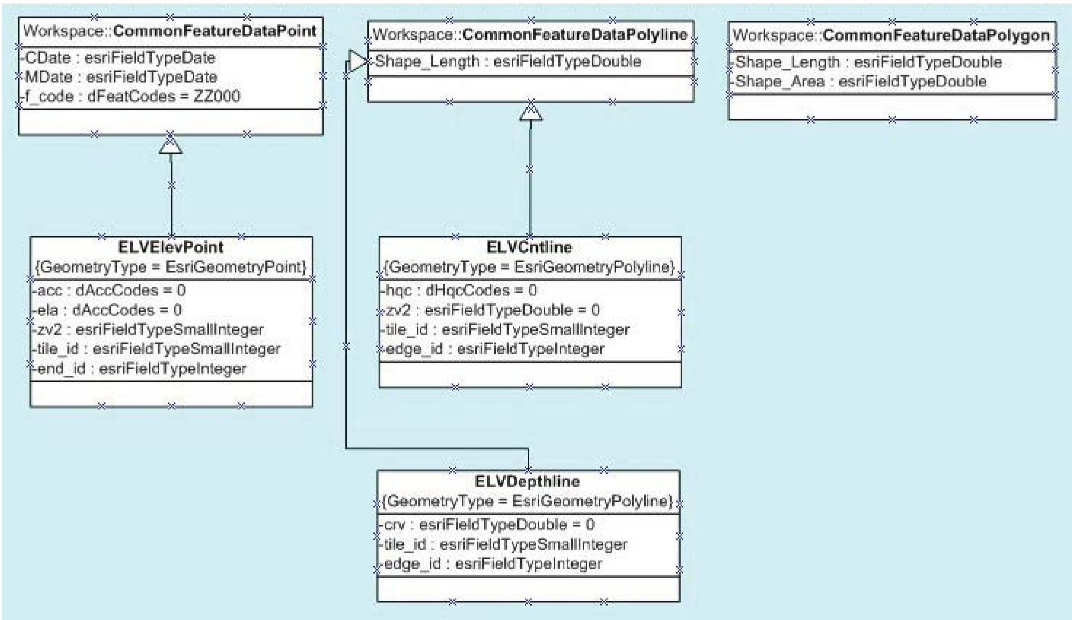


Figure 5.28 – UML Elevation static structure.

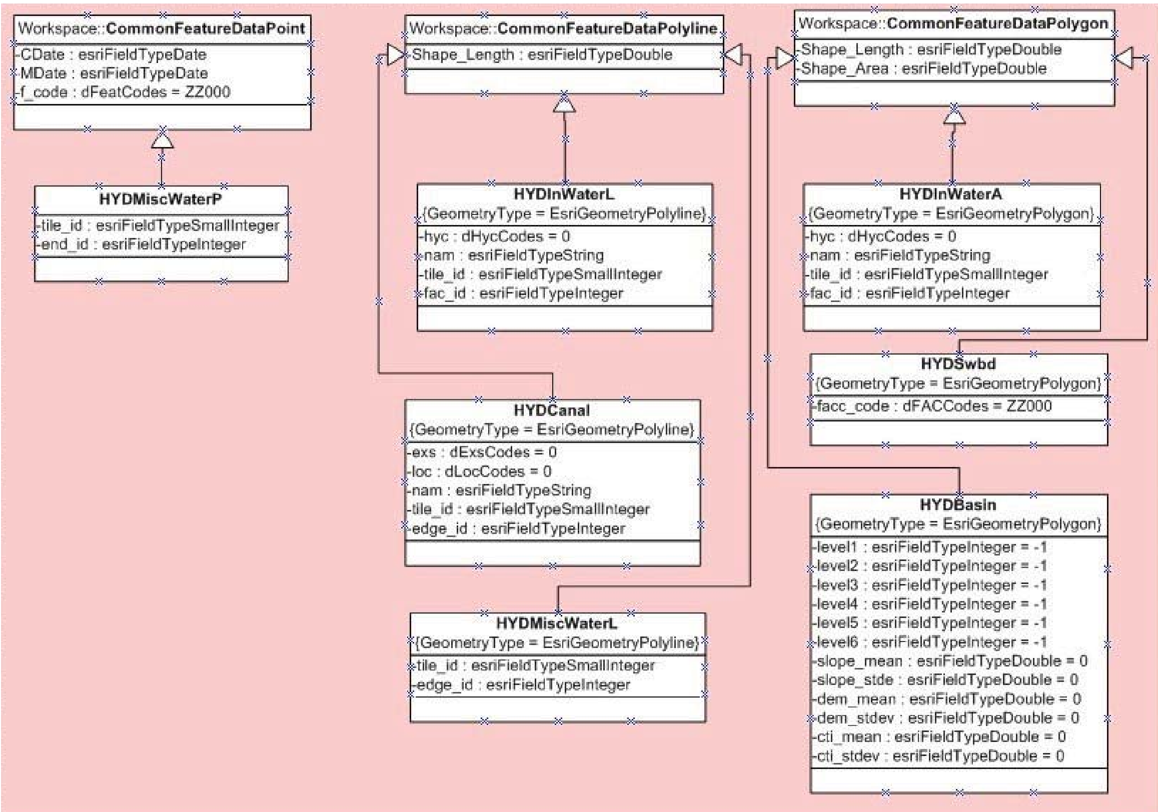


Figure 5.29 – UML Hydrography static structure.

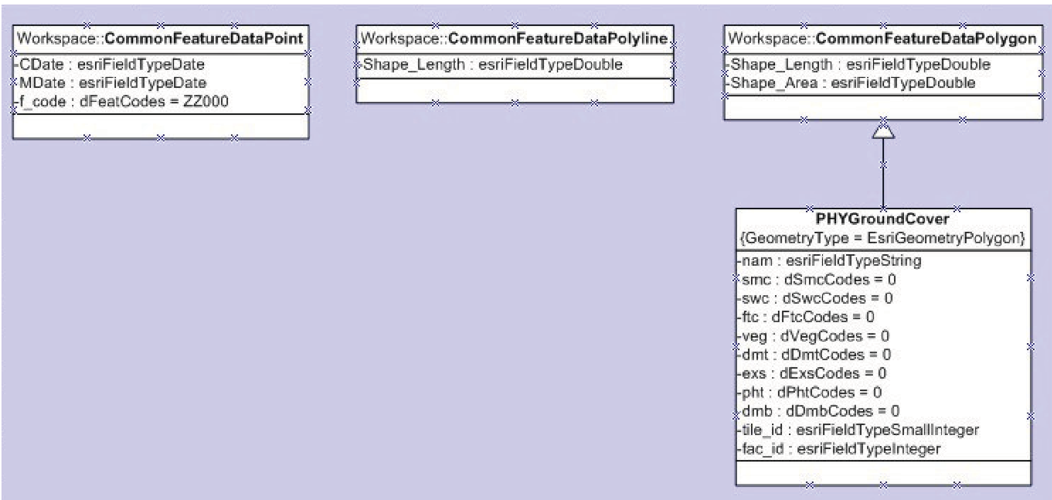


Figure 5.30 – UML Physiography static structure.

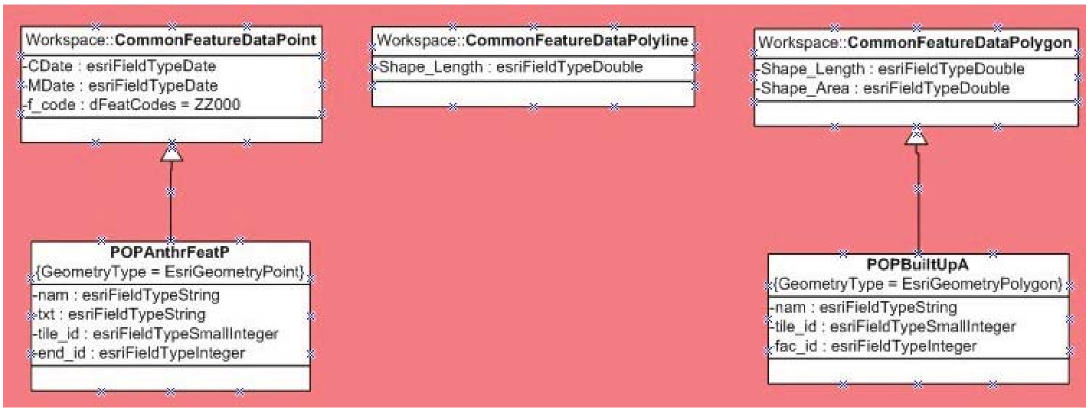


Figure 5.31 – UML Population static structure.

Figure 5.32– UML Industry static structure.

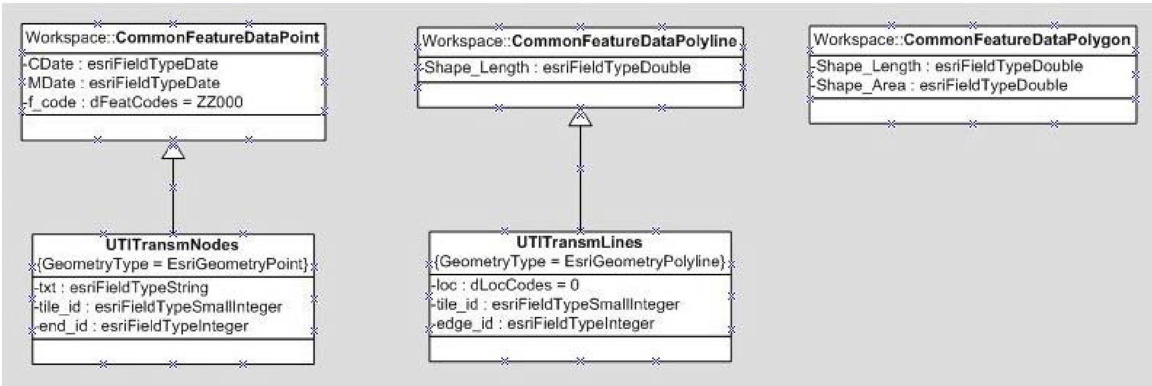


Figure 5.33– UML Utilities static structure.

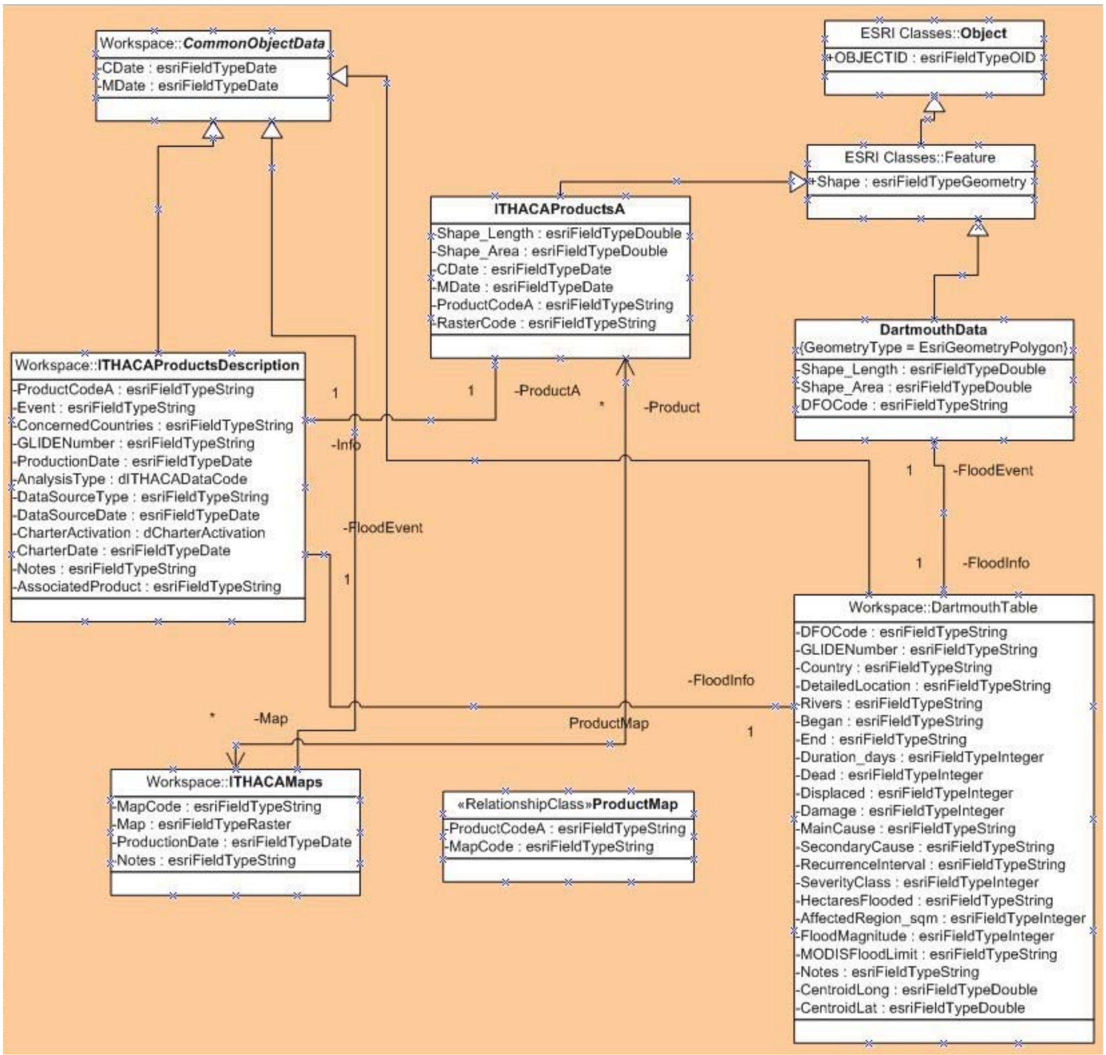


Figure 5.34 – UML Flood Analyses static structure.

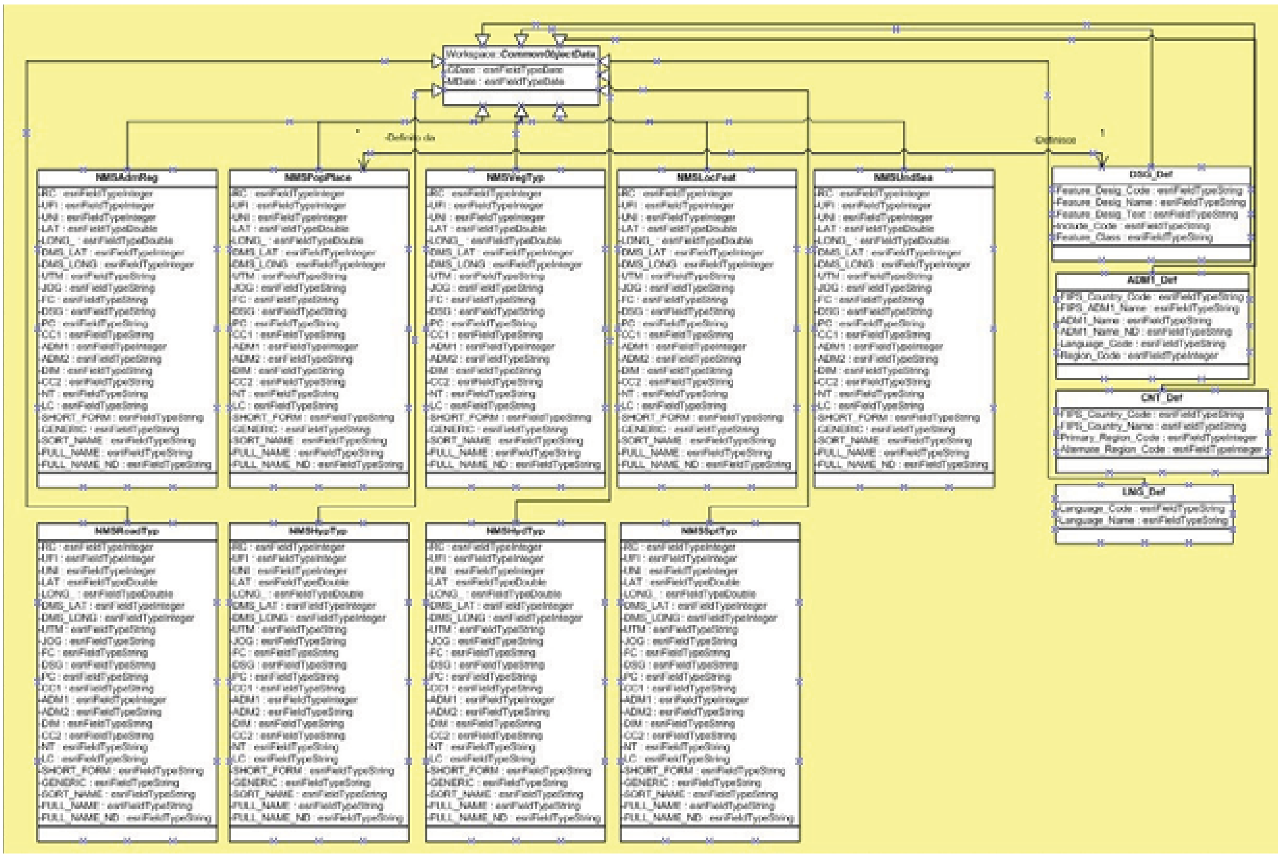


Figure 5.35 – UML Names static structure.

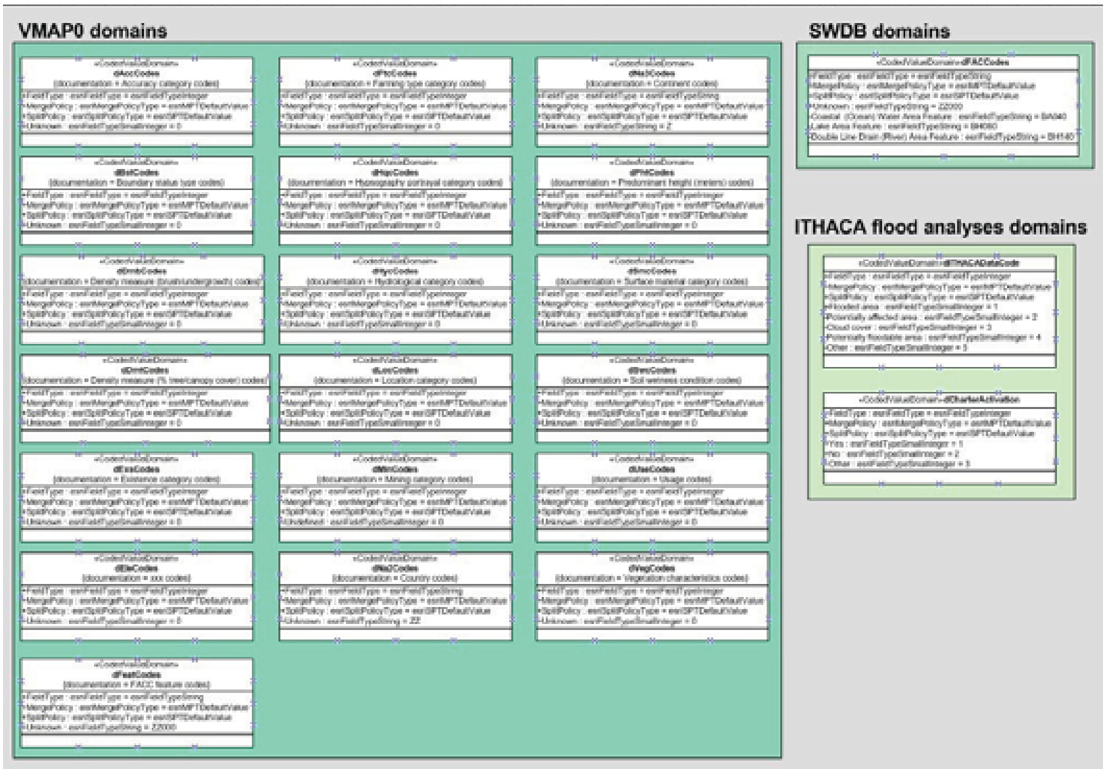


Figure 5.36 – UML implementation of geodatabase domains.

Figure 5.36 shows the UML implementation of the defined geodatabase domains. As you can see, a large part of them is derived by original VMap data, that constitute the primary source of geographic data contained in the planned geodatabase. A domain example is presented in figure 5.37.

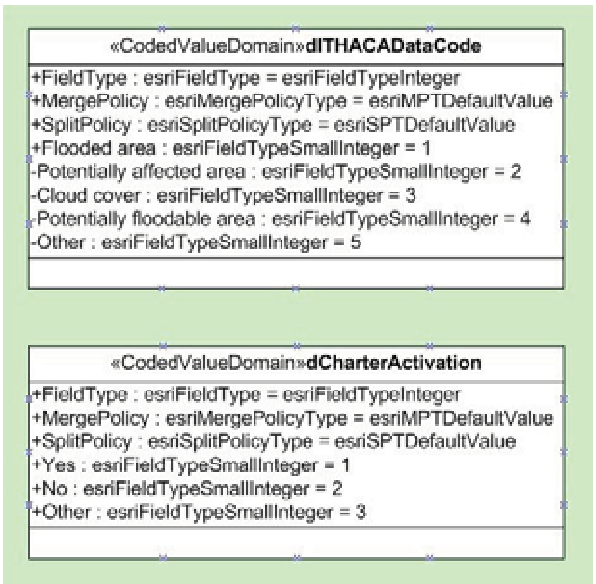


Figure 5.37 – UML implementation of geodatabase domains (example).

The raster component of the geodatabase

As shown in Tables 5.4, 5.5 and 5.6 some raster data have been included in the planned geodatabase. Generally, there are three ways to organize raster data in a geodatabase:

- *raster datasets*: is any valid raster format organized into one or more bands. Each band consists of an array of cells, and each pixel has a value. A raster dataset has at least one band. More than one raster dataset can be spatially appended (mosaicked) together into a larger, single, continuous raster dataset. When partially overlapping raster datasets are mosaicked, the overlapping area contains only one set of cell values. Storing the raster datasets individually is often the best method when the datasets are not adjacent to each other or are rarely used on the same project.
- *raster catalogs*: is a collection of raster datasets defined in a table format in which each record represents an individual raster dataset in the catalog. A raster catalog can be large and contain thousands of images. A raster catalog is typically used to display adjacent, fully overlapping, or partially overlapping raster datasets without having to mosaic them into one large raster dataset. Each raster in a catalog can have its own coordinate system that can be used to project each image for map on the fly. There are many uses for raster catalogs: raster catalogs can be used to hold a time series of raster images of the same area, to hold raster datasets partially or fully overlap preserving the common areas, and to record and manage additional attribute columns that describe each image in the catalog.
- *raster attributes:* rasters can also be an attribute of a feature in a feature class. This means that a field of type raster can exist as one of the columns within a feature class (geodatabase feature

classes only). This is similar to having a hyperlink of a file-based image in a field, except the raster image is stored and managed within the geodatabase.

The choice among these solutions is one of the major design decisions in managing raster data. The features of considered raster data, described in Chapter 4, dictated which approach to use. Because of the necessity to store large raster data and satellite images time series, we selected raster catalog format in order to collect needed imagery in the geodatabase.
The raster component of the geodatabase has not been modelled within the UML Class Diagram, but it has been added after the ESRI ArcInfo geodatabase creation.

Topology rules and reference system definition

The ArcInfo UML model do not support the spatial reference system definition and the topological rules implementation. Therefore they are defined after the ESRI ArcInfo geodatabase creation.
According to the global nature of the planned geodatabase, a global reference system has been defined for all contained geographic objects. Its features are presented in the following Table 5.7.

Table 5.7 – The geodatabase reference system

Geographic Coordinate system (GCS) - WGS84	
Angular unit	degree
Prime meridian	Greenwich
Datum	Ellipsoid: WGS84

When we created the first version of the geodatabase, no topological rules have been yet proposed. At present, this topic still needs more discussions among the geodatabase users, in ITHACA and in the WFP. Therefore, no topological rules have been implemented.

5.3 Geodatabase construction

5.3.1 Logical and physical database modelling: basic concepts

In the whole database design process, after the conceptual design phase, logical and physical design are performed (see Chapter 3, paragraph 3.2.3).
During the logical design phase, the conceptual schema is first translated from the high level data model into the data model of the chosen DBMS. Therefore, the DBMS typology must be defined. This step of the logical design process is system independent, but DBMS data model dependent.
Database Management Systems (DBMSs) can be classified according to the data model they use. In the past, there were the *hierarchical* and *network* data models, which have now being replaced by three models, namely, *relational*, *object-oriented* and *object-relational:*

- *The relational model*. Relational databases are based on the relational model. The relational model was first defined in 1970, when Codd introduced the idea of using the mathematical concept of relations (in the set theory) as the means for data model. The relational data model is based on mathematical *relations*, where data is logically structured in tables. The Entity-Relationship (ER) concepts are used in relational databases. Entities and relationships among them are stored in *tables*. They are tabulated into rows and columns. In the formalisms of the relational model a table row is called a *tuple*, while a column heading is an *attribute*. The table is called *relation*. Degree of a relation is the number of attributes it contains. Cardinality is the number of tuples it contains. Relations between entities are implemented through foreign keys. The foreign key is a code, which is stored in a table and refers to rows in another table. The primary key of the linked table is stored in the foreign code column of the other table.

Relational databases can also be defined as normalized relations. Normal forms are guidelines for relational database design that increase the consistency of data. In the relational database systems methods and integrity constrains, which are defined in the conceptual data model, are realized through external transactions.

This model is implemented in many commercial systems. At the moment, very common relational DBMSs (RDBMSs) are: DB2 and Informix Dynamic Server (IBM), Oracle and Rdb (Oracle), and SQL Server and Access (Microsoft). The relational database model presents operations on algebraic expression language in order to realize data query and manipulation. With data manipulation languages the types of operations allowed on the data is defined through relational algebra. The Structured Query Language (SQL) is the international standard query language of relational databases.

- *The object-oriented model*. Object-oriented database management systems (OODBMS) integrate object orientation and database functionality. An object-oriented database is characterised by having an object-oriented logical data model and by using an object-oriented programming language as its principal interface. One of the advantages of an object oriented data management system is that real-world phenomena can be modelled closer to reality with non-atomic objects. Object models could be structured around conceptual objects rather than geometrical properties. With the object-oriented approach, geometry can be modelled like other information. Explicit stored spatial information, which are required in GIS, can be realized without redundancy.

  Object-oriented databases were introduced as a means of overcoming the limitations of the relational model and to handle more complex application requirements, which the relational model cannot capture.
- *The object-relational model*. The object-relational approach is a compromise between the concepts of the object-oriented and relational models. In this approach, object-oriented features are incorporated into relational databases or in a manner that utilises the power of object orientation while maintaining the full functionality of the relational model. The object-relational database management system is an extended relational database supporting abstract data types, procedures, encapsulation and complex objects. Limited operations of relational databases can be extended, defining new operations and methods. The advantages of relational databases such as standardized query languages, security, versioning and referential integrity facilities can still be used. The object-oriented messaging approach is not required between objects. Additionally, the enhanced spatial query possibilities of relational databases are easily performed, since the complex objects are stored in tables associated with their object identifiers.

Different DBMSs implement the same data model using different constraints and modelling constructions. Therefore, the database schema obtained in the first step of the logical design process must be then adapted in order to make them consistent to the specific features of the chosen DBMS data model.
Finally, in the physical design phase, database specifications are designed in terms of physical storing structures and records and indices positioning. Database and applications are then implemented and tested.

5.3.2 Geodatabase logical and physical model development

As aforementioned, according to the analyses carried out during the needs assessment phase, we decided to implement the defined geographic database model as an ESRI ArcSDE Enterprise Geodatabase based on Oracle 10g as DBMS. For the implementation, a commercial GIS software package has been used, in conjunction with a commercial database system. Therefore, the basic structure of the logical schema (e.g., a relational data schema) and the entire physical schema were already predetermined. So, the main task we accomplished was to prepare the conceptual schema that properly described the entire geographic database and that was suitable for translation into the logical schema of the proposed GIS and database software.
The ESRI ArcInfo geodatabase can be thought, in our application, as the logical data model adopted in order to translate the developed conceptual schema into a logical schema. Some features of the geodatabase logical and physical data model are presented here.

The ArcInfo ArcSDE geodatabase is a collection of geographic datasets of various types, stored in a relational database such as Oracle, Microsoft SQL Server, IBM DB2, or IBM Informix.
The ArcInfo geodatabase storage model is based on relational principles. Key concepts are:

- data are organized into tables;
- tables contain rows;
- all rows in a table have the same columns;
- each column has a type, such as integer, decimal number, character, date, and so on;
- relationships are used to associate rows from one table with rows in another table. This is based on a common column in each table;
- relational integrity rules exist for tables. For example, each row always shares the same columns, a domain lists the valid values or value ranges for a column, and so on;
- a series of relational SQL functions and operators are available to operate on the tables and their data elements;
- the SQL operators are designed to work with the generic relational data types, such as integers, decimal numbers, dates, and characters.

As already discussed, DBMS tables can be used to store *feature classes* where each row in the table represents a geographic object or *feature*. The columns in each row represent various characteristics or properties of the feature, and one of the columns holds the feature geometry.
A homogeneous collection of common features, each having the same spatial representation, such as a point, line, or polygon, and a common set of attribute columns, is referred to as a *feature class* and is managed in a single table.

Raster and imagery data types are managed and stored in relational tables as well. Raster data is typically much larger in size and requires a side table for storage. The raster is tiled into smaller tiles, or blocks, and each block is stored in individual rows in the separate block table.
SQL operates on the rows, columns, and types in tables. The column types (the numbers, characters, dates, BLOBs, spatial types, and so on) are objects in the SQL algebra. The DBMS manages these simple data types and tables, while additional application logic implements more complex object behaviour and integrity constraints.
Particularly, the geodatabase storage in a DBMS contains two sets of tables, *system tables* and *dataset tables* (also called *user-defined tables*):

- the geodatabase *system tables* keep track of the contents of each geodatabase. They also contain and manage all the metadata required to implement geodatabase properties, data validation rules, and behaviours;
- each dataset in the geodatabase is stored in one or more tables. The *dataset tables* work with the system tables to manage data.

A *shape* column in each table is used to hold the geometry or shape of the features. The shape column holding the geometry is typically one of two column types:

- a spatial column type, if the DBMS supports it;
- a *binary large object* (BLOB) column type.

Each of aforesaid DBMS types has a slight variation in the set of tables and columns used to store and manage a geodatabase. Therefore, the type of DBMS adopted to store the geodatabase will impact the physical storage schema.
Figure 5.38 shows the ArcSDE Geodatabase system tables, which maintain information about and manage the data in the geodatabase.

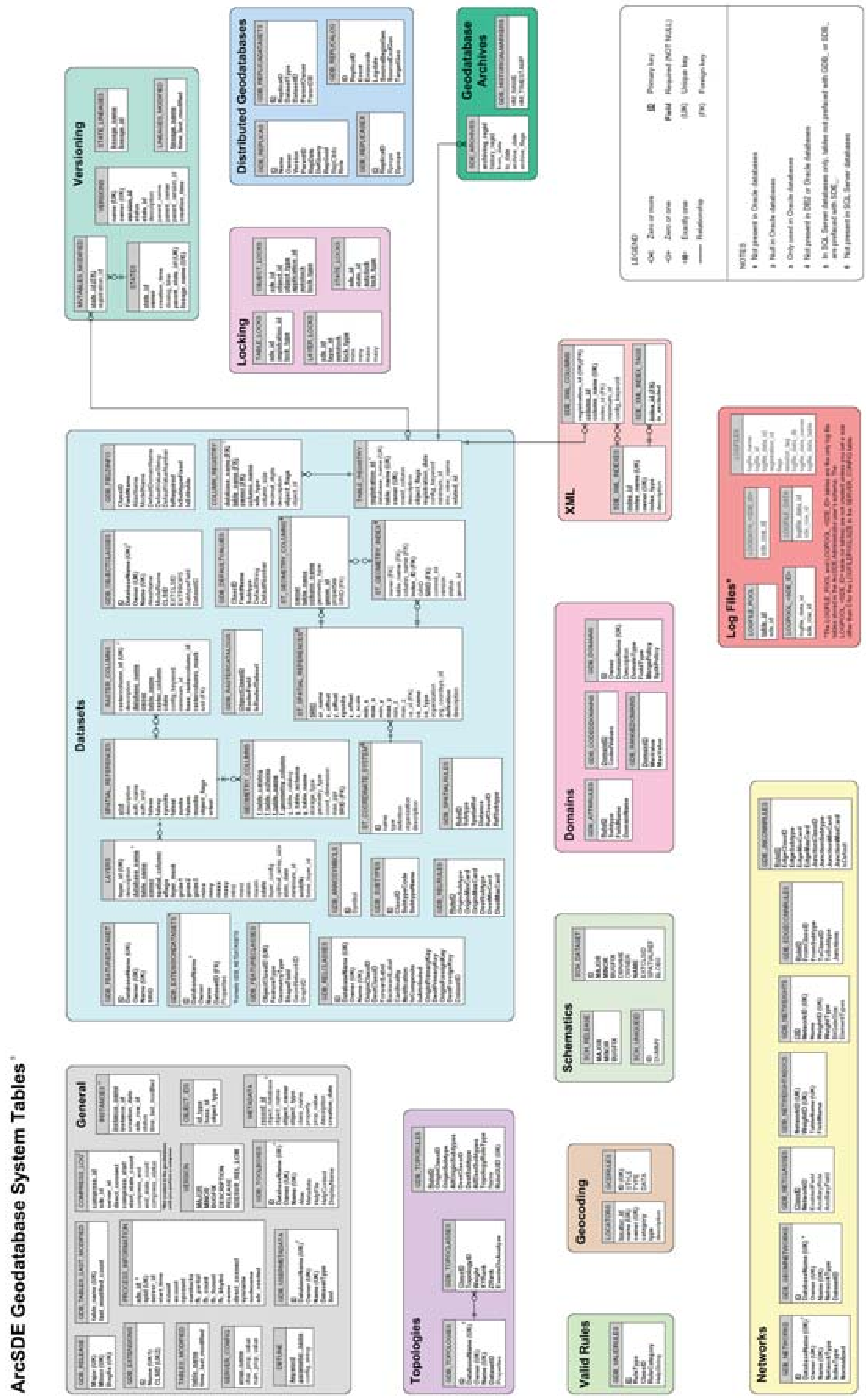


Figure 5.38 – ArcSDE Geodatabase system tables.

The ArcInfo geodatabase is object-relational. The geodatabase employs a multilevel application architecture by implementing advanced logic and behaviour in the *application level* on top of the *data storage level* (managed within DBMS). The geodatabase application logic includes support for a series of generic GIS data objects and behaviours such as feature classes, raster datasets, topologies, and networks, among others. This multilevel geodatabase architecture is sometimes referred to as an *object-relational model*.

Responsibility for management of geographic datasets is shared between ArcGIS software and the generic DBMS. Certain aspects of geographic dataset management, such as disk-based storage, definition of attribute types, associative query processing, and multiuser transaction processing, are delegated to the DBMS. The GIS application retains responsibility for defining the specific DBMS schema used to represent various geographic datasets and for domain-specific logic, which maintains the integrity and utility of the underlying records.

This could be considered a multilevel architecture (application and storage), where aspects related to data storage and retrieval are implemented in the data storage (DBMS) level as simple tables, while high-level data integrity, management of spatial relationships, geographic behaviour, and information processing functions are retained in the application and domain software (ArcGIS).

All ArcGIS applications interact with this generic GIS object model for geodatabases, not with the actual SQL based DBMS instances. The ArcSDE software provides a gateway for ArcGIS to interact with the DBMS.

The ArcSDE geodatabase is multiuser. ArcSDE is the multiuser data access extension to ArcInfo. In ArcSDE geodatabases, there is no limit to size or numbers of users. ArcSDE geodatabases support multiple users and editors and support long transactions in GIS using *versioning* approach.

Transactions are packages of work that make changes to databases. GIS databases, like other database applications, must support update transactions that enforce data integrity and application behaviour. In many cases, users can exploit the DBMS's transaction framework for managing edits and updates to geodatabases. The geodatabase mechanism for managing these and many other critical GIS workflows is to maintain multiple states in the geodatabase and, most important, to do so while ensuring the integrity of the geographic information, rules, and behaviour. This ability to manage, work with, and view multiple states is based on *versioning*. As the name implies, versioning explicitly records versions of individual features and objects as they are modified, added, and retired through various states. Each version explicitly records each state of a feature or object as a row in a table along with important transaction information. Versions enable all transactions to be recorded as a series of changes to the database through time. This means that various users can work with multiple views or states of the geodatabase. Any number of users can simultaneously work with and manage multiple versions.

Geodatabase creation

The geodatabase logical schema includes the definitions, integrity rules, and behaviour for each geographic dataset. These include properties for feature classes, topologies, networks, raster catalogs, relationships, domains, and so forth.

The translation from the developed UML conceptual model to the logical one has been performed using valuable software tools, according to steps shown in Figure 5.39. First of all, the UML conceptual schema of the geodatabase has been exported in the XMI (*XML Metadata Interchange*) format. This format is a

standard ISO format which specifies how the UML model must be saved in a XML file, used for the information exchange. ArcInfo can then use it to generate the geodatabase schema. The XML file generation has been carried out using the *ESRI XMI Export* tool available in MS Visio.
Before exporting to XMI format, the whole model has been checked for semantic errors by using Visio UML *Semantics Checker* tool. This prevents lots of errors during the automatic schema generation.
Afterwards, we created in ArcCatalog a new empty geodatabase and we used ESRI *Case Schema Creation* tool to generate the geodatabase schema form the UML model. Table 5.8 shows how ArcInfo maps the elements of the object model into relational database elements.

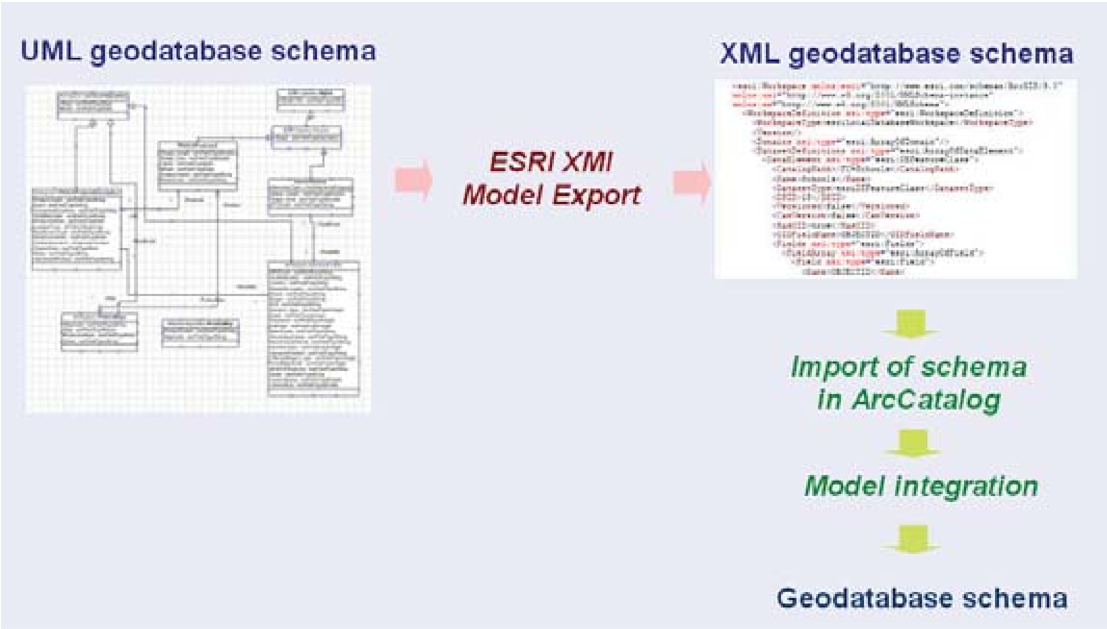


Figure 5.39 – Creation of the geodatabase schema.

Table 5.8 – Object-relational mapping (source: [M. Zeiler, 1999])

Object model element	**Relational database element**
Object	Row
Attribute	Column, Field
Class	Table

In our case, the GIS platform used for implementing the whole system has been the ArcInfo version of ArcGIS 9.2 software.
In the subsequent model integration phase, we developed raster catalogs, raster features, and we defined the reference system according to the choices presented in the previous paragraph. When applicable, also topological rules and other constraints are defined in this phase.
Finally, Figure 5.40 shows a global view of the created geodatabase in ArcGIS ArcCatalog, with its content. Further descriptions of the geodatabase content will be provided in Chapter 6.

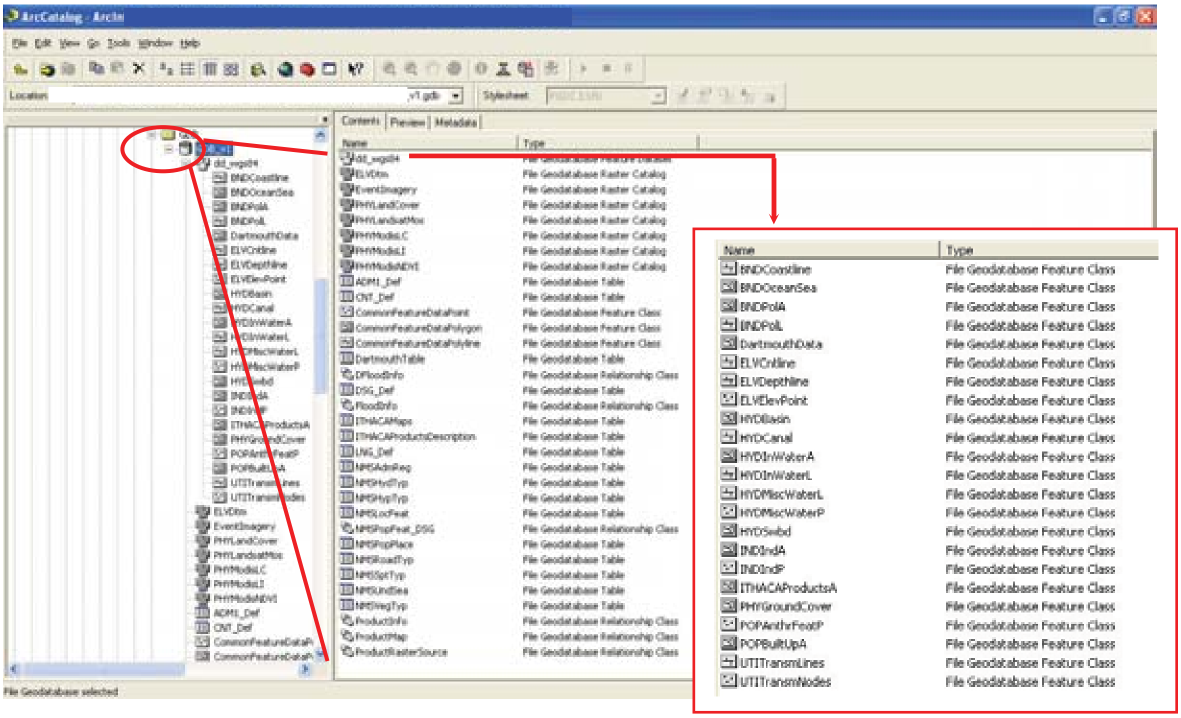


Figure 5.40 – The created geodatabase in ArcCatalog.

5.3.3 Geodatabase implementation

The details concerning the implementation of the planned geodatabase are out of the scope of the conducted research work, principally aimed at performing the geographic database design. Nevertheless, for completeness, a document about this topic, produced by ITHACA's *Geodatabase/SDI user group*, is supplied in Appendix.

As discussed in Chapter 2, the planned geodatabase not only contains all data used by all ITHACA's user groups, but also constitutes the first of the activities needed in order to develop a *Spatial Data Infrastructure* for the WFP. Therefore, in the documentation proposed in Appendix the reader will also find information about the choices concerning the development of the whole system architecture, based on the planned geodatabase, which will grant the correct development of a SDI. These choices deal mainly with:

- the definition of the different read/write privileges for the various kinds of geodatabase users;
- the definition of the data distribution schema and of the system architecture;
- the development of valuable Web services and applications, based on the planned geodatabase, allowing an interoperable level of access and management of the database and the development of tools in support to Early Warning and Early Impact activities.

6

DATABASE CONTENT AND APPLICATIONS

An overview of the spatial features contained in the developed geodatabase are here presented. Moreover, an example showing the validity of the implemented solution in usual ITHACA Early Impact activities are also presented.

6.1 Geodatabase content

The following Figure 6.1 summarizes the feature classes, tables and raster data used to implement the conceptual model previously described, classified according to the different defined themes.

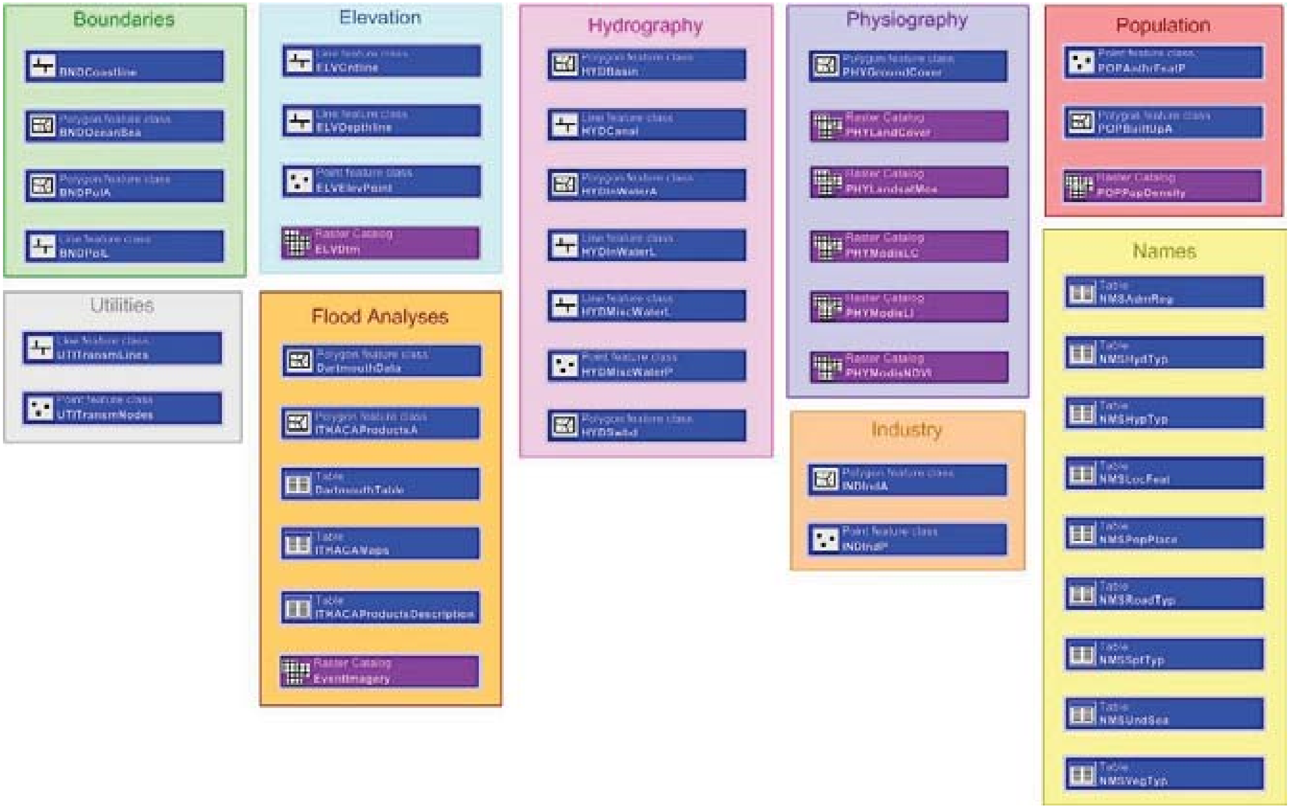


Figure 6.1 – Overview of the developed geodatabase content.

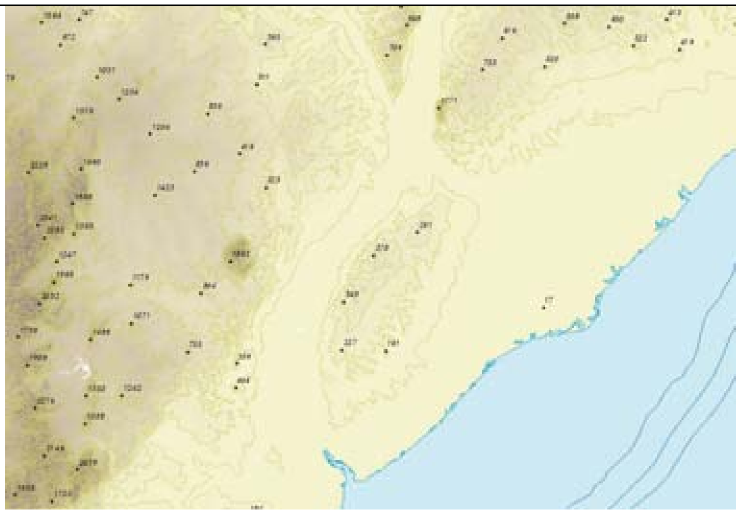
In the next figures, some examples of the content of each theme are proposed.

Boundaries

Classes	Data type	Example	Legend
BNDCoastline ***BNDOceanSea*** ***BNDPolA*** ***BNDPolL***	Polyline Polygon Polygon Polyline		Administrative boundary Level 2 Boundary Level 2 Boundary (in dispute) Level 1 Boundary Level 1 Boundary (in dispute) Level 0 Boundary Level 0 Boundary (in dispute)

Boundaries data are contained in the geodatabase as both line and polygon features. Linear feature classes are used primarily for cartographic purposes. The polygon feature classes are used in maps for cartographic labelling. They are also used in some Early Impact activities, as described in Chapter 4.
In the shown example, these data have been symbolised based on their boundary type (national or sub-national, first and second level, regions, provinces and districts), according to the UNGIWG *Map Production Guidelines*. In order to exploit the geodatabase data content in a more effective way, symbolic representations of the different feature classes can be defined and saved as ArcGIS layer files (.lyr). *Map layers* specify how datasets must be drawn. A layer is a set of rules for displaying and working with datasets; it includes symbol assignments, classifications, labeling rules, and other used map properties.

Elevation

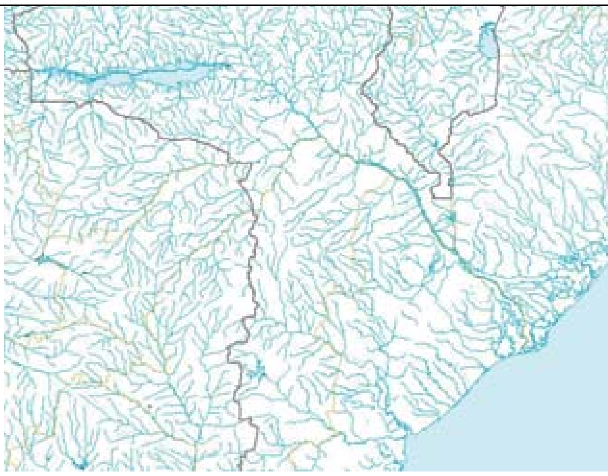
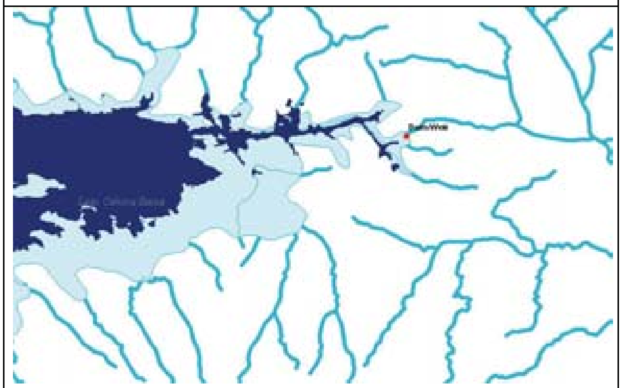
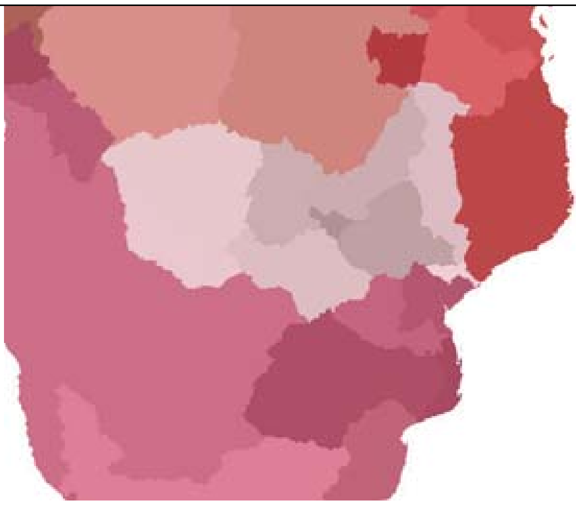
Classes	Data type	Examples	Legend
ELVCntline ***ELVDepthline*** ***ELVDtm*** ***ELVElevPoint***	Polyline Polyline Raster Point		ELVElevPoint ELVCntline ELVDepthline

Vector topographic data include elevation points and contour lines. Elevation points are usually benchmarks for ground control reference or elevation of significant features. What constitutes a significant feature can vary with the surrounding terrain. Contours are lines of constant elevation.
Raster topographic data is constituted by a satellite derived Digital Terrain Model (*Shuttle Radar Topography Mission*, SRTM, of NASA) with a spatial resolution of 90 m.

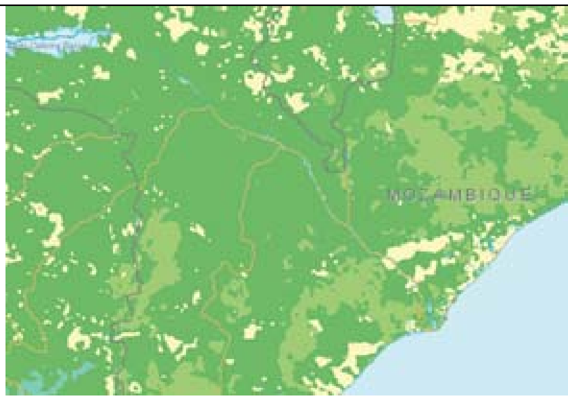
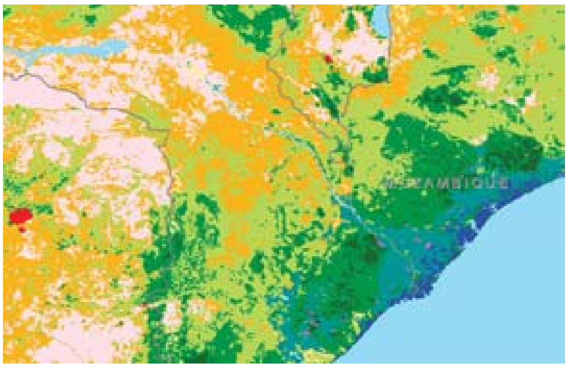
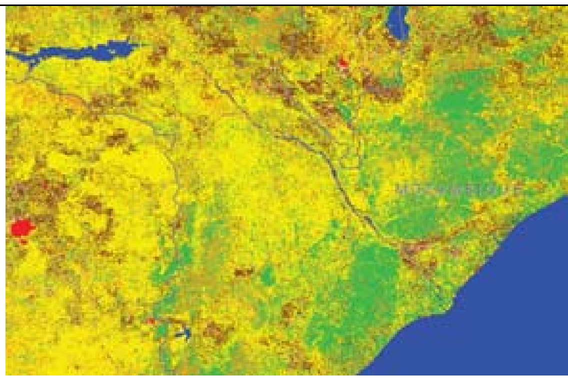
Hydrography

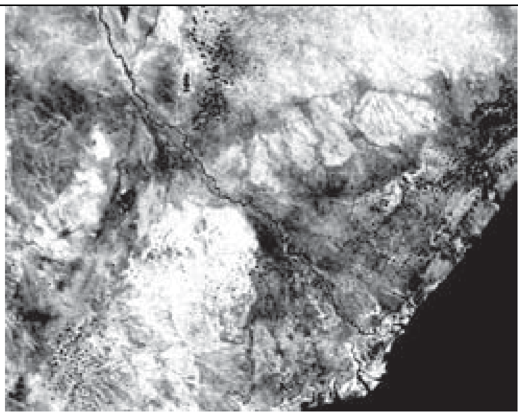
Hydrographic features include:

- point features, such as Springs, Water-holes, Waterfalls, Dams, Weirs, Locks and Sluice Gates;
- linear objects, such as Aqueducts, Canals, Flumes or Penstocks *(HYDCanal),* Water courses (rivers and streams), perennial e non-perennial (*HYDInWaterL*), and miscellaneous water elements;
- areal objects, such as Inland Water, Wetlands and Lands subject to the Inundation (*HYDInWaterA* and *HYDSwbd*). The *HYDInWaterA* feature class contains data derived from the VMap0 source, with map scale of 1:1,000,000, while the *HYDSwbd* contains STRM *Water Body Dataset* (SWBD) derived from the *Shuttle Radar Topography Mission* (STRM) of NASA.
- drainage basins data, subdivided in six levels.

Classes	Data type	Examples	Legend
HYDCanal ***HYDInWaterA*** ***HYDInWaterL*** ***HYDMiscWaterL*** ***HYDMiscWaterP*** ***HYDSwbd***	Polyline Polygon Polyline Polyline Point Polygon	 	
HYDBasin	Polygon		**Drainage Basins (I and II levels)**

Physiography

Classes	Data type	Examples	Legend
PHYGroundCover	Polygon		Bog Cropland Grassland Ground Surface Element Land Subject to Inundation Marsh/Swamp Rock Scrub/Brush Trees Tundra
PHYLandCover	Raster		Artificial surfaces and associated areas Bare Areas Cultivated and managed areas Herbaceous Cover, closed-open Irrigated Agriculture Mosaic: Cropland / Shrub and/or grass cover Mosaic: Cropland / Tree Cover / Other natural vegetations Mosaic: Tree Cover / Other natural vegetation Regularly flooded shrub and/or herbaceous cover Shrub Cover, closed-open, deciduous Shrub Cover, closed-open, evergreen Snow and Ice Sparse herbaceous or sparse shrub cover Tree Cover, broadleaved, deciduous, closed Tree Cover, broadleaved, deciduous, open Tree Cover, broadleaved, evergreen Tree Cover, burnt Tree Cover, mixed leaf type Tree Cover, needle-leaved, deciduous Tree Cover, needle-leaved, evergreen Tree Cover, regularly flooded, fresh water Tree Cover, regularly flooded, saline water Water Bodies
PHYModisLC	Raster		Fill Value Water Evergreen Needleleaf Forest Evergreen Broadleaf Forest Deciduous Needleleaf Forest Deciduous Broadleaf Forest Mixed Forest Closed Shrubland Open Shrubland Woody Savannas Savannas Grasslands Permanent Wetlands Croplands Urban and Built-Up Cropland/Natural Vegetation Mosaic Snow and Ice Barren or Sparsely Vegetated Unclassified
PHYModisLI	Raster		

Classes	Data type	Examples
PHYLandsatMos	Raster	 
PHYModisNDVI	Raster	

Physiographic features include:

- *Land cover data*: this kind of data is used for distinguishing broad categories of vegetation and surface material. In the developed geodatabase three different types of land cover data have been stored. The *PHYGroundCover* class contains land cover data derived by the VMap0 source and this is in a vector format (polygon features). The *PHYLandCover* class contains a land cover raster dataset (1 km of spatial resolution) distributed by the *European Commission Joint Research Ce*ntre (JRC), based on the SPOT-4 VEGETATION VEGA2000 satellite data. The used legend is the Land Cover Classification System of the FAO (LCCS). The LCCS is a comprehensive, standardized, a priori classification system. The LCCS system describes land cover according to a hierarchical series of classifiers and attributes. These separate vegetated or non-vegetated surfaces; terrestrial or aquatic/flooded; cultivated and managed; natural and semi-natural; anthropic; cover; height; spatial distribution; leaf type and phenology. The use of the LCCS system allows a map legend to be progressively more detailed for regional, and in some cases, national level uses. Due to its hierarchical structure, it is possible to translate the regional classification into a more general, or global, legend.

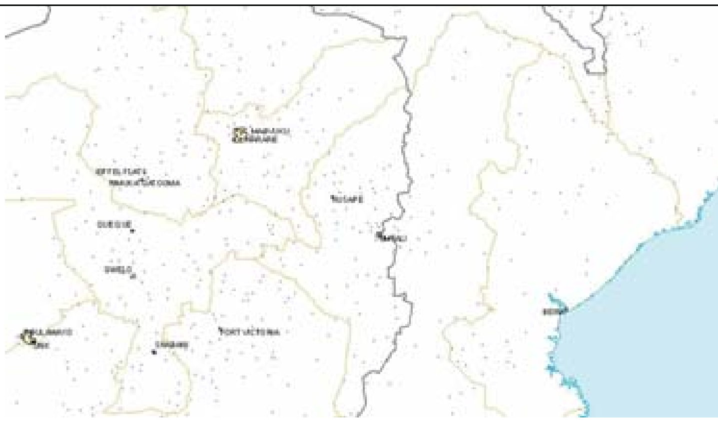
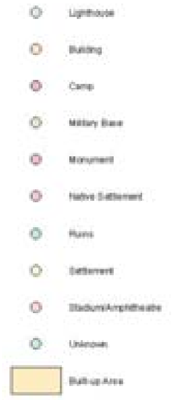
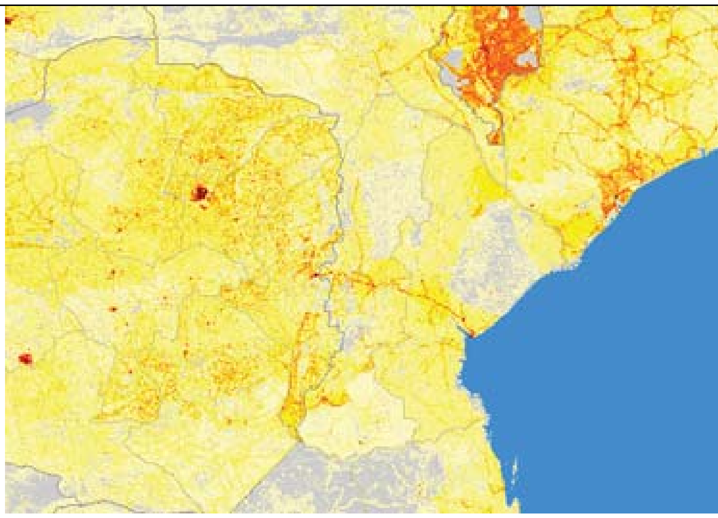
  Finally, a third land cover class, the *PHYModisLC* one, contains land cover data derived from the MODIS sensor (NASA Terra and Aqua satellite platforms; 1 km – 5 km of spatial resolution). The used MODIS land cover scheme identifies 17 classes of land cover defined by the International Geosphere-Biosphere Programme (IGBP), which include 11 natural vegetation classes, 3 developed land classes, one of which is a mosaic with natural vegetation, permanent snow or ice, barren or sparsely vegetated, and water. The major advantage of this dataset is its updating frequency: as a matter of fact this dataset is produced yearly. In the first developed version of the geodatabase only the most up-to-date version of this data has been stored. Nevertheless, this kind of data will require future updating operations, according to the availability of new source data.

- *Satellite imagery*: there are several satellite derived raster datasets in the developed geodatabase. In the physiographic theme, it has been included satellite imagery collected in order to supply general background layers in some ITHACA map products and applications. Considered satellite data are an Orthorectified Landsat TM Mosaic (false colour synthesis, 30 m of spatial resolution; it dates back to year 2000) and satellite imagery acquired by the MODIS sensor (false and real colour synthesis; 250m, 500m and 1 km of spatial resolution) which is downloadable free-of-charge from dedicated web-sites. The MODIS sensor acquires images of the whole Earth surface with about a daily frequency. Therefore, it will constitute a very up-to-date source of data in the geodatabase maintenance phase.
- *NDVI data*: the NDVI (Normalized Difference Vegetation Index) values indicate the amount of green vegetation present in the scene. NDVI data are obtained transforming multispectral data, using the Red and the Near Infrared bands, into a single band representing vegetation distribution. Valid cell values fall between –1 and +1, and higher NDVI values indicate more green vegetation.

All raster data included in the geodatabase, after the necessary pre-processing operations, have been stored in raster catalog datasets. A raster catalog is a collection of raster datasets defined in a table format in which each record defines an individual raster dataset in the catalog. In particular, using raster catalogs allows to manage collections of images as integrated sets and constantly to add new images to the geodatabase.
Some of the most benefits of storing raster data in the geodatabase are:

- huge datasets requiring multiuser access: when many users are accessing the same raster files simultaneously, better performance is possible from a centralized database than from file-based system. This is especially important with massive raster datasets. The chosen ArcSDE geodatabase solution supports huge data volumes, fast display and good multiuser performance.
- data security: a DBMS permits multiple security levels to be established and enforced. Users can be given access to the imagery relevant to the job on which they are working.
- data query: a DBMS enables a common query environment. Queries can help to locate all data related to an area for a particular subject matter or during a particular time period, depending on the metadata provided with the datasets. With raster catalogs, additional attributes can be associated with each dataset to support more extensive queries.

Population

Classes	Data type	Examples	Legend
POPAnthrFeatP ***POPBuiltUpA***	Point Polygon		
POPPopDensity	Raster		

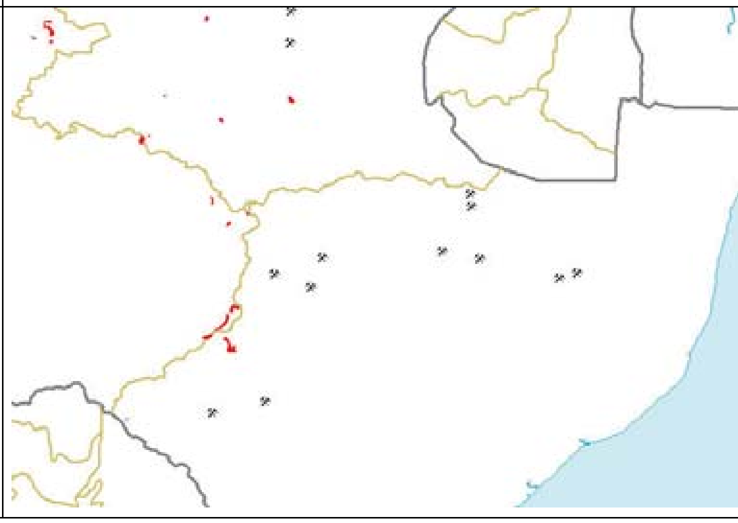
Vector population data includes built-up areas and anthropic feature points. Raster populations data is constituted by the LandScan 2005 global population database (1 km of geometric resolution), developed by the *Oak Ridge National Laboratory* (ORNL) for the *United States Department of Defence* (DoD). This dataset contains global population distribution (the dataset values are people per cell).

Utilities

Classes	Data type	Examples	Legend
UTITransmLines ***UTITransmNodes***	Polyline Point		

This theme includes transmission nodes, such as Power Plants, Substation/Transformer Yards, Pumping Stations and Communication Buildings, and lines, such as Pipelines (below surface/underground and on ground surface).

Industries

Classes	Data type	Examples	Legend
INDIndA ***INDIndP***	Polyline Point		

This theme includes point features, such as Mines/Quarries, Oil/Gas Facilities, Tanks, Water Towers, Depots, and Processing Plant/Treatment Plant, and areal features, such as Mines/Quarries, Oil/Gas Fields and Salt Evaporators.

Flood Analyses

This theme contains the following objects:

- *ITHACAProductsA:* results of flood events analyses performed by ITHACA during Early Impact activities (polygon features). Usually this feature class contains flooded areas, affected areas and wet areas, as detected from satellite base imagery acquired during flood events;
- *ITHACAProductsDescription:* description of ITHACA flood products and information about related flood events (table);
- *DartmouthData:* historical areas hit by flood events (polygon features). These data are produced by the Dartmouth Flood Observatory;
- *DartmouthTable:* description of DFO historical data (table);
- *ITHACAMaps:* maps produced by ITHACA during flood events (table containing raster maps with related information);
- *EventImagery:* base satellite imagery used in the flood analyses performed by ITHACA (stored in raster catalog structures). Common used satellite data are those acquired by the MODIS sensor. Other kinds of satellite data are available only for catastrophic events which required *International Charter* system activation.

Some instances of the content of this theme will be proposed in the following paragraph, while an overview of proposed data model is shown in Figure 6.2.

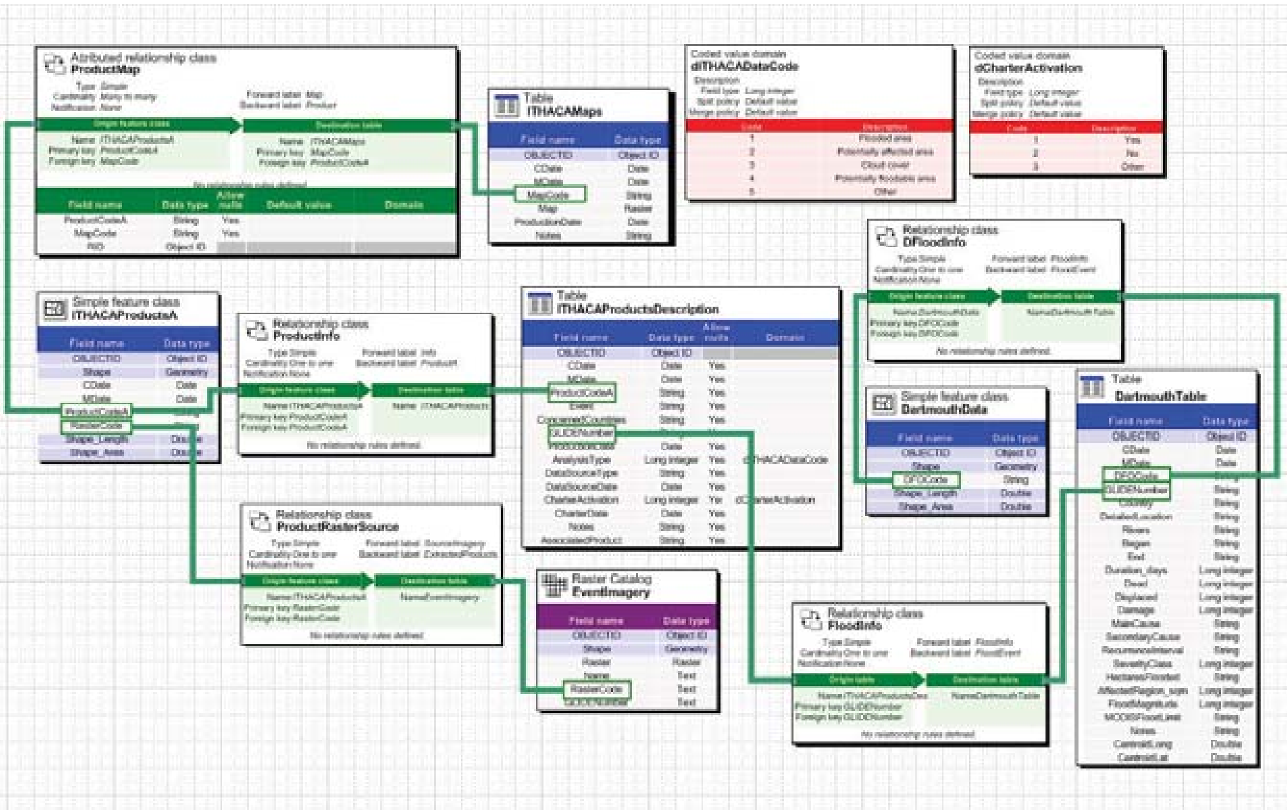


Figure 6.2 – Data model developed for the Flood Analyses dataset.

Names

This dataset contains many table objects:

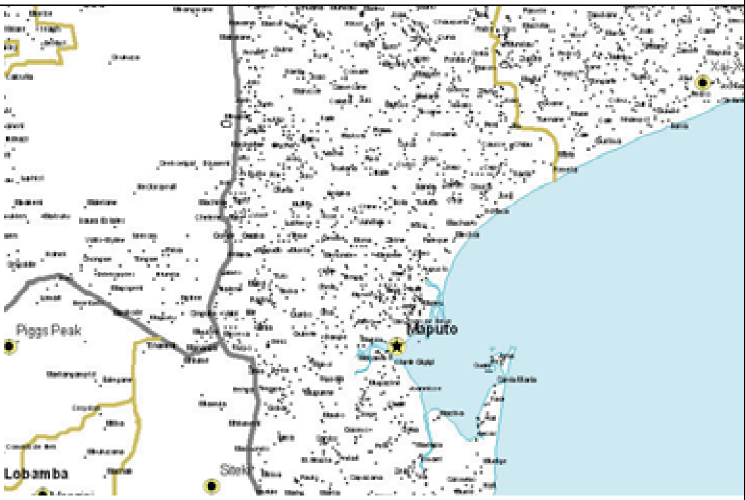
- *NMSAdmReg:* names of administrative region features;
- *NMSHydType:* names of hydrographic type features;
- *NMSHypType:* names of hypsographic type features;
- *NMSLocType:* names of locality or area type features;
- *NMSPopPlace:* names of populated place features;
- *NMSSptType:* names of spot type features;
- *NMSStrType:* names of street, highway, road or railroad type features;
- *NMSUndType:* names of undersea type features;
- *NMSVegType:* names of vegetation type features.

The *GEOnet Names Server (GNS) Database* developed by the *US National GeoSpatial-Intelligence Agency* (NGA) is the source of this dataset. The GNS database provides the baseline for many, if not all, of the gazetteers available either commercially or from private sources.

These table objects containing coordinates and other data can be converted to point events. The point events created from the table behave just like a point feature class, and can be symbolized and labelled using attributes in the table.

The full encoding attribution for each location includes: latitude and longitude in degrees: minutes: seconds (DMS) and decimal degrees; diacritical and non-diacritical representations of the name; sort name attributes; two topical feature classifications, e.g. a generalized hydrographic class followed by a specific subclass such as reservoir.

The following table shows an example of populated places contained in the *NMSPopPlace* table.

Classes	Data type	Examples	Legend
NMSPopPlace	Table with X,Y fields		

6.2 An application example: map production using the geodatabase

As already discussed in Chapter 4, the developed geodatabase supports all the activities performed by ITHACA's units. Nevertheless, among these activities, cartographic production operations, carried out by the Early Impact unit, are those that most take advantage from the use of the geodatabase.
This unit mainly deals with support supply of flood disasters mitigation operations conducted by the WFP. The support in early impact activities for flood events is conceived mainly in provision of *map products* showing the impact of the flood on population and on infrastructures. In operative terms, map-supporting is required at two different stages: at the first state of alert and just after the event. In both cases the response time after help request triggering is short, lasting generally from 24 to 48 hours. The general workflows adopted in Early Impact unit's activities have been described in detail in Chapter 4. In order to correctly face the activities required when a flood emergency starts up, activities provided in the previously described workflows are carried out by two sub-units (the *Remote Sensing user group* and the *GIS/Map Production user group*) which produce, obtain and modify several data and exchange them each other through the developed geodatabase.
The map layout arranging and production operations performed by the *GIS/Map Production sub-unit* will be discussed here because, evidently, they are those which may realize the maximum profit from the structuring and the organization of the used geographic data within the central geodatabase. Particularly, we will consider the activity of production of a standard map product. For instance, the production of a *Flood Monitoring Map,* which has been described in the Map Sample Form N°2 (see Figure 4.20) proposed in Chapter 4. The software environment used to perform map production activities is the ArcMap ArcGIS 9.2 one.

The primary need of ITHACA's map production activities, is to have technologies and instruments allowing the quick carrying out of the necessary activities, according to their emergency response purposes. Therefore, with the aim of automate much, if not all, map activities, the *GIS/Map Production user group* has developed:

- several map templates (Figure 6.3) containing commonly used map layout elements, such as:
  - map title, logos, contact information, and version date;
  - map window, latitude-longitude reference graticule with labels;
  - geodetic datum and projection parameters descriptions;
  - north arrow and scale text and bars;
  - legend and used data disclaimers.

  These map layout elements have been chosen according to the UNGIWG OCHA *Map Production Guidelines* (see Chapter 4) and can be used with any number of the geodatabase datasets.
- several map layers that implement the cartographic specification rules for classes contained in the geodatabase. A critical part of each dataset is the specification for how it is symbolized and rendered in maps. Map layers are typically defined as layer properties, which specify how features are assigned to map symbology (colors, fill patterns, line and point symbols) and labeling specifications. Feature layers, defined according to the UNGIWG OCHA *Map Production Guidelines,* constitute an

important geodatabase model completion. All data examples shown in paragraph 6.1 have been defined using the developed map layers.

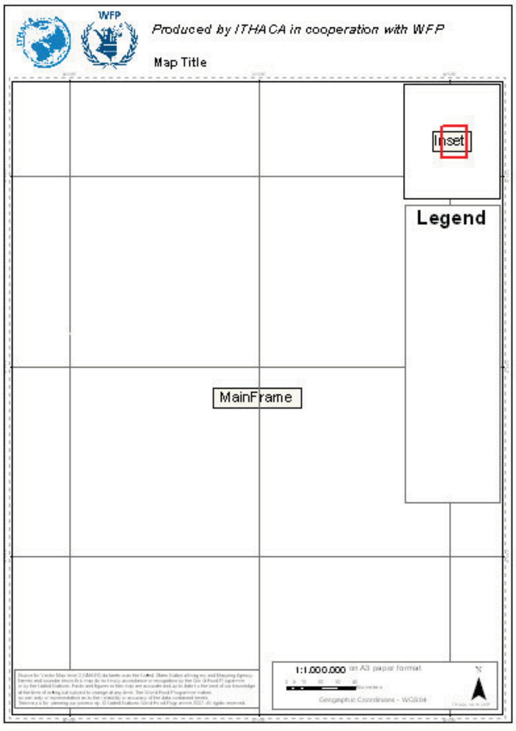


Figure 6.3 – An example of map template developed by the ITHACA's GIS/Map Production unit.

Using a suitable *spatial database connection* users of the *GIS/Map Production sub-unit* can load into a previously defined ArcMap map document all geographic data they need in order to produce the desired *Flood Monitoring Map* product (see Figures 6.4 and 6.5).

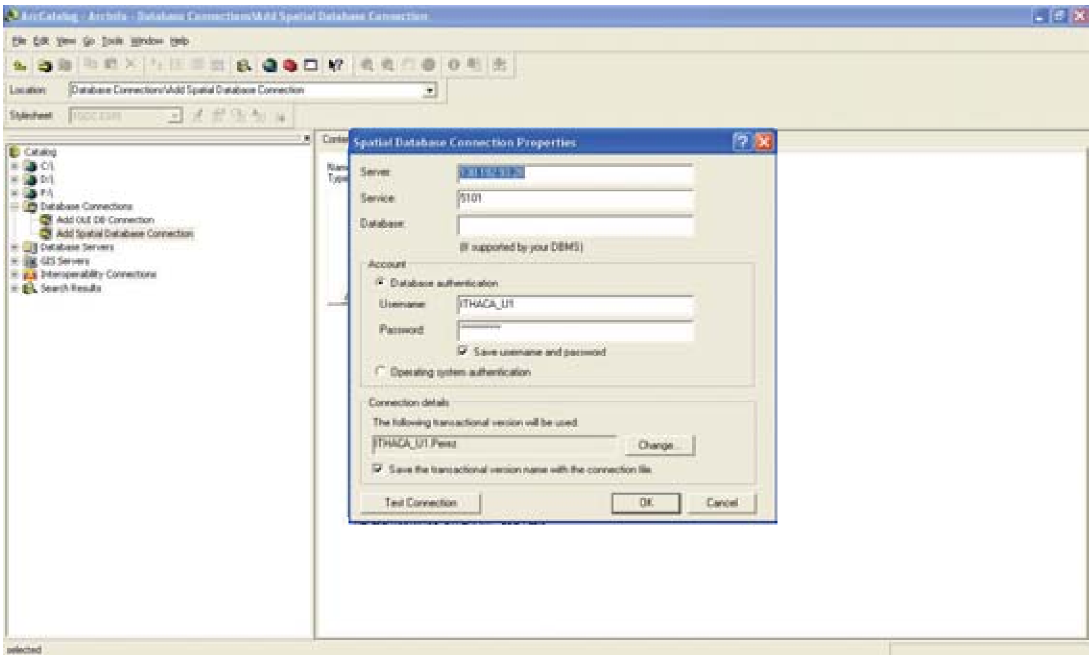


Figure 6.4 – Creation of a spatial database connection in ArcCatalog.

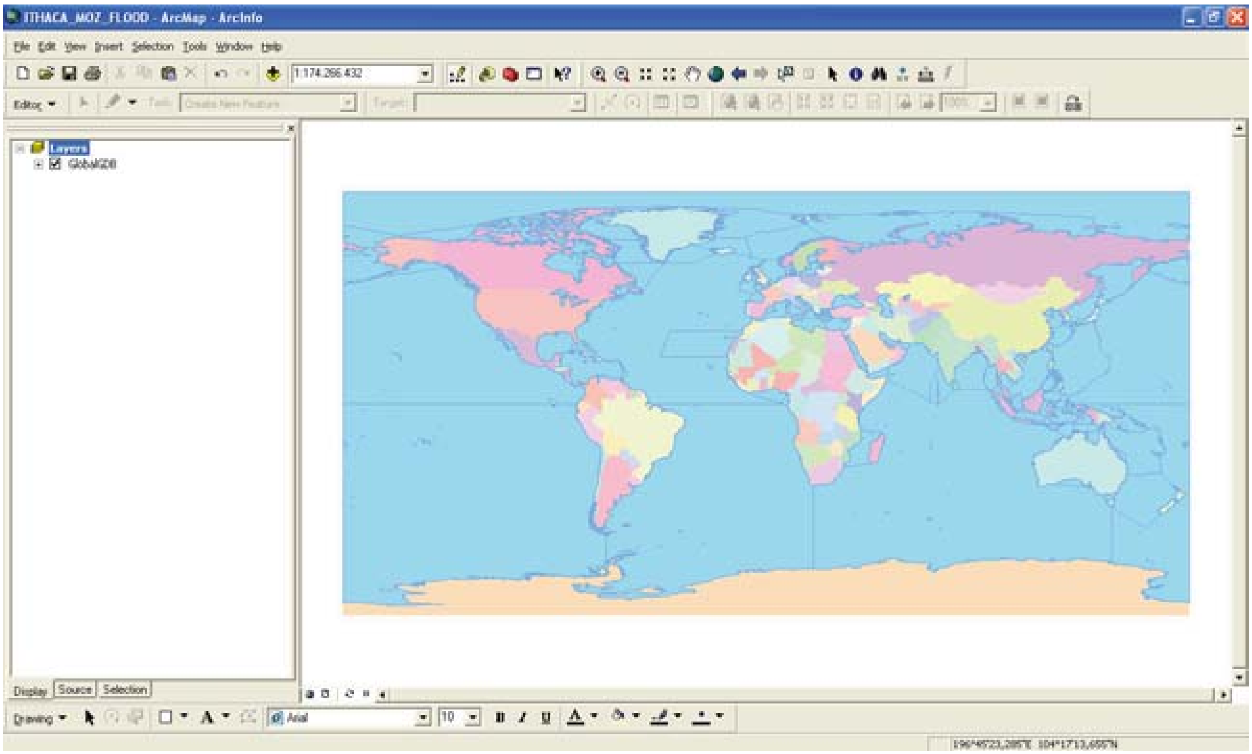


Figure 6.5 – Access to the global geodatabase content.

Adding a connection to an ArcSDE geodatabase service in ArcCatalog, users can manage the developed ArcSDE geodatabase in a remote way. The ArcSDE service spawns a process on the server to broker the connection between ArcInfo and the database. When user add a new connection to an ArcSDE geodatabase service, connection file on disk is created. This file contains the information needed to establish the connection. Database administrator can set up connection files for interested geodatabase users and distribute these so that end users will not require any information about the geodatabase server to which they are connecting.

In this way, all geodatabase users, either in the ITHACA and in the WFP's units, have a fast, direct and easy access to the global geographic data contained in the developed geodatabase. A major advantage of the ArcSDE geodatabase solution implemented, is that each user disposes of a unique and central repository of all data needed in his activities. Moreover, to have a central repository of data produced and/or used by different users allows to retain always the most up-to-date version of geographic data themselves.

After loading spatial data in a suitable map document, users of the *GIS/Map Production sub-unit* can then navigate the geodatabase and select, for a particular area of interest, the specific geographic data which they must include in the final map product.

Figure 6.6 shows the choice of a particular area, that is, the Mozambique state which was hit by a severe flood event in January and February of this year. As a matter of fact, immediately after the event, the WFP's humanitarian workers, engaged in the rescue operations, asked ITHACA for producing cartographic products showing the impacts of the flood, that is, the areas covered by the flood water (usually called "flooded areas") and the affected population.

As already discussed, flooded areas data are produced by the *Remote Sensing user g*roup (see Chapter 4), which performs satellite data processing and analysis in order to extract the desired information. To correctly accomplish its task, this user group needs some preliminary produced data (the floodable areas), reference satellite data and reference geographic vector data. All these data are easily found in the central geodatabase, where subsequently they are placed the final results of analyses carried out by this group, in order to share them with the *GIS/Map Production user group*.

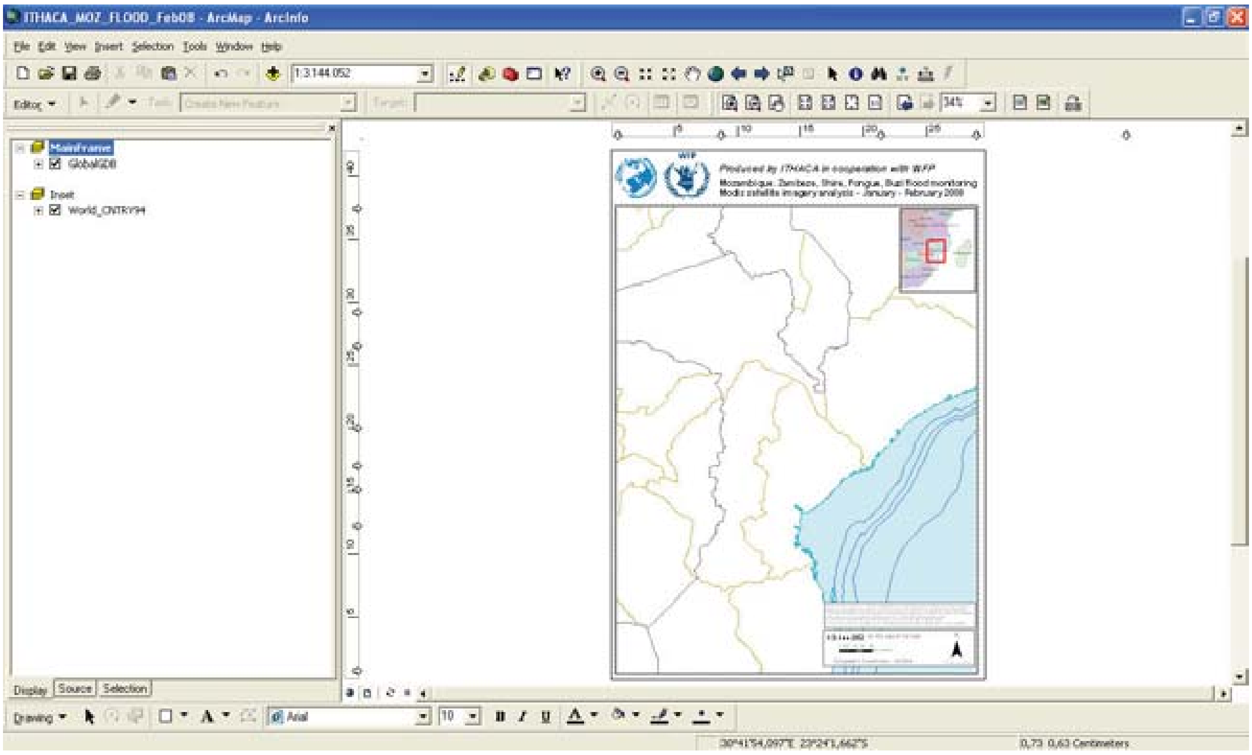


Figure 6.6 – Access to the Mozambique geodatabase content.

Following Figures 6.7 to 6.11 show common map creation operations, using the ArcSDE geodatabase connection. During these operations, all base geographic data (such as road networks, hydrography classes and antrophic features, and so on), which are necessary in order to obtain the final desired map type, are superimposed and represented according to the defined map layers. In this frame, the advantage of the developed geodatabase is that, during map production activities, all needed geographic data are found in an unique repository. Therefore, users do not have to find data stored in different files that sometimes belong in different positions, saving one's time.

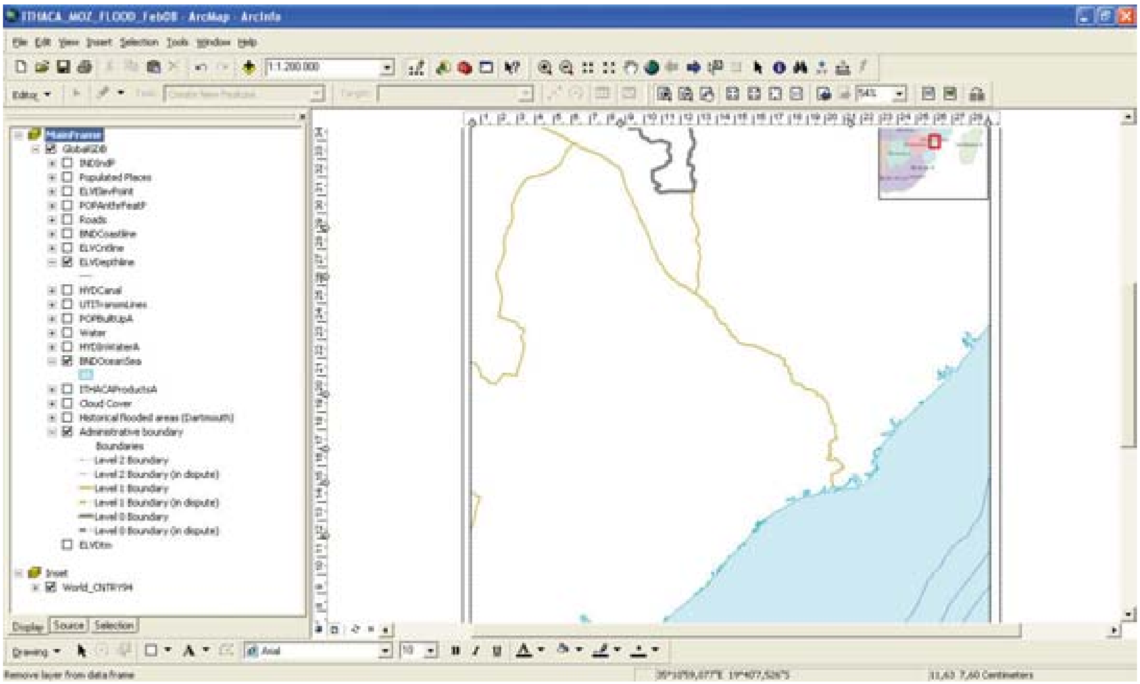


Figure 6.7 – Boundaries features overlay.

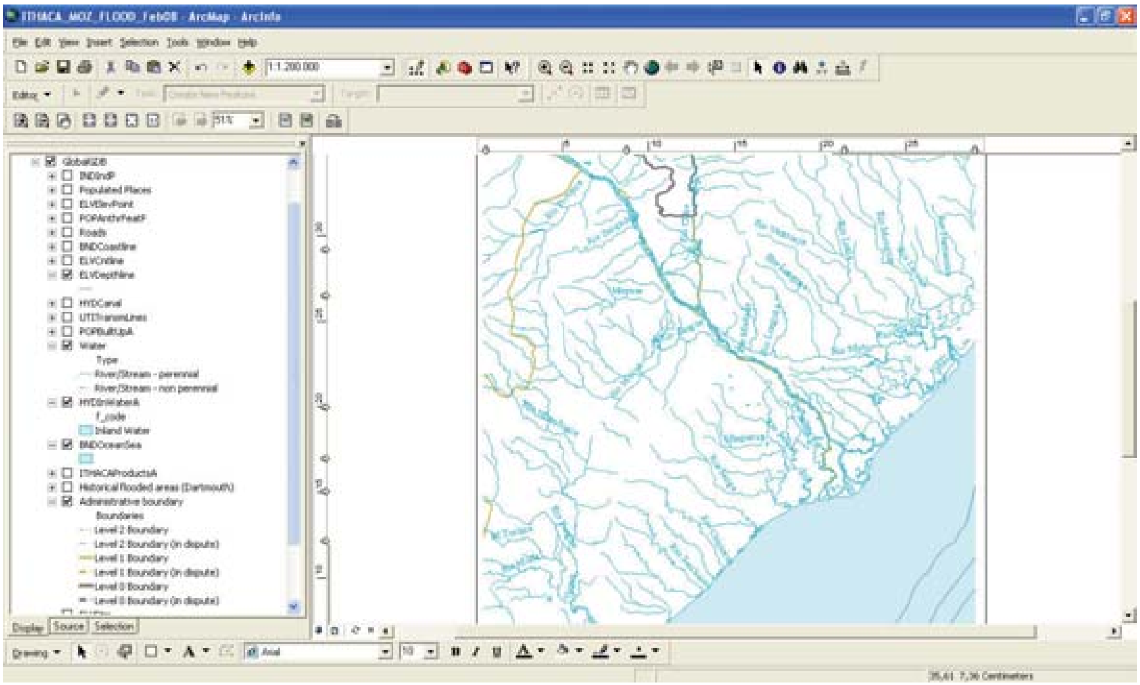


Figure 6.8 – Hydrography features overlay.

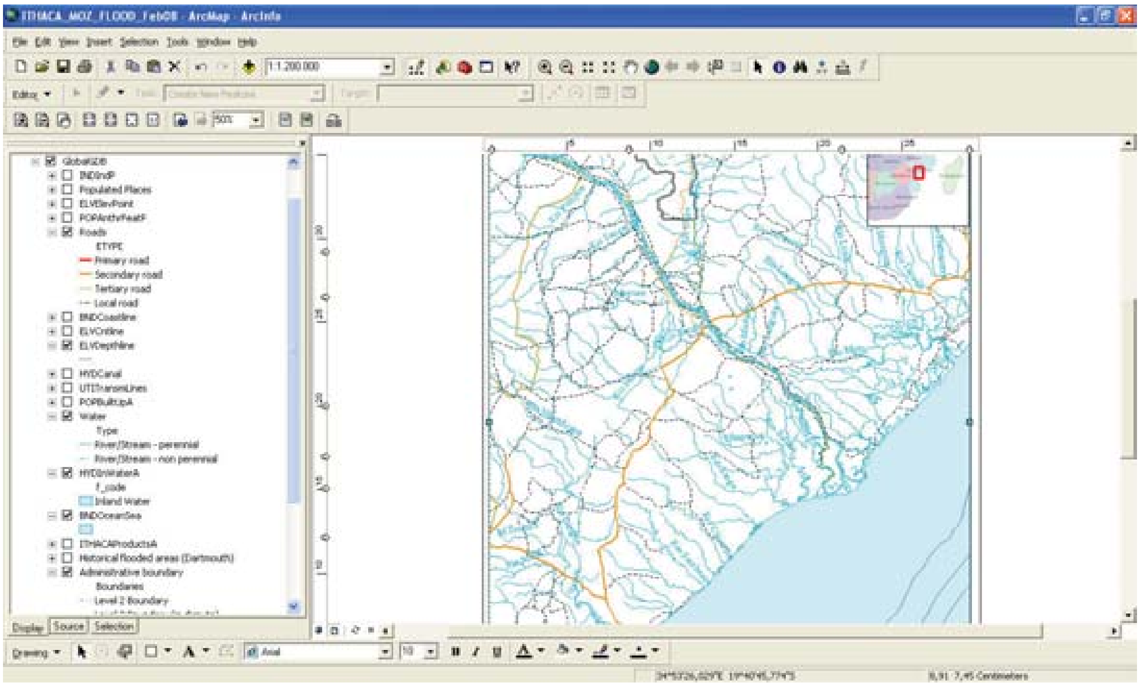


Figure 6.9 – Transportation features overlay.

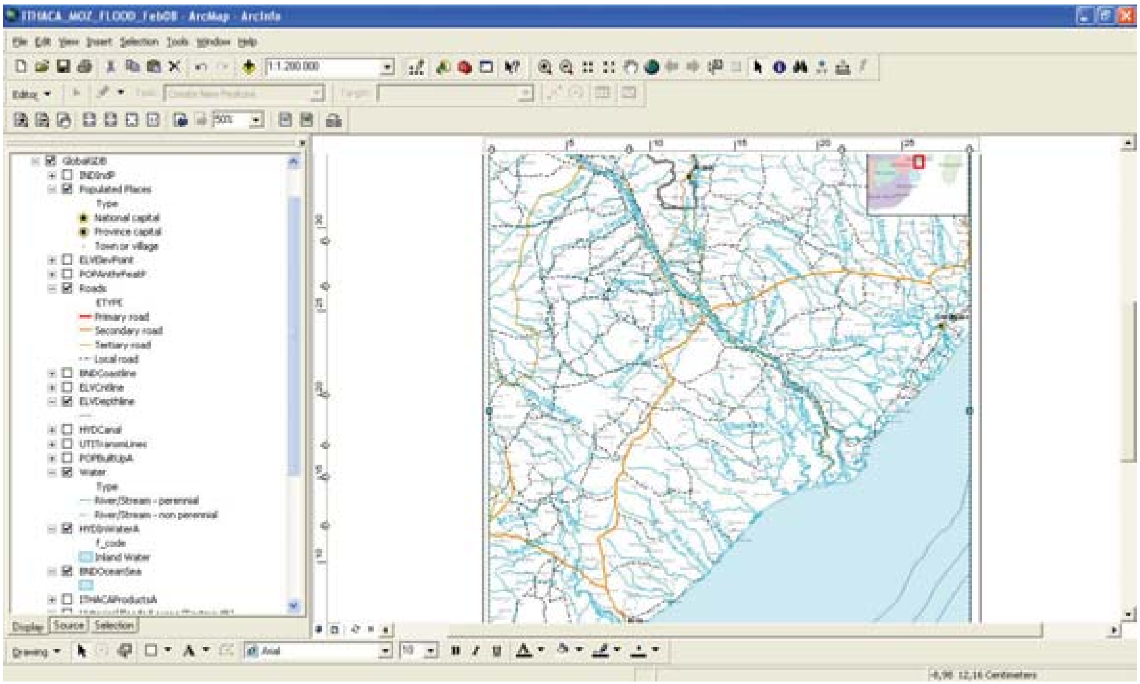


Figure 6.10 – Populated places features overlay.

Afterwards, fundamental thematic data are added to the map. In this considered case, fundamental data are constituted by flooded areas polygons supply by the *Remote Sensing user group*. As aforementioned, the flood polygon features produced during the various flood events analysed by the Early Impact unit are stored in the geodatabase. A different product code, indicating each particular catastrophic event, is associated to these data in the geodatabase. Therefore, in the map production phases, simply query operations on data contained in the database allow to easily retrieve flooded areas corresponding to the chosen flood events. This is fundamental in order to develop maps showing flood evolution monitoring results. In this way, the developed geodatabase guarantees the correct and fast data sharing and reuse among all the different users. Figure 6.11 shows the superimposition of flooded areas defined in two different dates during the considered Mozambique flood event.

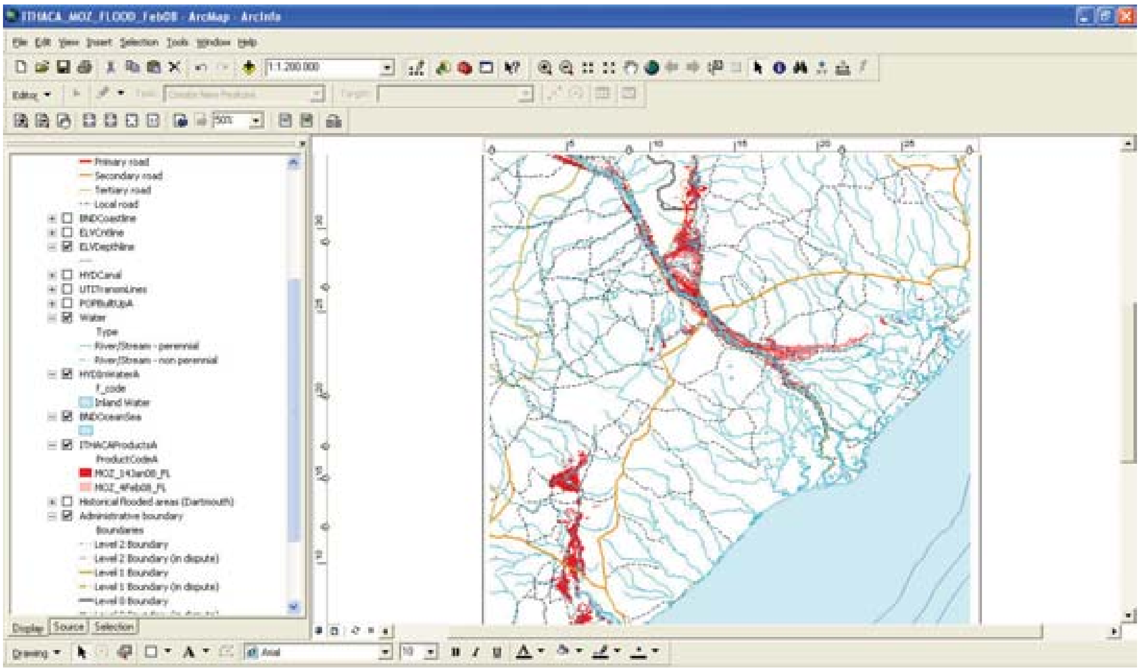


Figure 6.11 – Retrieving of historical flooded areas features in order to develop a Flood Monitoring map.

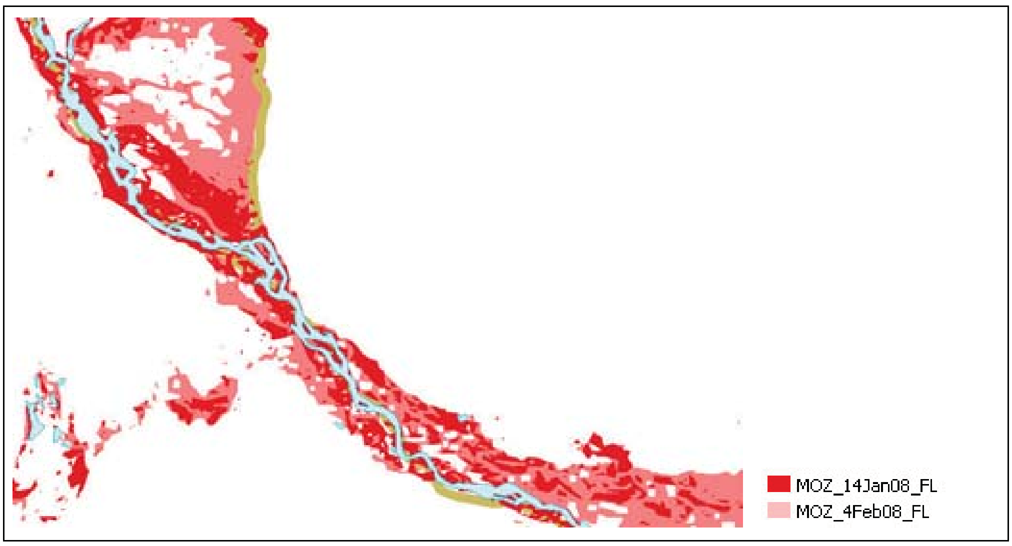


Figure 6.12 – Retrieving of historical flooded areas features in order to develop a Flood Monitoring map (particular).

Other data produced during the *Remote Sensing unit* analyses, such as the areas which can be considered wet or the areas obstructed by cloud cover, are also saved within the geodatabase. They can be then used in order to complete the map definition (Figure 6.13).

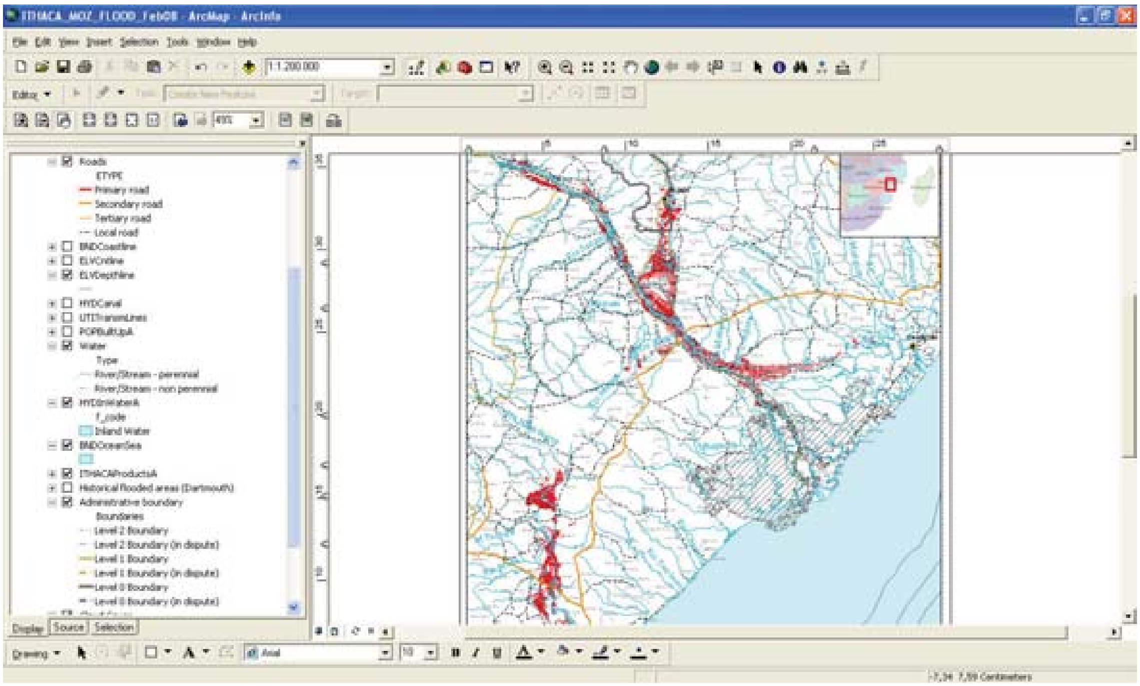


Figure 6.13 – Retrieving of cloud covered areas in order to develop a Flood Monitoring map.

Another important operation carried out by the *GIS/Map Production unit* is the estimate of the population hit by the flood. This operation, as already discussed in Chapter 4, is based on a GIS specific procedure developed by the unit using flooded polygons, population raster data and boundary data contained in the geodatabase. Obtained affected population figures are then included in the map (Figure 6.14).

Country	Province	District	N. of people living in estimated flooded areas (+/-10%)
Malawi	Southern		155,800
Mozambique	Manica	Guro	50
		Sussundenga	5,000
		Tambara	1,800
	Sofala	Buzi	39,200
		Caia	6,800
		Chemba	5,000
		Cheringoma	250
		Chibabava	700
		Dondo	23,400
		Gorongoza	7,900
		Marromeu	4,100
		Muanza	300
		Nhamatanda	28,600
	Tete	Cahora Bassa	150
		Changara	200
		Moatize	300
		Mutarara	14,400
	Zambezia	Chinde	1,900
		Inhassunge	1,000
		Mopeia	8,000
		Morrumbala	4,700
		Namacurra	0
		Nicoadala	2,000
		TOTAL	**311,550**

Figure 6.14 – Population figures added to the Flood Monitoring map.

Finally, some map layout arranging operations are carried out (Figure 6.15) in order to obtain the final map product (Figure 6.16).

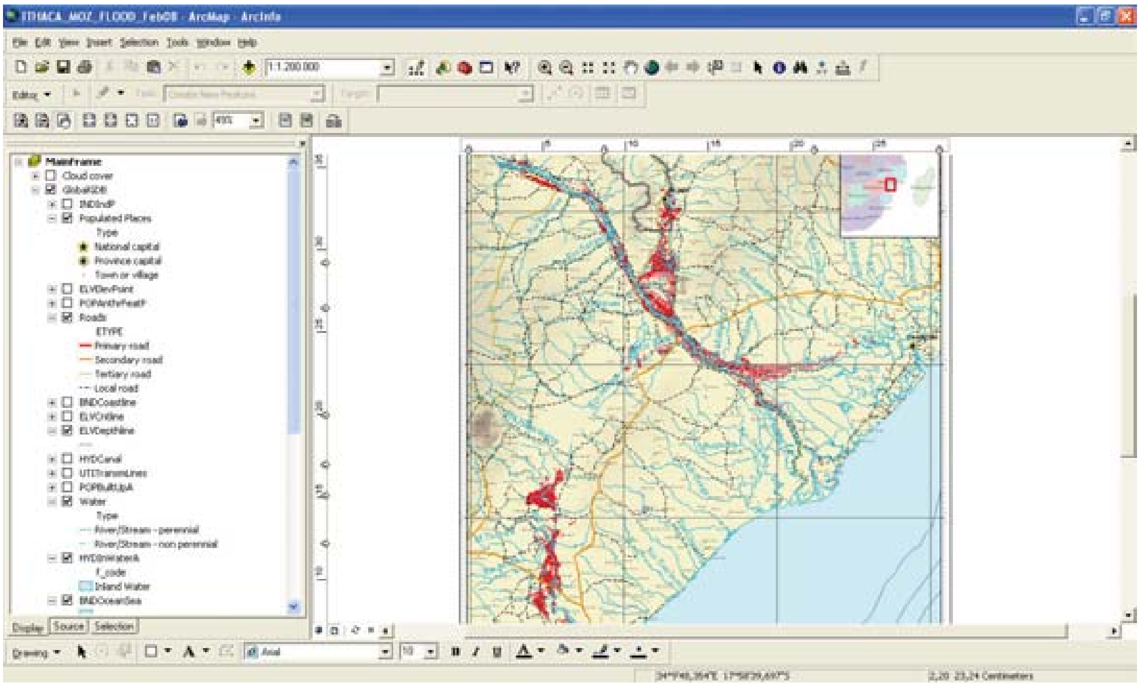


Figure 6.15 – Addition of the DTM data as background and other graphic elements, such as the geographic graticule.

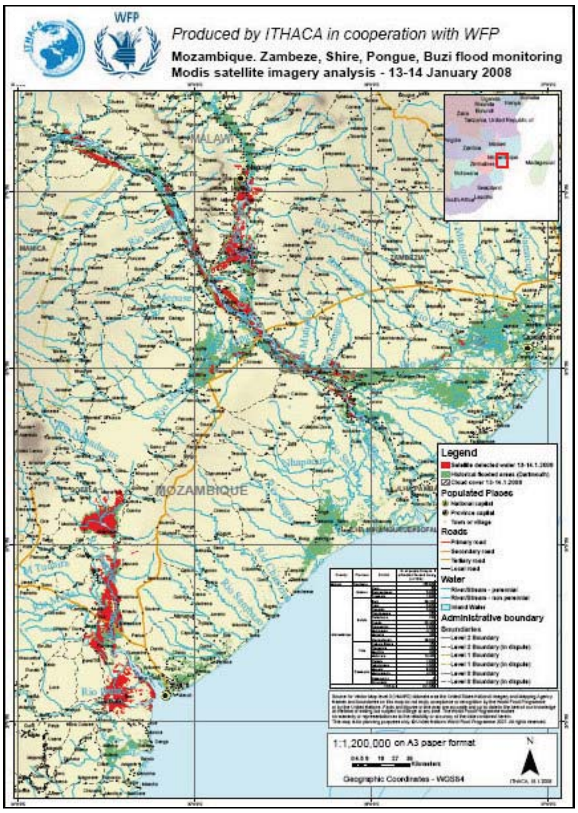


Figure 6.16 – The final produced Flood Monitoring map.

Final cartographic products obtained are then exported in a raster format and subsequently stored in the geodatabase, where they are connected to the corresponding vector flooded areas features using a many-to-many relationship.

Produced raster maps are supply to the WFP's users in two ways. Authorized users can connect to the geodatabase in order to look for the desired products. For this purpose, geodatabase queries about a specific flood event can be performed using the international GLIDE number attribute. This is an international code assigned to each catastrophic event, used in a global context. Other users can download maps through the ITHACA web-site, where a suitable *Map Repository* service (which will be in future connected to the developed geodatabase) has been implemented by the ITHACA's *WEB Application unit* (Figure 6.17).

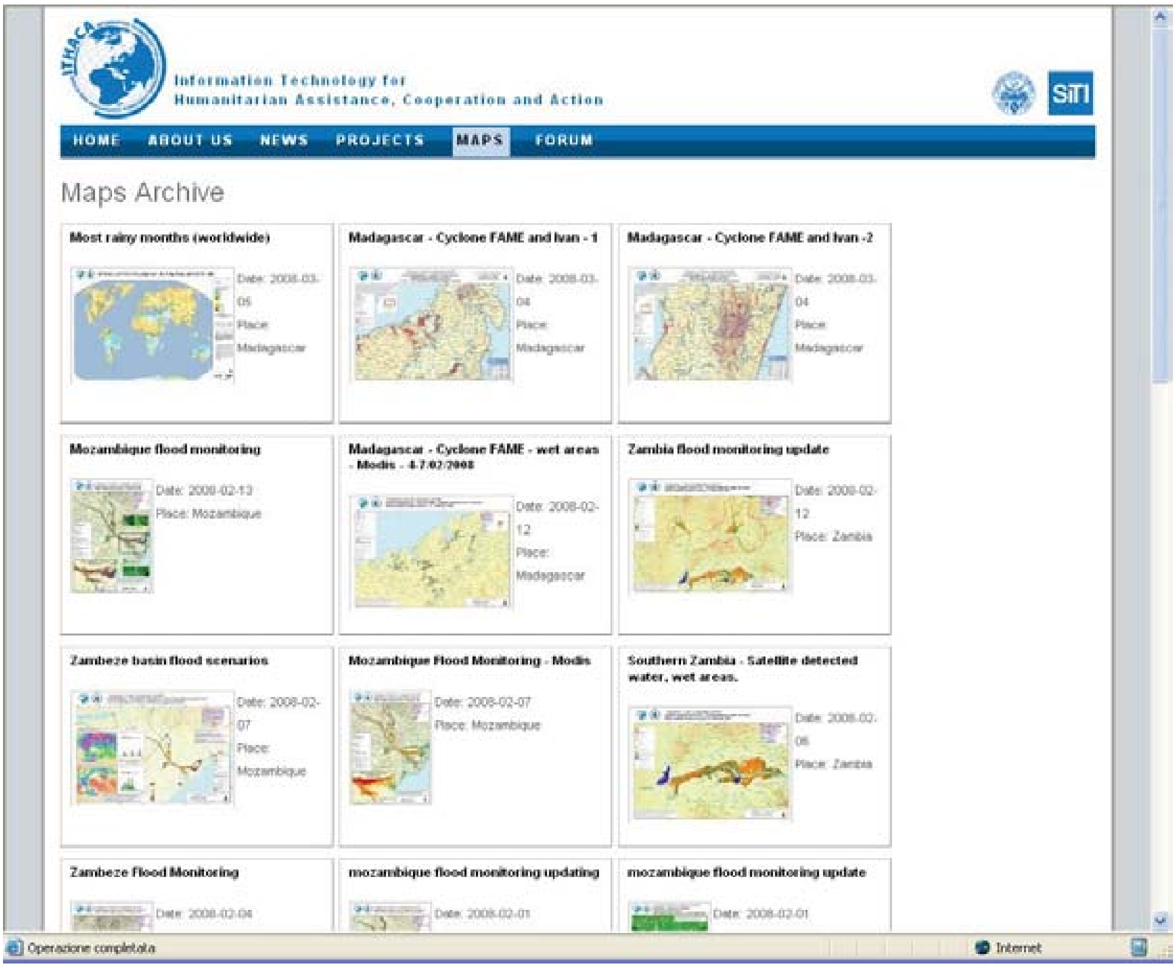


Figure 6.17 – The ITHACA web-site: the Map Repository.

Finally, flooded areas vector data are also distribute to the WFP's users using KML files to visualize in Google Map (Figure 6.18).

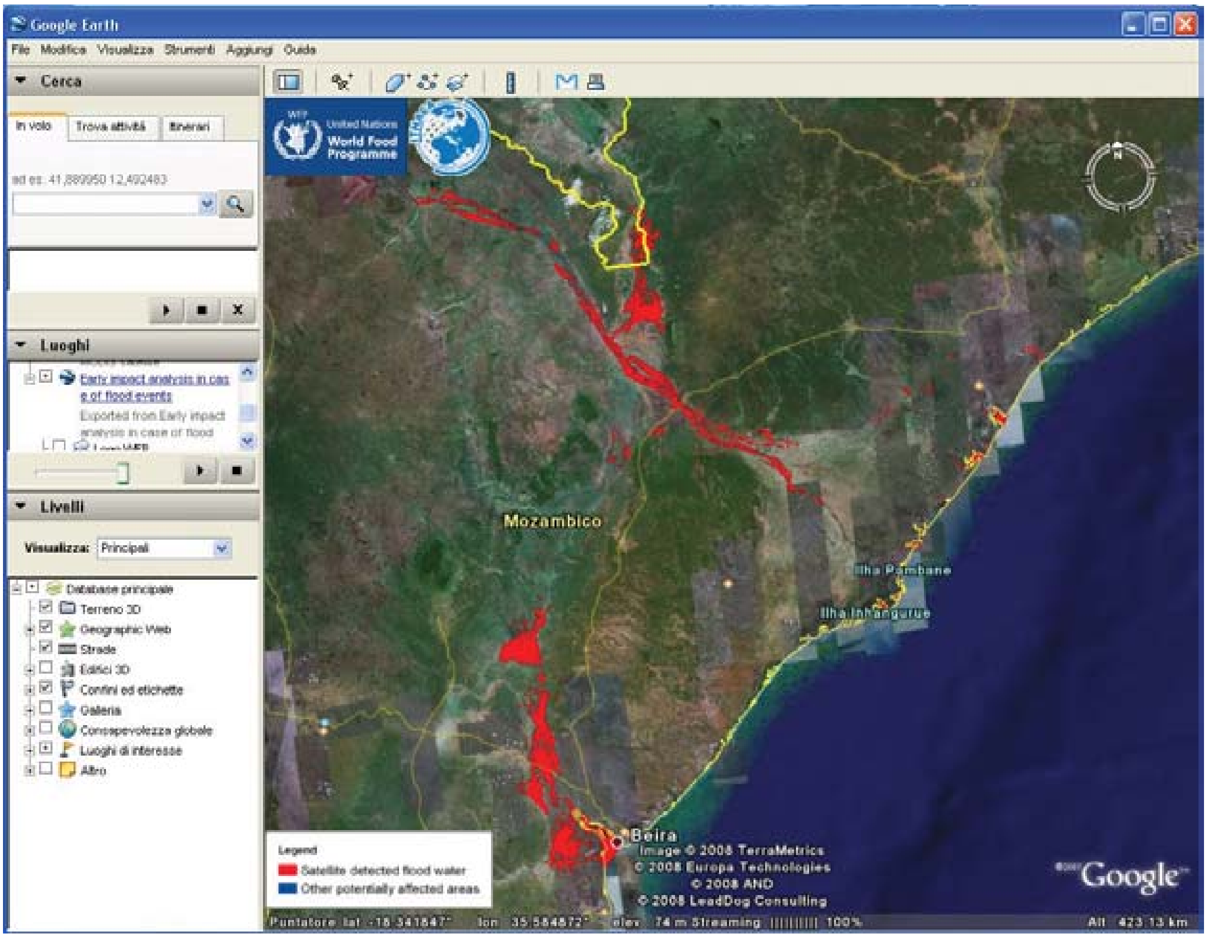


Figure 6.18 – Flooded areas data produced by ITHACA in Google Earth.

CONCLUSIONS AND OUTLOOK

The evident output of this study has been the development of a database containing all the global geographic data which are necessary to correctly carry out both the WFP and ITHACA units activities. The final aim of the planned geodatabase is to support activities connected to the management of emergencies due to natural disasters.

The faced major research topics have been:

- the development of a conceptual schema of the geodatabase. The conceptual design phase has been carried out according to specific needs and demands of the involved user groups within the different interested organizations, using a suitable conceptual data model;
- the implementation of the defined geographic database model as an ESRI ArcSDE Enterprise Geodatabase based on Oracle 10g as DBMS.

Nevertheless, the main results of the conducted study include:

- In order to correctly carry out geodatabase design operations, a needs assessment phase has been performed. The main aim of the activities carried out during this phase has been to identify all the *applications* which would have been supported by the geographic database, all the providers and consumers of geographic information and all the key data flows among users and processes. This was the starting point for defining the spatial *data required* to fulfil the identified applications and, for each data type, identifying the most suitable source of data.
  This phase has been devoted to ensure a common understanding between the design team and those users that have some interests in the implementation of the geodatabase. The main result of this phase has been to obtain a systematic look at how each involved organizations function. In addition, the needs assessment activity itself can be considered as a useful learning tool. As a matter of fact, during this phase, potential users in each participating organization learn about the geodatabase and the different applications and operations, involved with geographic data, which can be considered necessary in ongoing and planned activities. Moreover, in particular for ITHACA, which has been recently established, the development of the needs assessment activity constituted an interesting opportunity to check and structure in detail suitable and effective work-flows.
- The developed geodatabase constitutes a *uniform repository* which allows collecting and organizing all currently used geographic datasets. This geodatabase supports many different types of geospatial information, and it can be entirely placed within its structure. Therefore, all geographic data used by the different users are stored in one *centralized location*. This makes it easier to manage and access. As a matter of fact, users do not have to keep track of various data in different formats stored at different locations.
  Afterwards, the implemented solution allows to define suitable methods in order *to share common datasets* among ITHACA research groups and between ITHACA and external agencies involved in the humanitarian sector (e.g. WFP units or UNOSAT). Users can exploit geodatabase replication functionalities to easily share the contents of the geodatabase with other users at other locations. Solutions devoted to the distribution of the data contained in the geodatabase based on Web GIS applications and services are, at the moment, under consideration. In this way, redundant datasets are eliminated, optimizing performances, avoiding the inconsistencies due to the data duplications, and finally reducing costs.
  Another important point which can be noted is that, in order to define a central geodatabase which supports different user groups, it has been necessary that, in the design phase, all different interested users, in ITHACA and in WFP, agreed on a unique data model. In this way, the geodatabase design activity has certainly contributed also to the *normalization* of features and content of all considered geographic data and to the *definition of a unique and coherent vision* of them, granting better future collaboration results.
- The developed geodatabase guarantees the *management and the distribution of geospatial data within a very controlled environment*, granting to maintain *data reliability and integrity*. The particular adopted choice to implement an ESRI multi-user geodatabase allows to define tools to perform data entry and editing operations in a very efficient way. As a matter of facts, by storing data in an ESRI geodatabase, it is possible to apply rules and constraints on them. To specify rules and constraints

that geodatabase users must follow helps to enforce database integrity and hopefully reduces chances of errors occurring in stored data. For instance, a control system often implemented in the developed data model is the definition of domains, which are definitions of valid values for fields and tables.

Furthermore, the ESRI geodatabase solution allows to define and model spatial relationships among features in a feature class or between feature classes. When feature classes are grouped together within a feature dataset, advanced geospatial relationships such as topologies, can be easily modelled. Topology rules are another useful system which allows to maintain data integrity during data editing and entry operations. At present, a first prototype version of the planned geodatabase is accessible to the WFP offices in order to perform some tests about its effectiveness and to identify next improvements to be developed. After this test phase we expect to work with the WFP users in order to come to the definition of the necessary topology rules and other more customized constraints.

- As shown in the application example presented in Chapter 6, the content of the developed geodatabase is adequate for carrying out ITHACA Early Impact activities, warranting the availability of suitable geographical base data with global coverage, useful for referencing and map production purposes. The appropriate organization of all data internally produced by ITHACA during its activities, guarantees the definition of valuable and easily searchable historical archives exploitable for future analyses. Afterwards, the possibility to include raster data in the geodatabase into raster catalogs structures allows to create massive image repositories, accessible by many users simultaneously. Finally, as discussed in the proposed example, the definition of map layers and cartographic representations constitute an effective enhancement of the geodatabase which helps to speed up maps production activities performed during emergencies.
- Finally, there is another feature of the proposed solution which is very important in order to define a multi-user and distributed data managing system. The developed ESRI ArcSDE geodatabase supports a multi-user editing environment, without data duplication or data extraction. In other words, this system allows two or more users accessing the same data at the same time, which simultaneously make edits. This is possible with *versioning*, the ability to track and record all changes to the contents of a geodatabase. With versioning multiple editors can make changes to the data in the geodatabase and then, later in the workflow, the different edits can be reconciled and aggregated back together. Different users access and read-write permissions to the contents of a geodatabase can be also defined.

  This is a very important feature exploitable in our application. For instance, it may be necessary to check out a portion of a database for a particular area or district to a personal computer or mobile device and update that information in a disconnected session. Then, the updates must be posted to the main database. Moreover, implementing a *Spatial Data Infrastructures,* in which information is shared across a range of organizations, is collaborate effort that requires the sharing of updates across the Internet in a well-defined XML schema and the sharing of change-only updates between databases. Finally, the versioning approach may be exploited to distribute geographic database replicas which can be partial copies of a main central database for a particular geographic region or purpose. Periodically, the databases must be synchronized by exchanging updates.

> At present, some discussions are ongoing with the WFP offices in order to define how exploiting this geodatabase capability, for example in order to supply geodatabase replicas to the WFP Country Offices and Regional Bureaux and to the UNJLC field users, allowing effective data updating operations.

At the moment, future developments of the conducted study are mainly connected to two different activities, the geodatabase maintenance and the WFP SDI development ones. First of all, some geodatabase data model enhancement operations are expected in order to include additional spatial data required by new starting ITHACA activities. Other data model changes probably will be necessary after the completion of the geodatabase prototype testing operations.

As aforementioned, the geodatabase data model design and implementation activities performed constitute only the first step of the more complex project of development of a *Spatial Data Infrastructure* for the WFP, managed by ITHACA. As a matter of fact, the concept of SDI concerns not only spatial data structure, but also data management (the technology to allow multiple users to query, retrieve, manipulate, edit and save data in a common repository) and supply of services (what the users can get from the data repository). Therefore, the final complete system will contain catalogues, search engines, data processing tools, map generating tools, user identification and security measures.

At the moment, the *geodatabase/SDI ITHACA unit*, responsible for this project, is discussing with the WFP users in order to define the final system architecture which will make the geodatabase accessible to the interested users, according to suitable access permissions. Another important topic under discussion is the possibility to provide several services based on data contained in the implemented geodatabase. The services which will be provided are devoted to the development of Web applications and portals designated mainly for data consultation and analysis, for on-line archive catalogues distribution, and for fast automatic map products generation.

BIBLIOGRAPHY

S. Abiteboul, R. Hull, “IFO: a formal semantic database model”, ACM transaction on Database Systems, 12(4):525-565, 1987

A. Albano, G. Ghelli, R. Orsini, “Basi di dati relazionali e a oggetti”, Zanichelli, Bologna, 2007

A. Albano, G. Ghelli, R. Orsini, “Fondamenti di basi di dati”, Zanichelli, Bologna, 2005

A. Ajmar, F. Perez, G. Sartori, “Sviluppo e implementazione di una Spatial Data Infrastructure per il World Food Programme”, XI Conferenza Nazionale ASITA, Torino, Italia, 2007

D. Arctur, M. Zeiler, “Designing Geodatabases, Case studies in GIS data modeling“, ESRI Press, California, 2004

P. Atzeni, S. Ceri, S. Paraboschi, R. Torlone, “ Basi di dati” 2° ed, McGraw-Hill, 1999

P. Atzeni, S. Ceri, S. Paraboschi, R. Torlone, "Basi di dati: architetture e linee di evoluzione" 2° ed, McGraw-Hill, 2006

C. Batini, S. Ceri, S.B. Navathe, "Conceptual database design", Benjamin Cummings, USA, 1992

P. Boccardo., E. Borgogno Mondino, F. Di sabato, F. Giulio Tonolo, F. Perez, G. Sartori, "Utilizzo di dati MODIS per la gestione di emergenze", X Conferenza Nazionale ASITA, Bolzano, Italia, Novembre 2006

P. Boccardo, F. Disabato, F. Giulio Tonolo, F. Perez, "ITHACA project: thematic maps production for disaster management", 27th EARSeL Symposium "Geoinformation in Europe", Bolzano, Italia, 2007

G. Booch, J. Rumbaugh, I. Jacobson, "Unified Modeling Language Reference Manual", Addison-Wesley, 2004

H. W. Calkins, "Entity-relationship modeling of spatial data for geographic information systems", (from http://citeseer.nj.nec.com website)

G. Comoglio, "Cartografia numerica", dispense del corso, Politecnico di Torino, 2003

CRED EM-DAT, Centre for Research on the Epidemiology of Disasters Emergency Events Database. Website: http://www.emdat.be/Database/terms.html (last visited March 2007)

J.F. Dooley Jr., "An inventory of globally consistent geospatial databases and libraries", Environment Resources Series No.19, FAO, Rome, 2005

M. J. Egenhofer, R.D. Franzosa, "Point-set topological spatial relation", International Journal of Geographical Information Systems, 5(2), 1991

R. A. Elmasri, S. B. Navathe, "Sistemi di basi di dati" 5° ed, Pearson Addison Wesley, 2007

ESRI, ArcGIS Desktop Help, 2006

A. Friis-Christensen, N. Tryfona, "Requirements and research issues in geographic data modeling", (from http://citeseer.nj.nec.com website)

Global Map project. Website: http://www.iscgm.org/cgi-bin/fswiki/wiki.cgi (last visited September 2007)

Global Spatial Data Infrastructure (GSDI), "Developing Spatial Data Infrastructures: the SDI Cookbook" v. 2.0, editor: D.D. Nebert (GSDI), 2004 (from GSDI website: http://www.gsdi.org)

M. F. Goodchild, "Geographical data modeling", Computers and Geosciences, 18(4):401–408, 1992

D. Guha-Sapir, D. Hargitt, P. Hoyois, “Thirty years of natural disasters 1974-2003: the numbers”, CRED, Presse Universitaire de Louvan, 2004

T. Hadzilacos, N. Tryfona, “Evaluation of database modeling methods for geographic information systems”, (from http://citeseer.nj.nec.com website)

B. Henricksen, “UNSDI COMPENDIUM. A UNSDI vision, implementation strategy and reference architecture“, United Nations Geographic Information Working Group (UNGIWG), 2007

P. Hoyos, R. Below, J-K. Scheuren, D. Guha-Sapir, “Annual Disaster Statistical Review: Numbers and Trends 2006”, CRED, Brussels, 2007

Object Management Group (OMG), Unified Modeling Language. Website: http://www.uml.org/ (last visited November 2007)

A. Perencsik, E. Idolyantes, B. Booth, J. Andrade, “Introduction to CASE tools”, ESRI Press, USA, 2004

A. Perencsik, E. Idolyantes, B. Booth, J. Andrade , “Designing geodatabases with Visio”, ESRI Press, USA, 2005

A. Perencsik, S. Woo, B. Booth, S. Crosier, J. Clark, A. MacDonald, “ArcGIS 9, Building a geodatabase”, ESRI Press, USA, 2004

A. Rajabifard, I. P. Williamson, “Spatial Data Infrastructures: concept, SDI hierarchy and future directions”, (from http://citeseer.nj.nec.com website)

P. Rob, C. Coronel, “Database systems: Design, Implementation, and Management” 7° ed, 2006

J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, W. Lorensen, “Object-oriented Modeling and Design”, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1991

T.J. Teorey, S.S. Lightstone, T. Nadeau, “Database modeling and design: logical design” 4° ed, Morgan Kaufmann, 2005

M. Zeiler, “Modeling our world, The ESRI guide to geodatabase design”, ESRI Press, California, 1999

United Nations Geographic Information Working Group (UNGIWG). Website: http://www.ungiwg.org/ (last visited December 2007)

UNGIWG, “Report of seventh Plenary Meeting, Santiago, Chile, 2006”, 2006

UNGIWG, “Strategy for developing and implementing a UNSDI in support of Humanitarian Response, Economic Development, Environmental Protection, Peace and Safety”, 2007

UNGIWG, “UNSDI Global Partners Meeting, ESA-ESRIN, Frascati, Italy, 2007 – Meeting Report”, 2007

UNESCO, “Training Module on the Applications of Geographic Information Systems (GIS)”, internet resource. Website: http://gea.zvne.fer.hr/data/index.html (last visited March 2007)

UN Joint Logistic Center (UNJLC). Website : http://www.unjlc.org (last visited September 2007)

United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Website: http://ochaonline.un.org (last visited September 2007)

Vulnerability Analysis and Mapping (VAM). Website: http://vam.wfp.org/ (last visited September 2007)

VAM Spatial Information Environment (VAM-SIE). Website: http://vam.wfp.org/vamsie/srv/en/main.home, (last visited September 2007)

World Food Programme (WFP). Website: http://www.wfp.org (last visited September 2007)

WFP, Emergency Preparedness and Response Web (EPWeb). Website: http://epweb.wfp.org (last visited November 2007)

WFP, “Disaster Mitigation Guidelines for WFP Assistance”, 2002 (from WFP website)

WFP, “Emergency Preparedness and Response Framework (EPRF)”, 2003 (from WFP website)

WFP, “Disaster Mitigation: a strategic approach”, 2000 (from WFP website)

WFP, “How to work with WFP”, 2005 (from WFP website)

“Links between Natural Disasters, Humanitarian Assistance and Disaster Risk Reduction: A Critical Perspective”, UNDP Human Development Report 2007/2008

APPENDIX

The geodatabase implementation

In this appendix a document produced by the ITHACA's *Global Geodatabase/SDI Unit* is proposed (authors: A. Ajmar, O. Terzo, and F. Stupino).

This document summarizes the most up-to-date choices about the whole system implementation, the software components, and the data distribution features. It proposes, also, further developments which refer mainly to the data management operations refinement and to the definition of future web services and applications based on data contained in the geodatabase.

Information Technology for
Humanitarian Assistance, Cooperation and Action

WFP Spatial Data Infrastructure (SDI) in support to a UN SDI

Data model definition, geodatabase implementation and WebGIS applications development

In cooperation with:

22 February 2007 (draft 1.1)

Summary

Introduction

During the Sixth United Nations Geographic Information Working Group (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. In 2006 the UNGIWG sponsored a number of strategic and technical papers on SDI. In 2007 the UNGIWG Plenary Meeting endorsed the proposal to establish a UNSDI project. Key issues on Geospatial Information management elements of the UNSDI has been identified during the UNSDI Global Partners Meeting held on 1-2 March 2007 in Frascati (Italy) and are highlighted in the next table.

	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

Table 1 - UNSDI main issues and expectations

SDI architecture

As result of needs assessment round tables with WFP and UNJLC users, an architecture granting a solid back end and a flexible, interoperable and customized front end has been considered the best solution for managing data in a distributed environment (Fig. 1). Back end component is accessed by high level users, in charge of database management and of performing complex data analysis procedures. Front end applications are mainly dedicated to analysis, processing of project specific geodata and exploratory aspects; simple editing capabilities should be also included.

Technical constraint related with low performance internet connection required to develop solution for disconnected data management using database replica and guided procedures for data reconciliation.

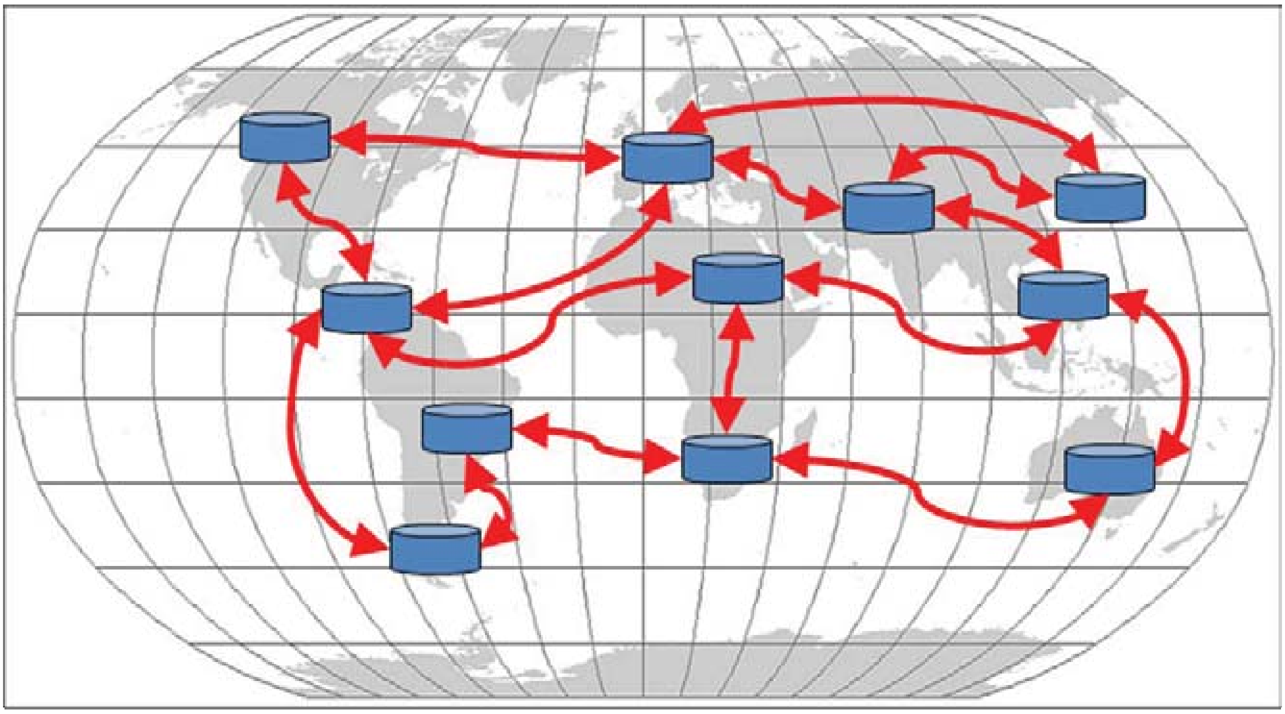


Fig. 1 - Schematic representation of a distributed data managing environment

Re-use and re-organization of currently managed dataset has been a priority in the data modelling phase, together with direct access to open geographic sources (SRTM, archive satellite images, etc.) without any need for data pre-processing.

The development of data management rules and map templates allow to create a "lowest common denominator" for geographic analysis and mapping, in support to decision making during emergencies.

Data model definition and consolidation

Based on the data structure submitted and approved by WFP (ODAP and VAM) and UNJLC, a UML data model has been consolidated and implemented using an ORACLE 10g database as DBMS platform.
The database schema is composed by 2 different entities:

- **Base data** includes basic geographic and alphanumeric information, in order to produce basic geographic outputs and analysis. Base map data sources are highlighted in Table 2;

BASE MAP Sources		
Name	Geographic Area	Ref. Scale
AVHRR NDVI	World	1000m
Basin	World	1:5.000.000
DCW	World	1:1.000.000
Geonames	World	1:25.000 to 1:1.000.000
GLC2000	World	1000m
Landsat Imagery	World	30m
Landsat mosaic	World	15m
LandScan	World	1000m
MODIS Imagery	World	250/500m
MODIS Land Cover	World	1000m
MODIS NDVI	World	250m
SALB	Near-global	1:1.000.000
SRTM	World	90m
STRM SWBD	Near-global	1:150.000
VMAP0	World	1:1.000.000
VMAP1	Local	1:250.000

Table 2 - Base map data sources

- **Transportation** is the translation of conceptual schema developed by UNJLC into a logical relational schema. The implemented version is 1.2.

TRANSPORTATION Sources		
Feature Class	Source	% of records
Roads	Digitized	6.1
Roads	Field Mission with GPS	1.4
Roads	Interview	2.4
Roads	Pre-existing geodataset	69.4
Roads	Report	19.2
Roads	Unspecified	1.4
Ports	Pre-existing geodataset	50
Ports	Report	50
Railways	Digitized	32.1
Railways	Pre-existing geodataset	21.4
Railways	Report	46.4
Stations	Report	100
WarehouseCompounds	Pre-existing geodataset	54.1
WarehouseCompounds	Report	45.9

Table 3 - Actual definition of data sources for Africa Transportation dataset

System architecture

System architecture definition (Fig. 2) has been obtained keeping in consideration three different tasks that the system must perform efficiently:

- ***Geodatabase network**:* the architecture of the geodatabase servers, including:
  - **The master Geodatabase** (Oracle 10g) that contains all the GeoDB schema and the data;
  - **The replica Geodatabase** (Oracle 10g) containing a two-way replica of the master geodatabase, for maintenance purpose and data consistency;

- **Publication GeoDB (**Postgre/PostGIS) a geodatabase replica to be accessed and used by web based services and applications.

➢ ***Internal backup and restore network***: internal support network used to backup sensitive data on a tape driver, to reduce the cost of the storage system and to assure the maximum flexibility of the service.
Policies and scheduling of backup operation should be discussed and jointly agreed, considering several different factors such as data volumes, update rates, data sensitivity and level of services that must be granted.

➢ ***Web Server***: server(s) that provide the publication service of the geodatabase using GIS application. The architecture of the system is composed by two servers with the same hardware configuration.
Three different hypotheses about the web-server publication service can be made:
- One server provides the effective service of publication while the second server supports the computing capacity to the first server;
- A cluster of two servers in active-active mode. They support each other to reduce the load and grant the servicecase of failure of one of the servers.
- A mixed approach: open-source GIS applications and enterprise applications like Arcgis Server.

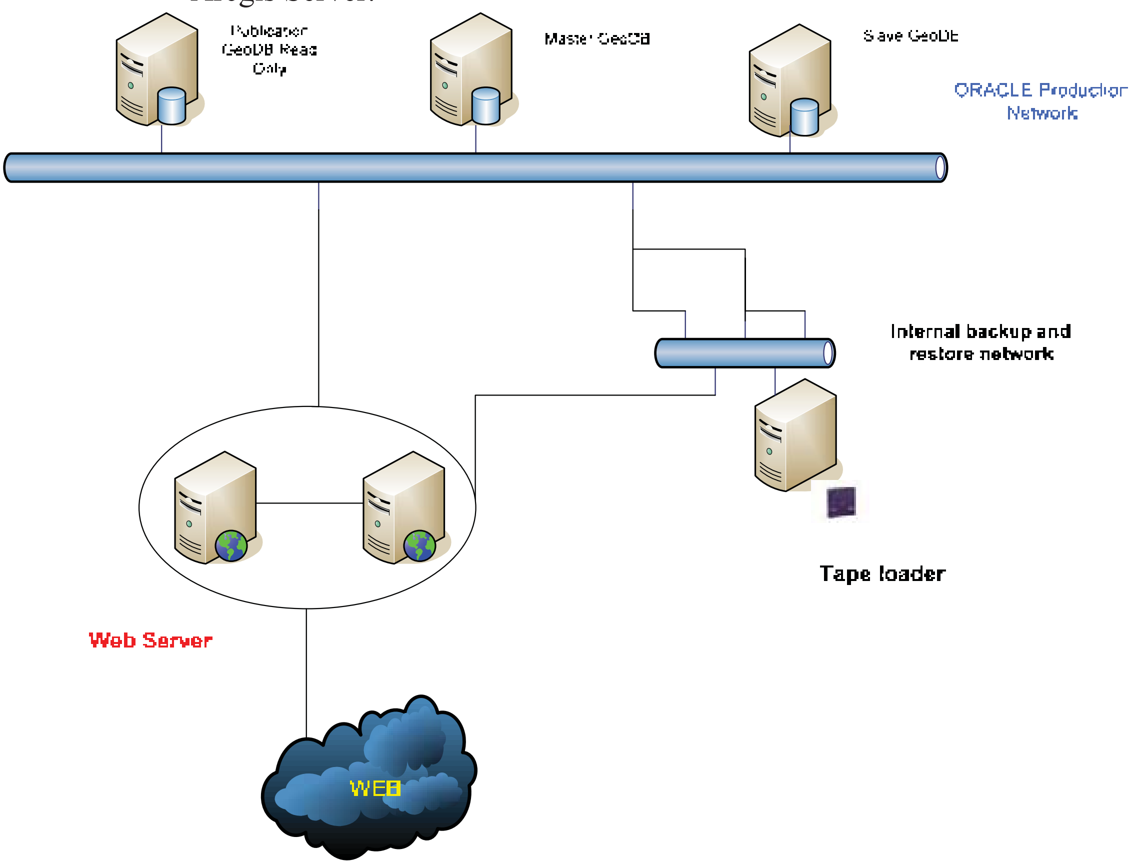


Fig. 2 - Proposed system architecture

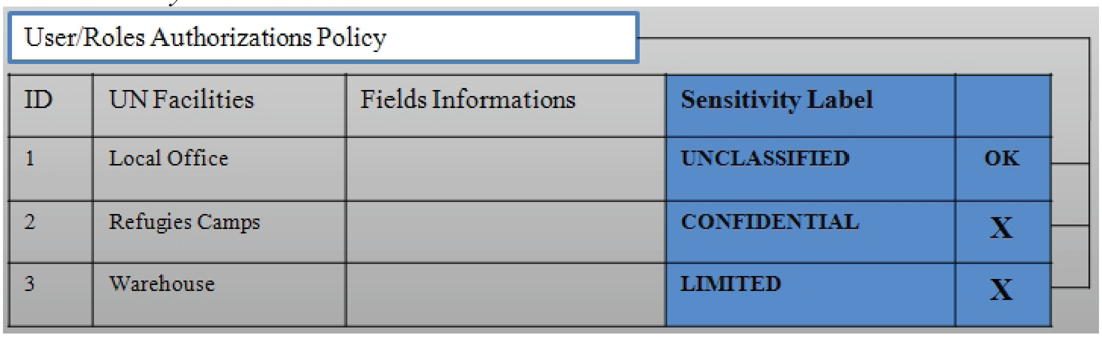
Software components

Table 4 present the proposed software list for the two operational environments:

- ***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, are main factors suggesting the implementation of a commercial products based platform. 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;

- *Label Security Authorizations*.



- ***Publication/Analysis environment***: front end applications, mainly dedicated to analysis, processing of project specific geodata and exploratory aspects; simple editing capabilities should be also included. 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.

Production/editing environment

Operating System	Ubuntu (linux)	7.10 (32bit)	Open Source
DBMS	Oracle 10g	10.2.0.1	Commercial
Gateway Software	ESRI ArcSDE [illegible] [illegible]	9.2	Commercial

Publication/analysis environment

	Description	Version	Type
Operating System	Ubuntu (linux)	7.10	Open Source
DBMS	Postgres-postgis	8.2	Open Source
Web Map Engine (*)	Mapserver	5.0	Open Source
GIS WMS Client	Any OGC Compliant	-	Open Source/Commercial

(*) OGC Web Services and/or specific and customized Web Applications (KA-Map and OpenLayers)

Table 4 - Software components for the two proposed environments

Software and data formats are defined in compliancy with OGC standards, as highlighted in Table 5.

Open Geospatial Consortium®		
Components	**Production environment**	**Publication environment**
DBMS	Oracle 10g	Postgis/PostgreSQL 8.2
Gateway	ESRI ArcSDE	-
Geometry storage	OGCWKB SDO_GEOMETRY	WKB WKT
GIS clients	ESRI ArcGIS Desktop	Desktop or WEB based
WEB Applications	ESRI ArcGIS Server	WMS/WFS
Metadata and Catalogue	GeoNetwork	

Compliant product
Registered product

Table 5 - Software Open Geospatial Consortium standard compliancy.

Data distribution

Following a discussion with WFP and UNJLC 4 levels of users and relative privileges has been identified (Table 6). Such analysis constitutes the starting point for privileges definition over the whole geodatabase.

		Visualisation	Interaction (query/geoprocessing/routing)	Download	Edit	Commit
Public		limited	limited	no	no	no
Humanitarian Partners	**Logs**	yes	yes	maps + reports	attributes only	no
	GIS	yes	yes	yes	yes	no
UNJLC/WFP	**Logs**	yes	yes	maps + reports	attributes only	no
	GIS	yes	yes	yes	yes	Yes

Table 6 - Users privileges schema

The system is made of two Oracle database instances, one acting as master database and another hosting a full two-way replica for data production; that solution, together with adequate back-up policies, should grant data protection and integrity (Fig. 3).

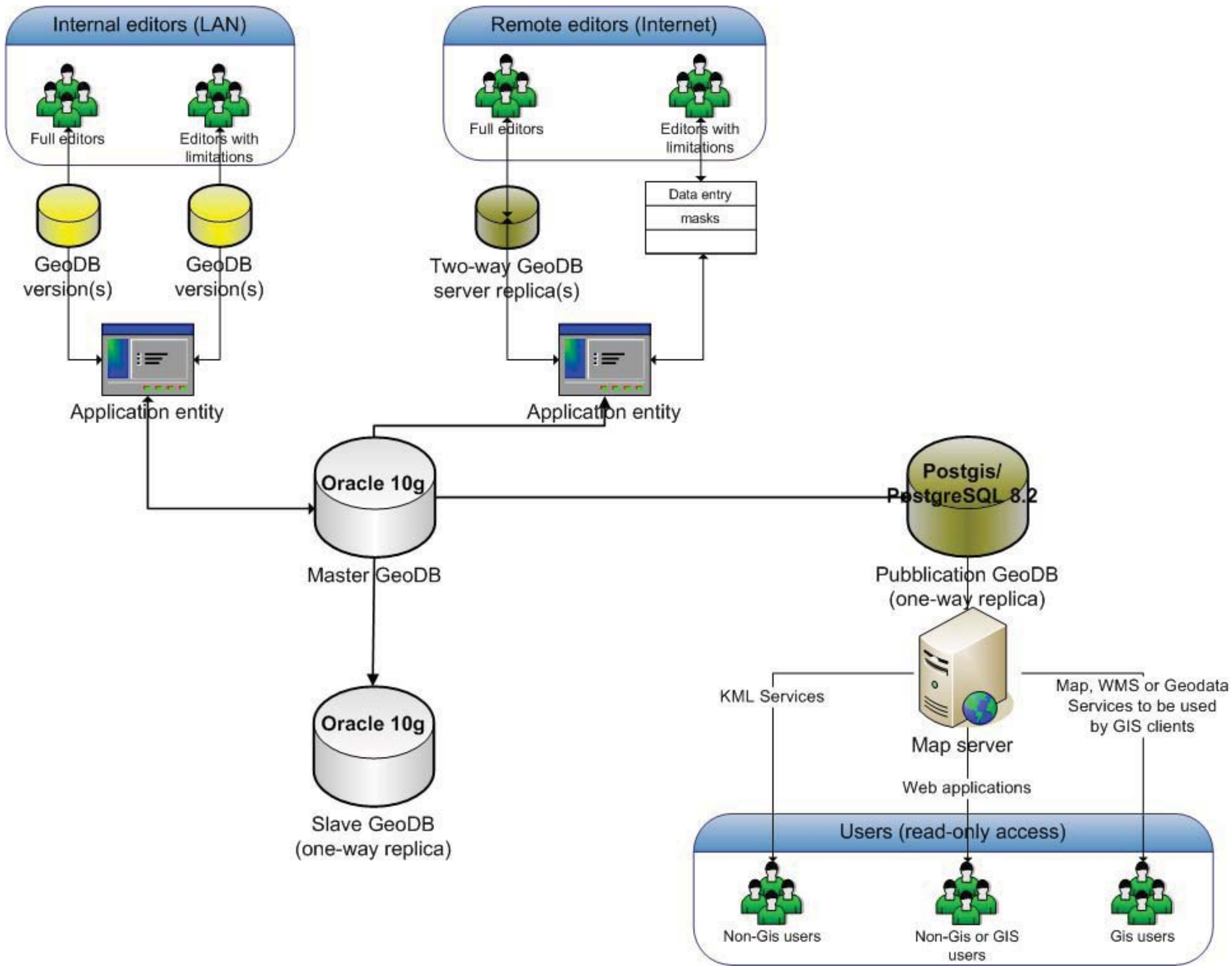


Fig. 3 - Data distribution schema

Future developments

While the WFP SDI phase one (data infrastructure) is under implementation and testing, the ongoing phases two and three (data management and web services) are still under discussion to undertake the following activities:

- **Update of database schema** in order to include more information components:
    - Image catalog for acquisitions performed during emergencies, and related data analysis results (in final stage of definition);
    - Lessons leanred and best practices identified during the testing phase of current data model.
- **Users privileges refinement:**
    - user privileges definition, acting at dataset or feature level;
    - creation and management of data view to filter unnecessary data for selected users (i.e. some of the transportation dataset attributes are not necessary for base data analysis and visualization).
- **Database fine tuning**;
- **Web services and applications** allowing an interoperable level of access and management of the database. Services are intended as user interface elements that accept input from the user selection, process it, and optionally put the result back in the clipboard. Table 7 contains a list of services requested by the users or proposed in the framework of ITHACA collaboration with WFP.
- **Distributed computing technologies**, in order to improve system performance using an available network of computers (ITHACA, UNOSAT, WFP, etc.);
- Full **interaction with Geonetwork** platform.

Type	Subtype	Description	Requirement	Requested by
Editing	Geometry	Perform editing operation on base geometries	• Check procedures before changes are committed in the master db	General
	Attributes	Update database attributes thought the definition of data entry masks	• Guided by database constraints (domains) • Following a predefined flow (UNJLC)	General (with specific requirements from UNJLC)
Data replication	Data management	Create replica of information contained in the geodatabase	• Synchronization • Replica of current displayed set	General
	KML/KMZ	Subset extraction and delivery	• Single layers or complete maps	UNJLC
Satellite image catalogue research functionalities	Disaster management	Research and access to satellite images and derived information produced in case of past emergencies		ODAP
Early warning tools	Flood modelling	Provide flood extent estimates during an emergency	• Usable with limited input datasets • Rapid set-up and result delivery	ODAP
	Climate drought estimate	Drought indicators, automatic alert triggering, food shortages estimates (?)		ODAP
Early impact tools	Affected areas	Delineation of affected areas from satellite identification of hit areas		ODAP
	Affected population	Estimates of population affected by an event	• Include simple socio-economic modelling	ODAP
Reporting	Logistic assessment	Automatic generation of reports containing maps and textual description		UNJLC

Type	Subtype	Description	Requirement	Requested by
Map Source Image Creation	Logistic in-field support	Generation of base layer representation as background map	• For GARMIN devices (UNJLC) • Vector and/or raster	UNJLC
Logistic tools	Snow cover	Automatic detection of areas covered by snow	• Intersection with logistic network components	ODAP

Table 7 - Services and relative description