THE DESIGN OF THE DEAFIN WEB-GEOGRAPHICAL INFORMATION SYSTEM: AN EXPERIENCE IN THE INTEGRATION OF TERRITORIAL RECLAMATION SUPPORT SERVICES

Fabio A. Schreiber[*], Alberto Belussi[§], Valeria De Antonellis[+], Maria Grazia Fugini[*], Giuseppe Pozzi[*], Letizia Tanca[*], Luca Tosi

Chapter VIII
in

WEB-ENABLED SYSTEMS INTEGRATION: PRACTICE AND CHALLENGES

edited by

Ajantha Dahanayake and Waltraud Gerhardt

IDEA GROUP PUBLISHING
2002

pp. 142 - 168

THE DESIGN OF THE DEAFIN WEB-GEOGRAPHICAL INFORMATION SYSTEM: AN EXPERIENCE IN THE INTEGRATION OF TERRITORIAL RECLAMATION SUPPORT SERVICES[1]

Fabio A. Schreiber[*], Alberto Belussi[§], Valeria De Antonellis[+], Maria Grazia Fugini[*], Giuseppe Pozzi[*], Letizia Tanca[*], Luca Tosi

[*] Dipartimento di Elettronica e Informazione, Politecnico di Milano, p.za L. da Vinci 32, 20133 Milano –MI- Italy,
fabio.schreiber/mariagrazia.fugini/giuseppe.pozzi/letizia.tanca@polimi.it

[+] Dipartimento di Elettronica per l'Automazione, Università degli Studi di Brescia, via Branze 38, 25123 Brescia –BS- Italy,
deantone@ing.unibs.it

§ Dipartimento di Informatica, Università degli Studi di Verona, strada Le Grazie 15, 37134 Verona –VR- Italy,
belussi@sci.univr.it

ABSTRACT

The design of a Web-geographical information system strongly requires methodological and operational tools for dealing with information distributed in multiple, autonomous and heterogeneous data sources, and a uniform data publishing methodology and policy over Internet web sites. In this chapter, we describe our experience for the activities of requirement analysis and conceptual design of the DEAFIN Web-geographical information system whose objective is to improve the quality and the comparability of information about available industrial vacant sites, coming from different regional data sources. Heterogeneity and web availability requirements have been taken into account in the system architecture design. The DEAFIN system is thus conceived as a federated web-based information system,

[1] In this document, sections are numbered. Cross-references between sections need to be updated. Keywords for the book index are underlined.

apt to manage and provide access to all the regional relevant information in an integrated and complete fashion. Furthermore, since the data available by a given DEAFIN region partner can be both spatial data and alphanumeric data, for each regional component system in the DEAFIN system, a Web-GIS system is defined.

1. INTRODUCTION

The DEAFIN (Development Agencies and their impact on Foreign direct Investments) project has been conceived in Spring 1997 by Province of Milano (Italy), Västernorrland County (Sweden) and the Freie Hansestadt Bremen (Germany). It aims at contributing to regional restructuring and reviving by improving the quality and comparability of information on available industrial sites and on environmental conditions, in order to improve each region's capability to attract foreign direct investments. Target users have been identified in the potential final investors – mainly small or medium sized companies - and in the regional agencies charged with the task of finding and assigning them vacant industrial sites.

The DEAFIN project comprises three actions and five sub-actions:

1. Regional environment analysis including:
   - inquiry towards the end users to understand their information needs in order to choose an industrial site;
   - a comparative SWOT (Strength Weakness Opportunity Threats) analysis to characterize the three partner regions;
   - a benchmarking activity, the aim of which is to show the development potentialities and the success factors in attracting investors to the successful areas and to suggest the regional authorities harmonious development strategies.
2. Operational tools definition including:
   - an information system with a data bank recording both macro economical and infra structural data on each regional area and detailed data about each vacant industrial site;
   - an eco-auditing action aimed at developing a methodology to classify, following unified criteria, the vacant sites in each region along the dimensions of their environmental criticality and of their state of pollution due to previous industrial activities.
3. Global strategy and new actions. This action tends to foresee a global strategy for a series of innovative actions which, even after the end of the project, should allow for an internationalization process of the small and medium sized companies and a strategic territorial marketing.

The aim of our contribution, within the project, was the study of a proper *information system architecture* and the production of the related *functional specification*, with main focus on "What functions are required within the system and what information is produced, used and consumed by each function". The produced functional specification constituted the technical annex of a tender to be issued by the end of February 2001 for the implementation of the system. Specifically, tenderers were required to provide solutions on "How to implement the information system for supporting required functions and information". The interested reader can find additional documentation about the DEAFIN project progress at the site http://www.deafin.org/.

In this paper, we describe the activities of requirement analysis and conceptual design of the DEAFIN information system and related data bank (action 2) as carried out by the DEAFIN Information System Design Group at the Politecnico di Milano. Section 2 states the problem we faced and describes the background of our intervention; Section 3 focuses on data gathering and access typologies; Section 4 describes the architecture for the proposed system; Section 5 describes the design steps of the federated database, by also providing a sketch of the developed conceptual schema; Section 6 describes geographical information system (GIS) basic concepts and the proposed web interface for the DEAFIN system; Section 7 describes advanced functionalities of the system, with respect to user typologies and to system monitoring and tuning; Section 8 provides final comments and conclusions on project development.

2. STATEMENT OF THE PROBLEM

The DEAFIN project has been launched with the purpose of allowing companies and investors to get a comprehensive information framework about areas located in European Regions suited for potential investments. The target is to make the regional data about possible investments homogenous and comparable, and internationally accessible. In particular, the three Regions (Bremen, Milano, and Västernorrland) participating in the DEAFIN project require to stimulate foreign direct investments related to industrial vacant sites available for redemption for various purposes, such as building or installation of new industrial, handicraft, service or educational activities. Potential investors need both a survey and a detailed view of vacant sites in different locations in order to compare different opportunities in the Regions and decide which investments can be carried on more conveniently. As a consequence, the main purpose of the DEAFIN project is to provide a common environment for comparison of information about available industrial vacant sites and for the analysis of elements of the Regions.

The architecture that better fits the given requirements is that of a *Federated Information System* (FIS) which grants local sites a great deal of autonomy while allowing interoperation among them by means of a common integrated conceptual schema. Furthermore, as to the access to the system, owing to the capillarity of the end user locations and to the need of a simple and widely known interface, a *web-based* structure is mandatory.

To define the functional specification of the DEAFIN Federated Information System (DEAFIN FIS), the following activities have been carried out:

- *analysis of the requirements* of the DEAFIN FIS, that is of a distributed web-based information system relying on a common conceptual database schema according to which Regional databases have to be built (most of regional information is available as paper based documents);
- *conceptual design* of the DEAFIN FIS, centered around the conceptual design of the DEAFIN federated conceptual database schema. The Regional databases have to be built according to the federated schema and then made accessible via web. The availability of data via web allows potential investors to navigate in the DEAFIN site according to various and customizable criteria, based on a benchmarking model developed within the project.

3. INFORMATION REQUIREMENTS

The DEAFIN project started with the *data gathering* phase, aimed at collecting requirements about data and processes managed at the DEAFIN project Public Administration partners' sites. A questionnaire-based tool has been circulated to collect common information to be stored in the DEAFIN FIS.

3.1 Data gathering

The basis of the questionnaire is a set of the *data categories* managed in Public Administration information systems. The DEAFIN-relevant data categories concern Land Use Plans (Master and Regional or specific), Territorial Services, Industrial Vacant Sites, Mobility data, Statistical and Social-Economic data, Base Cartography data, and Information on Cadastral Units data. Information on Vacant Industrial Sites is the main focus of the investigation.

For each category, the questionnaire has collected the data characteristics reported in Table 1.

DATA CHARACTERISTICS
Availability of the data category
Location of the data source
Support technology and name of products/tools
Data format
Data owner
User roles involved in data access
Restrictions applying to data access
Performances
Maintenance policies
Availability on the Web

Table 1. Summary of data characteristics collected in the data gathering phase

In order to identify the main processes to be supported by DEAFIN, the questionnaire had to collect a synthetic description of the procedures adopted at each site to manage the collected data. The basic process typologies are reported in Table 2, where the left-hand column identifies the process type and the right-hand column specifies the information to be provided by the Public Administrations for each process type. Links are specified for characteristics that eventually point to documentary databases or repositories of information about laws, regulatory plans, and regulation documents.

Information about procedures has been required only as a reference to the data categories (i.e., it has not been modeled or developed any further). The three Regions had to specify whether a process is automated or manual, which applications are under development or planned, and the basic information objects involved in each process.

Information obtained through the questionnaire has been the main input to the subsequent phase "Requirements Analysis for the Federated Information System" aimed at:

- specifying the common federated database schema (in Entity-Relationship and XML formats);
- specifying the Web-GIS functional requirements.

PROCESS TYPE	PROCESS CHARACTERISTICS
Polluted Sites Reclamation	
	Reference Organizations
	Basic steps of the procedure
	Timeliness
	Taxes and redemption costs
	Laws/regulations
	Available contributions/financing plans
Building Licenses Request	
	Reference Organizations
	Basic steps of the procedure
	Timeliness
	Taxes and redemption costs
	Laws/regulations
	Available contributions/financing plans
Production Activity Startup	
	Reference Organizations
	Basic steps of the procedure
	Timeliness
	Taxes and redemption costs
	Laws/regulations
	Available contributions/financing plans

Table 2. Summary of process characteristics collected in the data gathering phase

The result of this inquiry phase can be summarized as follows.

Automation of land use plans. This is a subject of ongoing projects in all Regions, in cooperation with the National Land Survey of each Country (the plans will be available as geo-referenced raster maps - tiff). More information on plan regulations is available in paper based documents (that is no plan to digitize them in the near future).

Regional plans and constraints. These are included in the land use plans, which contain all information relevant for a potential exploiter of an industrial site. Often, the maps are stored in ArcView shape-format and data are not stored in a database yet.

Territorial services. This information is available as associated to property units (from real estate registries), that is starting from a type code used for taxation denoting the type of activity on a certain unit (e.g., health care centers, educational activities, sport and resort activities). Other information in this category is stored in ArcView shape-format (e.g., for recreation/free time areas, parks/natural protected areas/historical sites, and so on).

Industrial vacant sites. As the main focus of the gathering phase, vacant sites are described through data stored as pointfeatures in ArcView (or similar) shape-file. No databases are available and most of the information is available as paper based documents. The requirement for DEAFIN is that a database about industrial vacant sites should be built at each regional site according to a federated conceptual schema in the FIS.

Mobility system. Road and railway maps are available, while water networks and interchange access nodes are not available data. A requirement for the DEAFIN project is that data on traffic, speed and mobility should be acquired, and become a central information in the DEAFIN FIS as strongly related to the infrastructures needed to connect a vacant site to the territory.

REQUIREMENTS FOR THE DEAFIN PROJECT
Acquisition of data and implementation of vectorial cartography of land use plans about the areas affected by the presence of industrial vacant sites. Development of the land use plans database. Digital acquisition, in the form of technical documents, of technical norms of land plans of selected sample Municipalities participating in the project.
Construction of cover GIS for areas of interest for DEAFIN and implementation of a related database.
Collection and synthesis of available data on mobility. Street and railway graphs are available in the GIS of each Region. Related data concerning name, type and hierarchy of connections should be collected and implemented in a database.
Database on companies could be acquired from the Chambers of Commerce and other Institutes for the purposes of DEAFIN. More surveys can be performed by DEAFIN partners to collect and aggregate data, especially for areas confining with sites of interest.
Further sets of data (e.g., about statistics, taxation, regional mobility, traffic, and so on) can be bought from relevant sources (National or Local level Public Administrations or Offices) for the purposes of DEAFIN. Information about frequency of data updates, performances, number of accesses, and in general volumes and quantities of information need to be collected in a further gathering round. In general, data are not used as a planning tool for regional development because they do not include any information about foreseen land use.

Table 3. Results of the data gathering phase in terms of DEAFIN project requirements

Statistical and socio-economic data. Data on administrative boundaries are available, while data on primary industrial and trade sites are collected on a Municipality basis. Data on income will be bought by the Regions from Central Statistics Offices or Tax Income Offices.

A requirement for the DEAFIN project is that a database on companies, trends, and market areas should be made available.

Base cartography data. This information, which is interesting for the DEAFIN Web-GIS part, exists in the Regions both in raster and vector data on different scales (tiff and ArcView shape-file). Image catalogs are used for rastermaps and orthophoto. Web access is available, usually only for Regional Technical Maps in raster format.

A set of requirements has emerged for DEAFIN, as reported in Table 3.

3.2 Comments on data gathering and on the federated database requirements

In general, the data collected at the three sites show uniformity with respect to the attention paid to cartographic availability, regulations and laws about reuse of vacant areas, centralization of resources and data. In particular, the need exists in each regional site to introduce tools able to treat heterogeneous data, since these data will more and more intensively come from various data sources, and will need to mapped into the federated schema.

Also the degree of automation is similar, since cartographic systems, and basic data management tools are available at the three sites. Several ongoing projects regard the digital acquisition of land use plans, the automation of document management, and the development of various thematic databases and web sites.

What is required is a unified common schema for the regional databases, where the following basic items are represented:

- available vacant sites;
- surroundings of sites in terms of infrastructure (transportation network, telecommunication etc.);
- profile of a Region in terms of “industrial clusters” and markets;
- regional services: public bodies, institutions, contact persons for several issues, financial or other assistance means, etc.
- regional qualification profiles: potential (qualified) personnel, universities and institutes, etc.
- regional research and development profiles;
- regional economic strengths: gross domestic products, consume data, etc.
- “weak location factors”: cultural, landscape and housing related facilities.

Moreover, the need for a uniform data publishing methodology and policy over Internet web sites emerges clearly from the three participants and involves almost all data categories. The need is stated to correlate data of various categories with one another (e.g., statistical with territorial/cadastral data), in order to augment the synergy among data. The purpose is to gain information for internal analysis (synthesis of data) and for availability to citizens and organizations.

3.3 Usage modes

The data gathering phase has detailed also a set of user profiles in order to specify how a variety of access demands can be supported by the DEAFIN system towards a variety of user groups. The design and implementation of profiles have obviously a deep impact on the usability of the DEAFIN system. Hence, a careful analysis of users typologies and profiles has been performed during the specification phase, while an enrichment of profiles and access modes has been planned in the project in a post-implementation follow up. The purpose is to have the DEAFIN system start with a set of pre-configured access typologies, and then to tune the access modes and user profiles against the more typical uses observed for a fixed period on the DEAFIN pilot implementation.

The interaction with the DEAFIN system differs according to the first broad difference between public administrations and private concerns, due to different data needs. Moreover, two basic modes have to be provided: browsing (using thematic areas and other refined search parameters) and querying (using simple and user-friendly interfaces). Accordingly, the access approach and query support tools can be very different but should substantially adhere to a common interaction mode that we have defined on the basis of user profiles (described in Section 7).

Moreover, the DEAFIN system can be regarded as:

- a *passive tool* when responding to user questions;
- an *active tool* when used as:
    - a decision support system (or a marketing tool) and its contents used to demonstrate the advantages of an area or
    - a *standard application* to which the new European partners should adhere when joining DEAFIN.

As a passive tool, the contents of the DEAFIN federated database can show the advantages and disadvantages of an area: the information provided can be related to the specific search needs of a specific user. These searches will be different according to user types, but apart from the function of location (or re-location) business activities, there probably exists the motivation to retrieve general information about a region.

4. THE DEAFIN SYSTEM ARCHITECTURE

The main objective of the DEAFIN FIS is to improve the quality and the comparability of information about available industrial vacant sites, coming from different data sources of the DEAFIN partner regions. In particular, the DEAFIN system is required to enable information management and access in an integrated way, while each regional site locally deals with information distributed in multiple, autonomous and heterogeneous data sources. In the design of the DEAFIN system architecture three different facets have been considered: *heterogeneity, web availability* and *data variety.*

In the literature, several approaches and tools for handling heterogeneous data sources have been developed and standards for distributed information systems have been defined [Wiederhold 1992, Mylopoulos 1997]. For these systems, the use of multiple layer architectures and of a common data model (mediator data model) able to support the integration and interoperability of the component systems is recommended. Specifically, systems for information integration, with architectures based on mediator data models, have been employed [Molina 1997, Fernandez 1998, Haas 1999, Castano 1999]. In these systems, wrapper/extractor and mediator tools are proposed to obtain a uniform data representation by abstracting from the formats in the origin data sources and to facilitate query and access interface. Following these directions, to handle heterogeneity at each regional site, an architecture based on extractors and mediator has been defined, where extractors translate data coming from local heterogeneous data sources according to a common reference format defined at the mediator level.

Furthermore, nowadays the use of the Web for improving interoperation and information interchange is becoming more and more attractive. Also for the DEAFIN system, the aspect of data publishing over Internet web sites has emerged clearly as a need from the partner regions, and involves most available information. The DEAFIN system is thus conceived as a web-based federated information system, apt to manage and provide access to all the regional relevant information in an integrated and complete fashion, thus satisfying the web availability need.

As for the third facet, we note that, at any given instant, the data available by a given DEAFIN region partner can be both spatial data and alphanumeric data. In the literature, GIS architecture and functions are widely adopted, and Web-GIS technology has been recently discussed and developed. Thus in the DEAFIN system, for each regional component system, a Web-GIS system is defined.

4.1 Overall architecture description

The DEAFIN system is composed of a set of web-interconnected sites:

- the *DEAFIN web site* operates as a central server that contains the DEAFIN interface and some aggregate data;
- each *Regional web site* contains a specific interface and stores regional data in a database, by means of a Web-GIS technology.

The connection modalities between the web-interconnected sites make the DEAFIN system a fully distributed system. Each regional web site is in charge of data publishing and certification over the DEAFIN system according to the common federated schema, while the central DEAFIN web site is the reference site for aggregate data and for requirements about new regional sites to be added in the federation or new applications to be developed in the component systems.

The overall architecture of the DEAFIN FIS is shown in Figure 1.

4.2 The DEAFIN web site architecture

The DEAFIN system has been conceived as a tool that allows users to explore vacant sites and decide where it is more convenient to make an investment or reuse a site on the basis of a set of search parameters. In order to build this tool, comparison capabilities are provided, able to show how the areas differ, using search parameters. The searches can be guided on the basis of predefined forms, search paths, predefined links, and so on. Moreover, the navigation can be free in order to let the user browse in the page structures to find data of interest.

Any user or application belonging to a given Regional site can access data belonging to, or services of, another Regional site by routing its request through the DEAFIN site. In particular, in the DEAFIN site, the user can obtain aggregate information about the vacant sites from Regional sites, using search parameters. The searches can be guided on the basis of predefined forms, search paths, or predefined links.

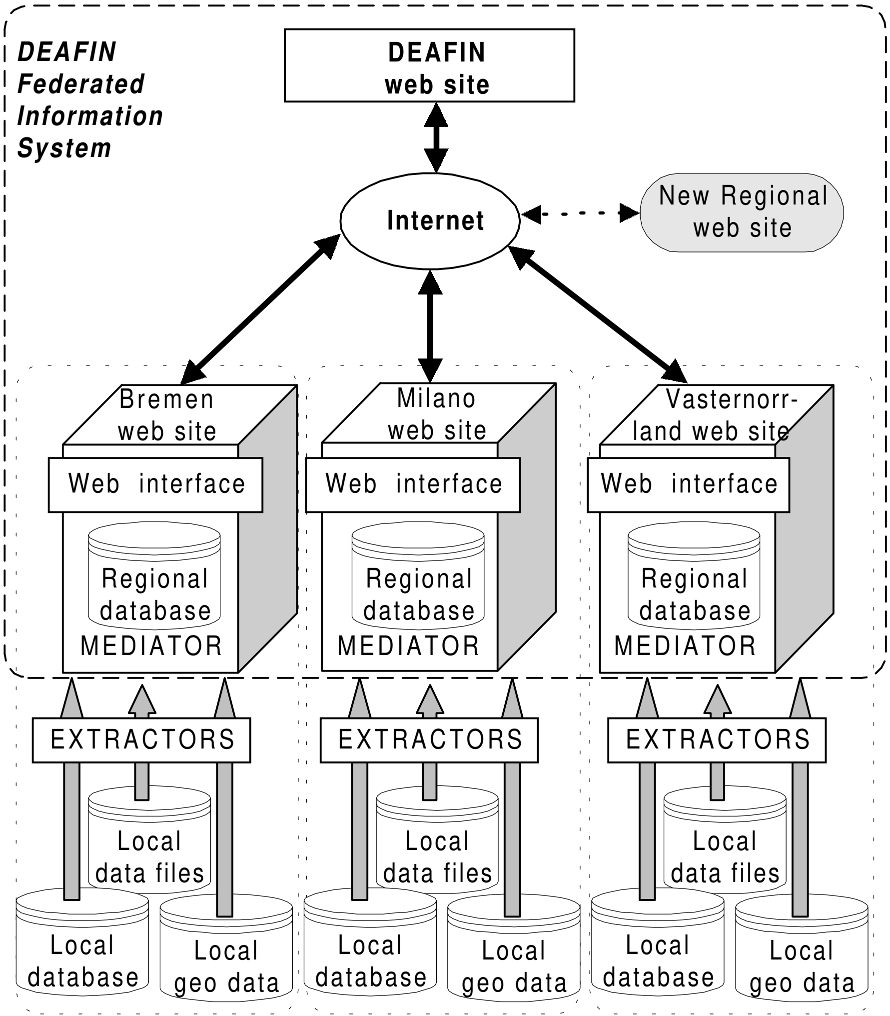


Figure 1. DEAFIN system overall architecture

4.3 Regional web site architecture

A fundamental requirement for the Regional sites is to be autonomous for the feeding and management activities over their local databases. Each regional database contains the set of data available at a certain instant by a given DEAFIN region partner; this is a geographical database including both spatial data in vector format and alphanumeric data usually structured in relational tables. It is populated by data coming from local heterogeneous data sources. Such data must be translated and unified according to a common reference format, employing automatic tools for data translation, properly customized to the data sources [Abiteboul 1999].

To this purpose, extractor modules that work off line and populate the regional databases are requested. Periodically, the extractor modules can be executed to update the database. The common reference format recommended to facilitate data customization, translation and mediation in the DEAFIN project is *XML (eXtensible Markup Language)*. Indeed, among the many data interchange formats proposed in the last years, XML, which was proposed by the World Wide Web Consortium [W3C 1998] for publishing documents over the Web, is the most important example, just because it provides powerful structuring mechanisms and overcomes several limitations of conventional markup languages. In the DEAFIN project, the use of XML is recommended.

5. FEDERATED DATABASE CONCEPTUAL DESIGN

The federated database conceptual schema provides an integrated high level representation of all data to be handled by the DEAFIN FIS. In order to achieve the highest readability and the widest possible diffusion, a common set of terms in needed. We adopted the dictionary of terms as defined by the EuroDicAutom [EuroDicAutom] of the EU: International System of measurements, the Euro currency and commonly agreed data categories were adopted as well.

The database conceptual design, carried out by means of the well-known Entity-Relationship conceptual model [Ceri 1992], produced a set of schematic requirement representations, each related to a specific information set. The reason for leaving these schemas separate is mainly readability. Such schemas are called *conceptual views*: they allow the smooth and uniform integration into the DEAFIN conceptual schema of geographic as well as conceptual information, and any further information need which might be identified in the future. In fact, the DEAFIN federated database is intended as a dynamic structure which, leaving from the preliminary structure, shall accommodate both new entities and attributes which will become useful in future applications and possibly new views adopted by new partners joining DEAFIN in the future.

5.1 Geographical Information Systems

As the data gathering phase has confirmed, a significant part of the data describing the vacant sites of DEAFIN partners are of geographical nature. The aim of the DEAFIN project includes the use of geographical data as an effective correlation means among the information about vacant sites collected from different sources.

Therefore, in the conceptual design of the DEAFIN FIS we have explicitly marked the entities of the database schema that have a representation on the territory, calling them “Geo-referenced entities”. The occurrences of such entities have a spatial representation that describes their shape, extension and location on the earth surface.

Geographical data always include both alphanumeric and spatial components. It is this spatial component that makes geography so difficult to manage by traditional database management systems and that generated at the end of the 80s a specific class of systems called Geographical Information Systems (GIS). This systems provided the data structures and algorithms able to represent and to query efficiently a collection of geographical data [Scholl 1989, Tomlin 1990, Gargano 1991, Güting 1994, Güting 1995].

In the last few years we see an increasing integration between GIS and web technologies and now many GIS products include in their standard suite also a web interface tool able to assist the user in publishing geographical data on the Web. Confirming this trend also the OpenGIS Consortium has recently proposed an XML standard for geographical data, namely Geographical Markup Language (GML) [OpenGIS 2001], and the same Consortium has defined the standard characteristics of a Web-GIS system [OpenGIS2001b].

DEAFIN FIS aims at being a Geographical Information System accessible via web. Thus it should enable the user to dynamically query the databases and in particular the geographical data they contain. This requires a Web-GIS system able to perform advanced operations guided by the client side (see Section 6.2).

5.2 Conceptual design

For the DEAFIN federated database, the following views have been defined: Vacant Site Description view, Land Use Plan view, Administration view, Eco/Reclamation view, Transportation view, Service view, Document view, Procedure And Task view. The global conceptual data schema is a merge of the different views obtained by superimposing entities with the same name.

In the following, we briefly show and discuss the schemas of the two most application peculiar views: the Vacant Site Description and the Land Use Plan. A complete description of the defined views is provided in [Deafin 2001].

The *Vacant Site Description* view is the central focus of the investigation and the core of the global schema: all other views are strictly related to this main view. In particular, the *Vacant Site Description* view (see Figure 2.a) provides information about the main vacant site features, including the site topography/geology features, the contained buildings, the hosted production activities and the serving technological networks; the *Land Use Plan* view (see Figure 2.b) provides information about the physical zone where the vacant site is located, including the neighborhoods and the plans, issued possibly by different administrations, ruling the entire area.

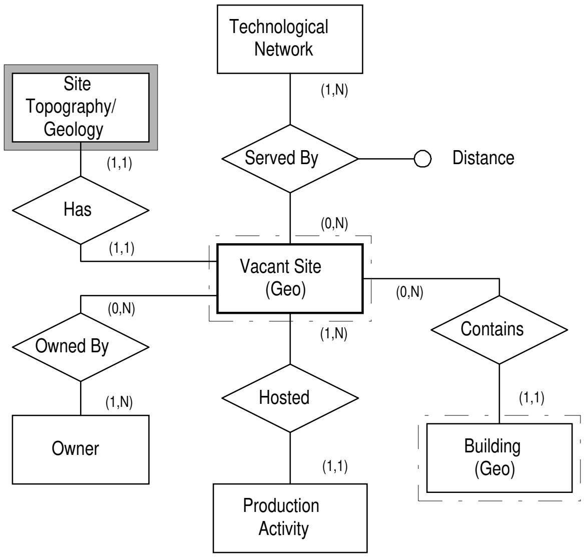


Fig. 2a: The VACANT SITE view

The conceptual design phase led to the definition of 36 entities, with a gross total number of about 201 attributes. Next, the *logical design* phase produced the *logical specification* of the DEAFIN conceptual schema ready for use by the appropriate commercial systems. For this purpose, XML (eXtensible Markup Language) was chosen as logical specification format, providing a common reference format to facilitate data customization, translation and mediation. XML has been proposed by the World Wide Web Consortium [W3C 1998] for publishing documents over the Web, just because it provides powerful structuring mechanisms and overcomes several limitations of conventional markup languages. An example of an XML code describing the VACANT SITE entity, its attributes and its relationships, is depicted in the following figure. An immediate correspondence can be drawn between this specification and the data structures proper of the commercial systems available at each single DEAFIN partner's site, such as a relational or object oriented database. In order to automatically derive the XML code from the conceptual schema we deployed a tool, namely SiteDesigner (see Figure 3), developed at the Politecnico di Milano within the EC funded project W3I3 [Ceri 1999].

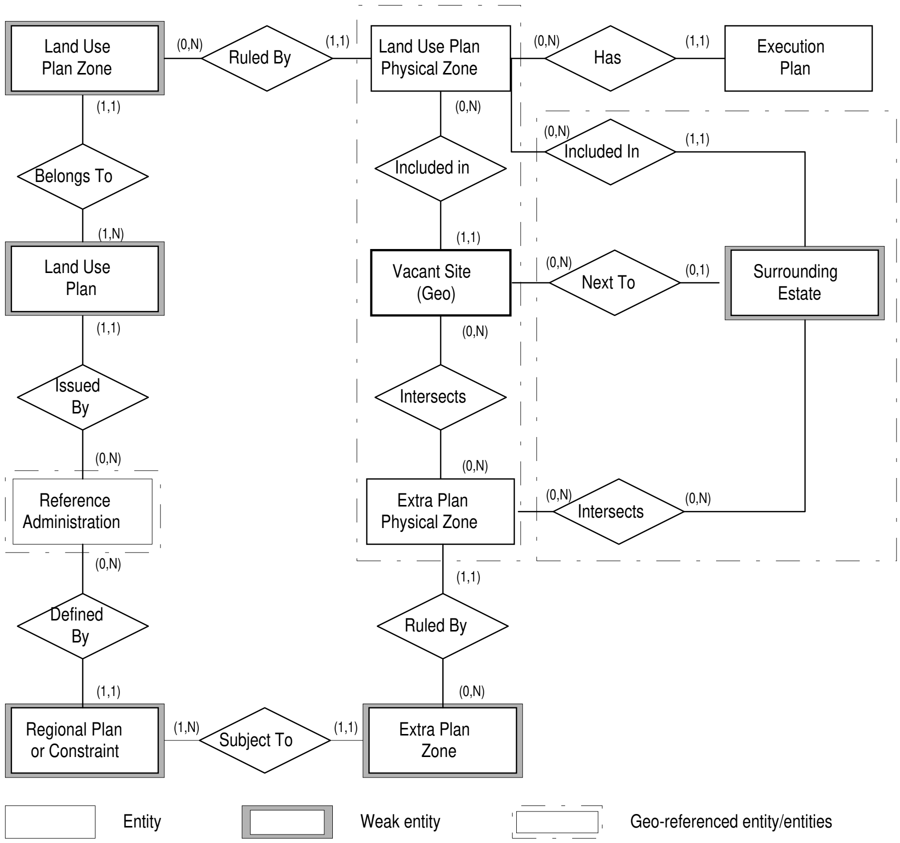
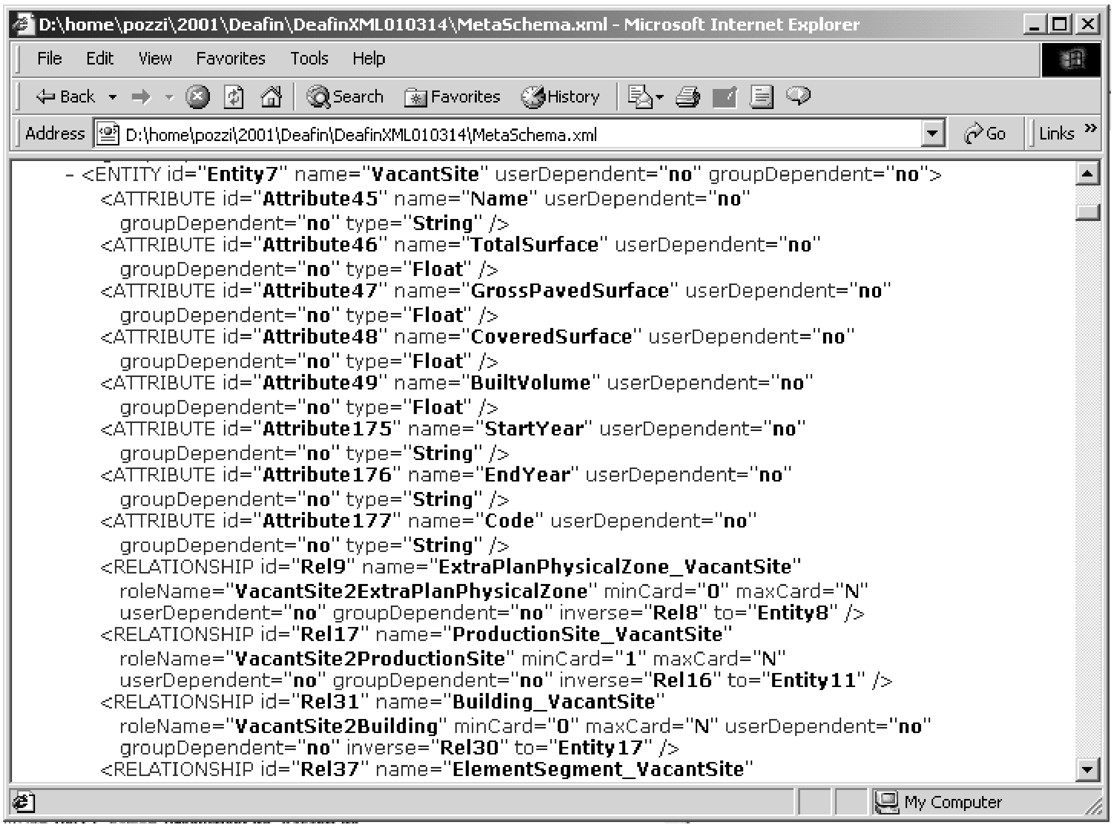


Fig. 2b The LAND USE PLAN view




```
- <ENTITY id="Entity7" name="VacantSite" userDependent="no" groupDependent="no">
    <ATTRIBUTE id="Attribute45" name="Name" userDependent="no"
      groupDependent="no" type="String" />
    <ATTRIBUTE id="Attribute46" name="TotalSurface" userDependent="no"
      groupDependent="no" type="Float" />
    <ATTRIBUTE id="Attribute47" name="GrossPavedSurface" userDependent="no"
      groupDependent="no" type="Float" />
    <ATTRIBUTE id="Attribute48" name="CoveredSurface" userDependent="no"
      groupDependent="no" type="Float" />
    <ATTRIBUTE id="Attribute49" name="BuiltVolume" userDependent="no"
      groupDependent="no" type="Float" />
    <ATTRIBUTE id="Attribute175" name="StartYear" userDependent="no"
      groupDependent="no" type="String" />
    <ATTRIBUTE id="Attribute176" name="EndYear" userDependent="no"
      groupDependent="no" type="String" />
    <ATTRIBUTE id="Attribute177" name="Code" userDependent="no"
      groupDependent="no" type="String" />
    <RELATIONSHIP id="Rel9" name="ExtraPlanPhysicalZone_VacantSite"
      roleName="VacantSite2ExtraPlanPhysicalZone" minCard="0" maxCard="N"
      userDependent="no" groupDependent="no" inverse="Rel8" to="Entity8" />
    <RELATIONSHIP id="Rel17" name="ProductionSite_VacantSite"
      roleName="VacantSite2ProductionSite" minCard="1" maxCard="N"
      userDependent="no" groupDependent="no" inverse="Rel16" to="Entity11" />
    <RELATIONSHIP id="Rel31" name="Building_VacantSite"
      roleName="VacantSite2Building" minCard="0" maxCard="N" userDependent="no"
      groupDependent="no" inverse="Rel30" to="Entity17" />
    <RELATIONSHIP id="Rel37" name="ElementSegment_VacantSite"
```

Figure 3. A screenshot of the XML code describing some properties of the VACANT SITE entity

6. THE WEB GEOGRAPHICAL INFORMATION SYSTEM

The aim of the web interface of the DEAFIN system is to provide a set of functionalities to explore the available information about the vacant sites, allowing the user to make a benchmark among different investment options. The web interface has been designed in order to allow the user to access data in two search modes:

- *direct search*: it leads directly to a vacant site of interest and then the user can view all the available information about it including also spatial data and enrich this view considering the information about the area surrounding the site;
- *navigation*: a list of vacant sites is available and can be filtered progressively in order to see a set of sites, to select a list of interesting sites and then make a comparison within the selected list.

The two search modes can be interleaved. Moreover, the user is free to enter a search session composed of alternate steps of browsing and querying; a session can be saved in terms of traversed links, search paths, and retrieved data. It is also possible to re-execute a search session by slightly varying the search parameters to implement a "what if" tool: DEAFIN should in fact allow to perform a "what if" analysis that provides a benchmark among the different vacant sites.

6.1 The web interface

As described in Section 4 the architecture of the DEAFIN web site is composed of a DEAFIN site and a set of DEAFIN regional sites.

The *DEAFIN site* presents information about the project: main targets of the project, list of partners, agreement rules, map of the web site and contact information. Moreover, it offers the user the following paths either to view and query some aggregate information about the areas or the vacant sites (navigation mode) or to go directly to a specific regional site (direct search):

- the exploration of general information about all the DEAFIN partners' areas. This path should present a static benchmarking table, where some attributes are listed with the correspondent values for each partner. From the benchmarking table the user can choose a partner and obtain some pages that describe the partner's area highlighting its most attractive features;
- the visualization of some aggregate data about the vacant sites of each area by means of a benchmarking table. On this table the user can execute some queries in order to filter the vacant sites on the basis of preferences or specific criteria such as: the type of usage, area size, costs, surrounding services, or accessibility. Consequently, the data of the benchmarking table are filtered for subsequent browsing operations;
- the direct link to each specific regional sites is also provided.

The *DEAFIN regional sites* contain detailed information about each partner area in general and about its vacant sites (Figure 4.a). Moreover, it offers the following paths:

- the exploration of the partner area by choosing a specific theme (Figure 4.b). The available themes are: land planning, business and market activities, mobility system, services, demographic/macroeconomic data and development strategies. Each theme shows some descriptive and spatial data, presented in form of reports and maps respectively;
- the access to business information: this path leads to pages describing contact offices and particular business opportunities that are currently available for some vacant site of the area;
- the access to aggregate data about vacant sites of the area: this path leads to a benchmarking table where the user can also apply some selection conditions based on the available comparison parameters;
- the selection of a vacant site. The path leads to a collection of pages that describes a single vacant site. Specific themes presenting the characteristics of the chosen vacant site are available in particular: buildings, eco quality, technological networks, land planning, mobility system, territorial services, procedures and tasks and documents;
    - the search of vacant sites based on forms that guide the user in the specification of the query (guided search pages). The interaction can be as follows: the user can specify some search conditions on the attributes proposed in each guided search page and decide the logical operator (and/or) to connect the different conditions; moreover, the guided search page can present a pre-defined spatial condition, like "search the vacant sites with a railway station within a radius of <parameter value> km" and the user can set the value of the parameters. The user can perform the selection on the entire data set of vacant sites or on the currently filtered data set. The second choice permits to perform a sequence of selections that restrict progressively the list of vacant sites of interest;
    - the search of vacant sites through advanced queries (advanced search pages).

These pages allow the user to specify a generic query on alphanumeric and spatial data. In particular, each advanced search page contains:

- a map section showing the map of the area with the localization of the vacant sites;
- the list of layers (set of spatial data) that can be visualized and queried;
- the attributes available for each layer;
- a tool to specify a generic search condition obtained by connecting (using logical operators) elementary conditions on the attributes of a layer;
- a tool to specify a spatial condition on the available layers. The spatial conditions can be: conditions on the spatial characteristics of the spatial data contained in a layer (area, length, shape, etc…); conditions on the spatial relation (or Euclidean distance) between a shape (point, line or polygon), drawn by the user on the map, and the

spatial data of a layer (for instance, “find all vacant sites that are within 10Km from a point that the user draws on the map”);

- a tool to specify a spatial joins among layers, where the join condition is based on: the Euclidean distance; the spatial relations (not empty intersection and containment); the distance function on the transport network;
- a tool to identify a location through its address on the road network (geo-coding).

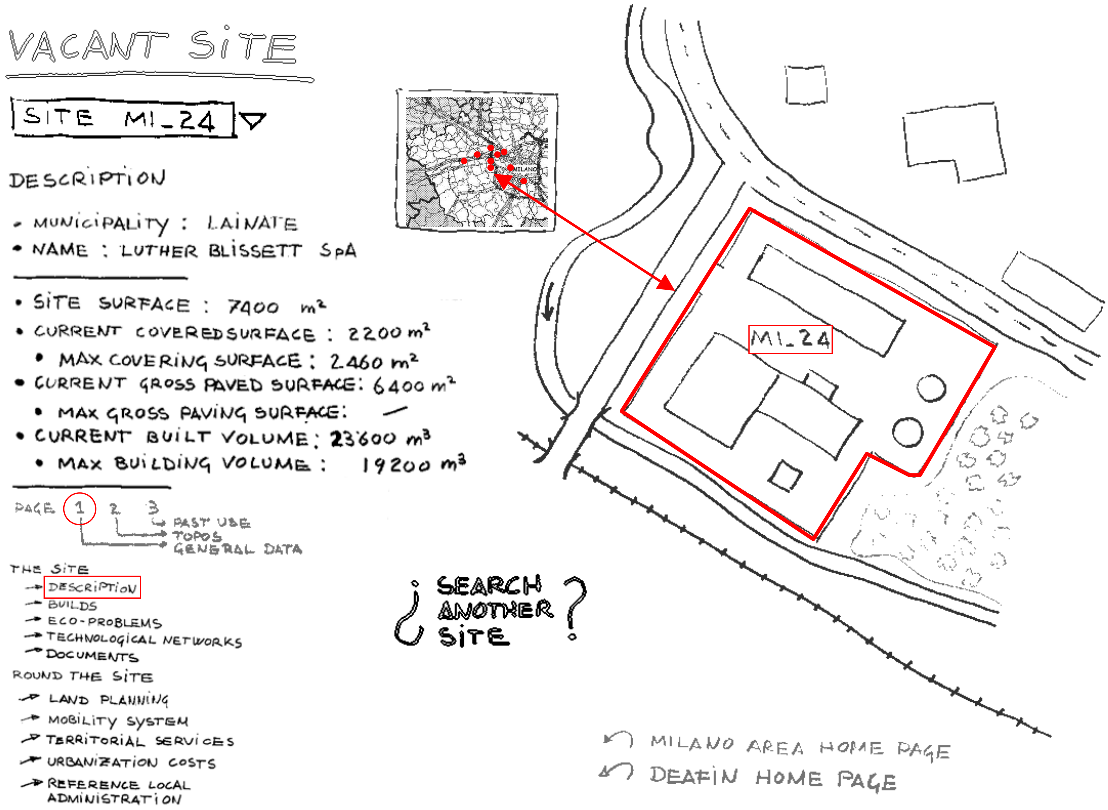


Figure 4.a Example of the web interface – general information about a vacant site is displayed

The last two paths offer interesting tools to query the spatial data describing the vacant sites. In particular the advanced query pages provide powerful operations for manipulating spatial data like, for instance, the spatial join, that permits the user to integrate data on the basis of their relations on the territory.

Furthermore, the content of a page can be affected by the user profile. This means that each section of a page and each item in the section can be shown or not according to the profile the user has specified in the home page of the DEAFIN site (or DEAFIN regional site). In the starting configuration all user profiles can access all pages and all information inside the pages. The system administrator can modify this configuration also considering the result of the monitoring activity.

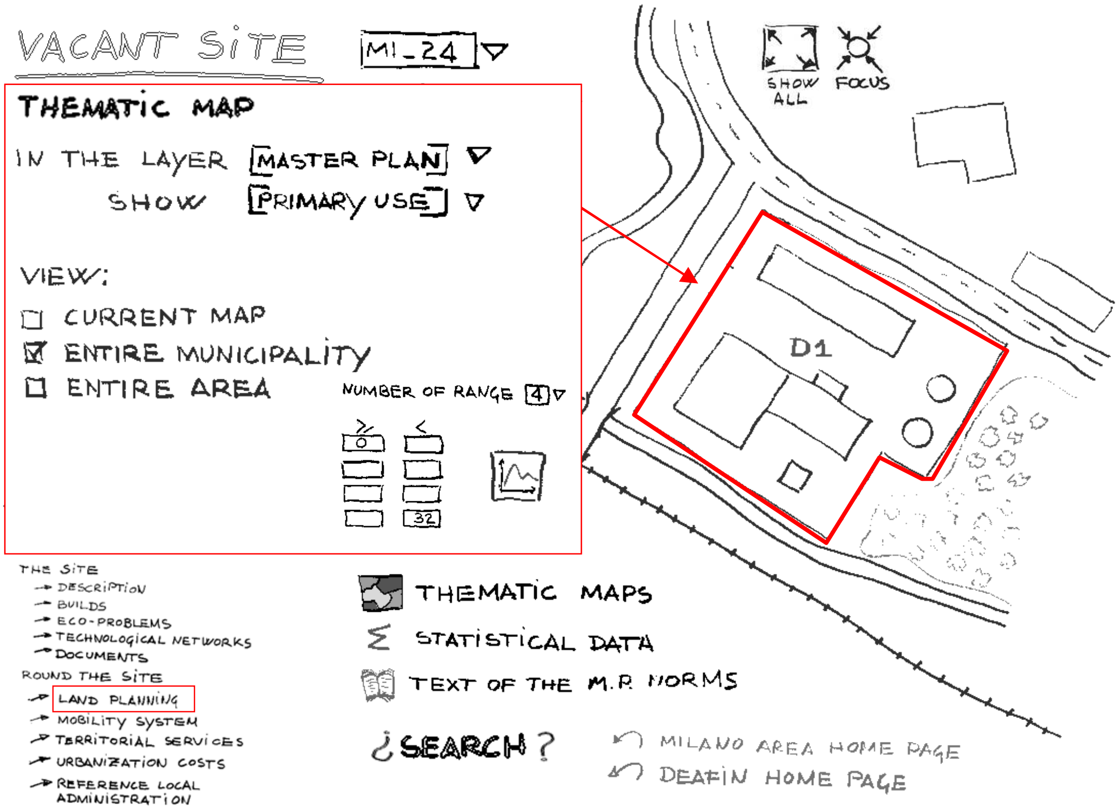


Figure 4.b Example of the web interface: after a vacant site selection, additional information are displayed

6.2 The functional requirements for the Web-GIS system

The choice of the Web-GIS system is strategic for the DEAFIN FIS since the main modules of the proposed architecture (see Figure 1), the mediator and the Internet interfaces have to be developed using a Web-GIS system.

The basic requirements for choosing the Web-GIS system can be identified by considering the main characteristics of any open system that generates web pages by extracting data from a Data Base Management System (DBMS). Among them the following ones are of particular importance for a distributed system that works on the Internet and manipulates geographical data:

- the system should run on different platforms
- the web-pages generated by the system should be readable from any internet browser (possibly only a light plug-in on the client side might be acceptable)
- the system should be able to connect to external DBMS's
- the system should be able to handle and generate XML documents
- the system should provide different strategies to efficiently send data describing maps (images/vector data) through the Internet
- the system should be able to minimize the amount of data which is exchanged between the http server and the browser.

Together with these basic requirements a collection of functional requirements have been defined by considering the design of the web interface of the DEAFIN system, which is illustrated in the previous subsection.

A. The data model of the Web-GIS system should include the following concepts:
   - layer: it is a set of feature in vector format (points, lines and polygons), where each feature has and identifier and might have some other descriptive attributes;
   - table (relation): it is a traditional table of the relational model with a schema (set of attributes or columns) and a content (set of tuples or rows). The feature identifier is usually used to link tuples (rows) of a table to the features of a layer.

B. The system should provide the programmer with the tools that allow the implementation of a dynamic interaction with geographical data. This means that the operations for querying, visualizing and analyzing geographical data which are usually present in the widely used GIS systems (for example: zooming operations, labeling operations or querying operations based on alphanumeric or spatial criteria, see for instance the GIS ArcView®, GeoMedia® or MapInfo®), have to be available also for the implementation of the client interface.

C. Some more advanced GIS functions should be provided by the system, in particular
   - the identification of a location through its address on the road network (geocoding) and
   - the calculation of distances on the transport network.

Note that the interaction with the user has to be done by a browser and therefore the development environment has to provide the tools that can perform this type of interaction on the client side.

Indeed the propagation of the traditional operations of a GIS interface to a remote interface with the Internet as transmission channel using the http protocol is not straightforward. Problems arise from the stateless property of the http protocol (each http request is completely independent from the previous ones) and from the large size of the geographical data usually rich of images. This situation produces the exchange of large amount of data between the http server and the client in both directions. A possible solution for this situation is to transfer a subset of functionalities on the client by means of Java applets or Javascript code. However, we think that this solution is not the better one since it requires the availability of additional resources on the client side, and this can limit the access to the web pages. The Web-GIS should provide alternative solutions to this problem for example by managing user sessions or by adopting different resolution levels when the interaction with the user involves the transmission of images.

7. ADVANCED SYSTEM FUNCTIONALITIES: PROFILING AND TUNING

This section describes the DEAFIN user typologies and access interfaces, giving the user profiles and the related access to the databases and to the Web, and briefly describes the DEAFIN Monitoring and Tuning System (MTS) that slowly adapts the system parameters according to the observed usage.

7.1 User typologies

DEAFIN users are of different nature and typologies: Public Administrations, public and private organizations, and citizens. Their purpose is to explore vacant sites and related investment opportunities, but queries are not limited to such purpose. We describe the basic user typologies and envision the use of DEAFIN on the basis of what emerged from the data gathering activity.

The user profiles in DEAFIN are defined at initialization time using the known categories of users. The user profile is made of two parts: <user ID, role>.

- the *user ID* comprises the user name and his/her authentication data consisting of a password, or a certificate allowing him/her to enter the parts of the system he/she has allowed access to. Allowed parts are views on the database and web pages;
- the *role* belongs to one (or more) of the pre-defined roles. Examples are: real estate agency users, public administration users, or private company users. Examples are "internal" roles, such as the DEAFIN site employee, the site administrator, the developer, while "external roles", are the users who browse the DEAFIN system to get information, such as a financing agent, a real estate agent. The role is characterized by a set of preferences seen as data and navigation paths for the user and are used to filter data and pages, using the thematic area the user is interested in. Initially, a preference filtering function sets the search parameters; then, using the DEAFIN Monitoring and Tuning System (MTS) described below, the DEAFIN system can store other preferences on the basis of the observed use modes.

Sample users and roles in DEAFIN, considering, as an example, the activity of *location or re-location companies*, are companies of various size and fields of business activities, real estate agents specialized in corporate clients, regional authorities or regional agencies.

The main difference between companies and other users is that a company:

- may already have information about a target region, based on business experiences with partners in a target region and may try to expand such information by acquiring additional data about a target region either for development of business activities or for location purposes (e.g. data about the regional support structure, financial support, available workforce, etc.);

- may have only some information about a target region and need strategic information on that specific region;
- will probably first search for contacts with regional authorities, agencies or chamber of commerce to draw further specific information and/or for professional support during the location process.

The processes of developing a strategy and taking location decisions may differ due to the size of the company and/or the field of business activities. The DEAFIN system is seen as a passive tool in that it simply responds to questions, which are usually not site-related questions. In fact, location decisions may be only the first step of decision making for business activities: companies may have developed business activities in a certain Region or plan business development in a Region. In this second case, a location decision comes after a survey of regional opportunities. In all cases, business activities and parameters, such as the proximity to Regional markets or infrastructures of importance for a certain company, have already been explored.

What discussed so far underlines that DEAFIN contains both general and specific information, that must be presented in a user-friendly way and allow for multiple-access to data, apart from the simple site-oriented view of information.

Regional authorities seldom use the DEAFIN system to collect specific information about other regions (or at least not site-related information), because regional authorities transfer the location business to affiliated agencies. Rather, regional authorities use DEAFIN for marketing their own regions. In this case, DEAFIN can be seen as a Decision Support System, in that it provides information for an unspecified target person whose search operations are not completely known. Therefore, DEAFIN should allow to expand the query search or to restrict it in order to let the user find a set of ranked candidates and then choose the "best fitting". We will discuss this aspect when describing the MTS, that observes the user behavior and tunes the search paths to ameliorate the precision and/or recall of the system.

Summarizing, the DEAFIN user profile can be: *statically determined* at system initialization; *dynamically modified* during the system life time to mirror the usage of DEAFIN performed by the users.

Modifications should be done semi-automatically, i.e., using a set of tuning system data about preferred query paths, entry points (pages, data sources), data formats (synthesis data, detailed data, and so on), most visited pages and data categories (mobility, territory, vacant sites, and so on). After a default preference setting, DEAFIN tunes its performances according to the observed user actions. This adaptive property is wrapped in a MTS module that will be shortly presented in the following.

7.2 System monitoring and tuning

A full exploitation of the DEAFIN system needs tuning techniques to make the use of the system convenient and user friendly. The proposed method consists in initializing DEAFIN to a default setting and then to employ a software module that continuously observes the user actions in order to tune:

- *the user profiles*;
- *the page sequences* according to preferred navigation paths and links;
- *the federated database contents and views*, according to the preferences and to the choices of users.

From one side, we have the *classification problem*, i.e., the problem of defining the web page contents and the user categories and the navigation paths seen in the previous paragraph. From the other side, we have the *retrieval problem* related to the search modalities and the query sessions.

The MTS (Monitoring and Tuning Systems) is the DEAFIN module that improves the efficacy of the search and navigation: it is functionally is represented in Figure 5. The user feedback is taken into account by recording the query/browsing element (or session) and the returned answer. The log file records the parameters of a query session. A possible set of parameters is:

- user profile;
- thematic area of the search;
- session id;
- traversing of marked (interesting) nodes;
- marking by the user of new nodes;
- set of answers with *rank* rates;
- position selected by the user among the returned set;
- difference between suggestion of system (best rank) and user choice (rank of user-selected elements).

These parameters are interpreted by the MTS to slowly modify the DEAFIN system. Both the DEAFIN system and the data dictionary are initialized with parameters set up by the developers. Then, the MTS automatically ameliorates the answer quality along the system life cycle employing a Quality Function (QF).

The properties of QF (details can be found in [Damiani 1999]) regard the observation of the user reactions to the query answers from the system, and in batch mode, has to slowly change the contents and structure of the system components, such as the data dictionary or the page and database contents.

In order to weight the quality of the answering system, ranks can be assigned to

- web (set of) pages
- search paths, or query sessions.

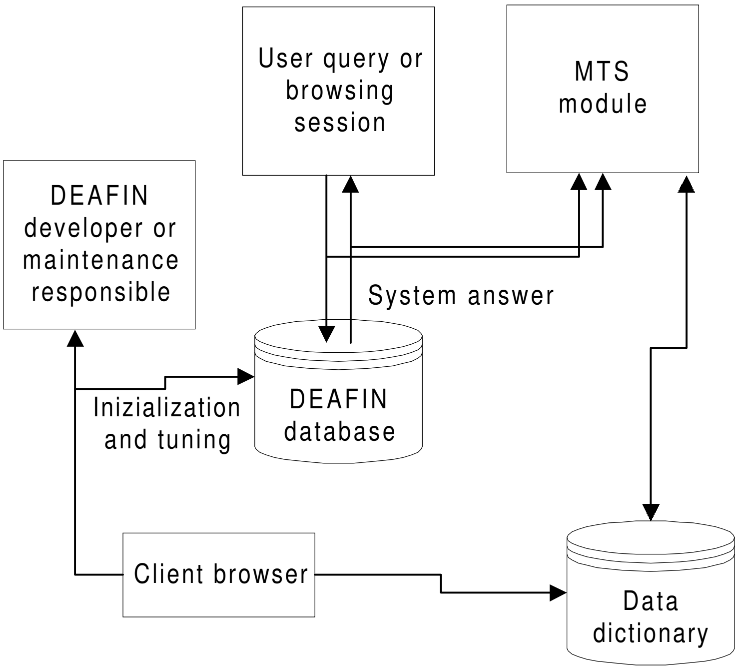


Fig. 5. The DEAFIN Monitoring and Tuning System. The MTS module observes user session, logs user query and choice in log file, supports query refinement and tunes the data dictionary and the user profile. The Data dictionary contains lists of contexts and themes, terms and user profiles.

Ranks are assigned to a system element when selected, or can be expressly assigned by the user upon request when completing a query or browsing session. Ranks are processed in batch mode by DEAFIN as a feedback of a navigation session. The modifications bias the system components in order to reflect the views of the DEAFIN user, who are considered as "experts" whose opinions about DEAFIN data and access modes are polled simply by registering their choices. Hence, the system learns by registering the user (implicit or explicit) ranking.

8. CONCLUSIONS

In this paper we presented the work of the design team of the DEAFIN Federated Information System. Starting from an analysis of the requirements and of the local availability of information, performed through a data gathering phase, we defined:

- the architecture of the overall system, in the form of a federated Web-GIS;
- the contents of the DEAFIN database, in the form of a set of conceptual views both as ER diagrams and in XML;
- the outline of the end-user interface;

- the functional requirements of the Web-GIS system and a set of advanced functionalities for the system management.

The proposed DEAFIN system offers a marketing and information platform in a European context based on advanced web functions integrating Geographic Information Systems (GIS), decision support tools, and database federation features. Analogous projects in the European area have been developed mainly in the field of tourism [Pühretmair 2001] where integration of data sources and access to GIS in a graphical context are needed.

Acknowledgements

The design of the DEAFIN FIS, while the responsibility of the authors, has benefited by many interactions with the regional partners, mainly as to the data gathering and to the view design activities. Therefore we want to acknowledge the cooperation given by Gian Paolo Corda of Politecnico di Milano, Marcello Correra and Luca di Giulio of Provincia di Milano, Christian Böttger, Wolfgang Schmidt and Oliver Steck of the Länd of Bremen, Anders Larsson of Länsstyrelsen Västernorrland and last, but not least, Leif Kärreby, the DEAFIN Project Director. We also want to thank the W3I3 team who developed the Site Designer tool we used.

References

[Abiteboul 1999] Abiteboul S., Cluet S., Milo T., Mogilevsky P., Siméon J.,& Zohar S. (1999). Tools for Data Translation and Integration. IEEE Data Engineering Bulletin, 22, 3-8.

[Castano 1999] Castano, S., &De Antonellis, V. (1999). A Schema Analysis and Reconciliation Tool Environment for Heterogeneous Databases. IDEAS conf., 53-62.

[Ceri 1992] Ceri S., Batini C., & Navathe S. B. (1991). Conceptual Database Design. Redwood City, Ca: The Benjamin/Cummings.

[Ceri 1999] Ceri S., Fraternali P., & Paraboschi S. (1999). Data-Driven, One-To-One Web Site Generation for Data-Intensive Applications. VLDB conf., 615-626.

[Damiani 1999] Damiani E., Fugini M.G., & Bellettini C. (1999). A Hierarchy-aware Approach to Faceted Classification of Object-Oriented Libraries. ACM Transactions on Software Engineering and Methodology, 8, 215-262.

[Deafin 2001] Politecnico di Milano (2001), Deafin Federated Information System: Technical Annex – Appendix 2.

[EuroDicAutom] Automatic Translation of Terms, European Community, (2001). http://eurodic.ip.lu:8086/cgi-bin/edicbin/EuroDicWWW.pl?page= expert

[European 2000] European Commission, (2000). Homepage of the IST.1999.12030 project: XML-based Mediator for Knowledge Extraction and Brokering, http://www.cordis.lu/ist/projects/99-12030.htm

[Fernandez 1998] Fernandez M.F., Florescu D., Kang J., Levy A.Y., & Suciu D. (1998). Catching the Boat with Strudel: Experiences with a Web Site Management System. SIGMOD Conference, 414-425.

[Gargano 1991] Gargano M., Nardelli E. & Talamo M. (1991). Abstract Data Types for the Logical Modeling of Complex Data. Information Systems, 16, 565-583.

[Güting 1994] Güting R.H. (1994). An Introduction to Spatial Database Systems. VLDB Journal, 3, 357-399.

[Güting 1995] Güting R.H., & Schneider M. (1005). Realm-Based Spatial Data Types: The ROSE Algebra. VLDB Journal, 4, 243-286.

[Haas 1999] Haas L.M., Miller R.J., Niswonger B., Roth M.T., Schwarz P.M., & Wimmers E.L. (1999). Transforming Heterogeneous Data with Database Middleware: Beyond Integration. IEEE Data Engineering Bulletin, 22, 31-36.

[Molina 1997] Garcia Molina H., Papakonstantinou Y., Quass D., Rajaraman A., Sagiv Y., Ullman J., Vassolos V., & Widom J. (1997). The TSIMMIS approach to mediation: data models and languages. Journal of Intelligent Information Systems, 8, 117-132.

[Mylopoulos 1997] Mylopoulos J. & Papazoglou M. (Ed.). (1997). Cooperative Information System [Special issue]. IEEE Expert, 12(5).

[OpenGIS 2001] OpenGIS Consortium. (2001). Geography Markup Language (GML), 20 February 2001, OGC Document Number: 01-029

[OpenGIS 2001b] OpenGIS Consortium. (2001) Web Map Service Implementation Specification, 21 June 2001, OGC Document Number: 01-047r2

[Politecnico 2001] Politecnico di Milano Working Group, DEAFIN Federated Information System Technical Annex, March 2001

[Pühretmair 2001] Pühretmair F., & Wöβ W. (2001). XML-Based Integration of GIS and Heterogeneous Tourism Information. In Dittrich K., Geppert A., Norrie M. (Eds.), Advanced Information Systems Engineering, Berln: LNCS Springer Verlag.

[Scholl 1989] Scholl M., & Voisard A. (1989). Thematic Map Modeling, SSD conf., 167-190.

[Tomlin 1990] Tomlin C.D. (1990). Geographic Information Systems and Cartographic Modeling. Upper Saddle River, NJ: Prentice Hall.

[W3C 1998] World Wide Web Consortium. (1998). Extensible Markup Language (XML), Version 1.0. W3C Recommendation, http://www.w3.org/TR/1998/REC-xml-19980210.html

[Wiederhold 1992] Wiederhold G. (1992). Mediators in the Architecture of Future Information Systems. IEEE Computer, 25, 38-49.