XML-Based Integration of GIS and Heterogeneous Tourism Information

Franz Pühretmair and Wolfram Wöß

Institute for Applied Knowledge Processing (FAW)
Johannes Kepler University Linz, Austria
{fpuehretmair, wwoess}@faw.uni-linz.ac.at

Abstract. With the tremendous growth of the Web, a broad spectrum of tourism information is already distributed over various Web sites. To fulfill the tourists request for an extensive data collection it is inevitable to make accumulated data from different sources accessible. In a first step towards a comprehensive integration of tourism data the official Austrian destination information and booking system TIScover is extended with a flexible data interchange adapter which allows interchange of structured data with other tourism information systems.
Beside the problem of distributed data sources tourists are also confronted with differences concerning information presentation on various Web sites. To cope with this problem our approach extensively uses maps for data presentation and federation of multiple structured and semi-structured tourism information sources on the Web. For this purpose touristic maps are generated dynamically including data resulting from database queries. This concept allows a clear and meaningful representation of up-to-date tourism information embedded in a geographical context.

Key words: geographic information systems (GIS), tourism information systems (TIS), information integration, electronic data interchange (EDI), meta data, extensible markup language (XML), scalable vector graphics (SVG).

1 Introduction

The aim of the official Austrian tourism information system TIScover [1], [2], [3] is twofold: First, tourists should be supplied with comprehensive, accurate and up-to-date tourism information on countries, regions, villages and all destination facilities they offer like hotels, museums or other places worth seeing. Second, it aims to attract tourists to buy certain tourism products either offline or even more important to allow tourists to buy them online.

The functionality provided by TIScover can roughly be categorized into three different components, the *public Internet* component, the *Extranet* and the *Intranet*. The *public Internet* component comprises that functionality of the system that is accessible to the public, the most important modules are Atlas and Booking. The *Extranet* provided by TIScover allows authorized tourism information providers, regardless whether it is a small guesthouse or a large local tourist office to update and extend their tourism information and products directly. Finally, the *Intranet*

K.R. Dittrich, A. Geppert, M.C. Norrie (Eds.): CAiSE 2001, LNCS 2068, pp. 346–358, 2001.

component of TIScover which is accessible at the system provider's side only allows to configure the whole system in various ways. For example, it is possible to extend the geographical hierarchy, to specify expiration dates for reports and to define the default language for all system components.

Currently, TIScover manages a database of about two gigabyte of data, more than 500.000 Web pages (composed of more than one million files) covering among others 2.000 towns and villages and over 40.000 accommodations. The system has to handle up to 21 million page views as well as up to 90.000 information and booking requests per month.

Although, these figures illustrate that TIScover manages a fairly huge amount of tourism information, it is of course far from being complete. To be able to satisfy also requests for certain information which is not part of the TIScover database, but already available at other information systems, the IST project XML-KM (eXtensible Markup Language-based Mediator for Knowledge Extraction and Brokering) [4], [5], [6] was intended to provide a proper technical basis.

Resulting of many years of project experience in the field of tourism information systems, we identify the following requirements having great impact on the quality of a tourism information system focusing the accessibility to the users (tourists):

- integration of geographical data with tourism data and
- integration of distributed data sources.

It is obvious, that in the actual situation to search for and to locate interesting tourism information distributed over various Web sites takes a great effort, is time consuming and burdens tourists. The situation gets even worse since Web sites usually differ very much especially concerning information presentation and information access. For a satisfying result the user needs a fast and straightforward access to tourism information with a reliable and clear data representation. Typically, if tourists search for objects they are interested in, they often have to deal with large result sets characterized by complex tabular lists where it is difficult to obtain an objective view. In the case of integrating tourism data with a geographic information system (GIS), maps offer a unique possibility to combine geographical data with plain tourism data at the visualization level. Maps are a proper medium to support the human perception in a convenient way and use people's inherent cognitive abilities to identify spatial patterns and provide visual assistance concerning geographic objects and their locations.

Today most graphics on the Web are still in raster format. The new XML-based graphic specification language scalable vector graphics (SVG) [7] allows the presentation of dynamic and interactive vector graphics in the Web. SVG offers substantial advantages for tourism information systems because maps change from static raster graphics to interactive graphical representations allowing the presentation of the most extensive information possible thus satisfying the demands of the users.

To fulfill the tourists request for an extensive data collection it is inevitable to make data from distributed data sources accessible. In a first step towards a comprehensive integration of tourism data TIScover is extended with a flexible data interchange adapter which allows interchange of structured data with other tourism information systems. Information about data interchange formats and data structures is transformed into a meta data structure represented by XML (extensible markup language) [8], [9], [15] document type definitions (DTDs). This concept allows a flexible administration and maintenance of several data interfaces to other tourism information systems.

The contribution of the presented approach is twofold: First, to provide an extensive data collection to a tourist, distributed and heterogeneous tourism information is accumulated and integrated. Thus, flexible data interchange mechanisms are introduced. Second, tourism information is combined with geographical information to achieve efficient visualization of tourism data. This purpose is enabled by the new SVG format, which allows to generate interactive tourist maps dynamically.

The paper is organized as follows: Section 2 illustrates the new possibilities yielding from the integration of tourism data with geographic information systems. The basic concepts of XML and meta data based tourism information integration are presented in Section 3 including an overview about the objectives of the XML-KM project. Section 4 concludes the paper by discussing further improvements and pointing to future work.

2 XML-Based Integration of GIS and Tourism Data

Geographic information systems (GIS) is a rapidly expanding field enabling the development of applications that manage and use geographic information in combination with other media. Among others, this technology offers great opportunities to develop modern tourism applications using maps to present information to the user, thus exploiting the two dimensional capabilities of human vision and present the information in a compact and easy to read way [10].

GIS technology has broadened our view of a map. In contrast to paper-based maps, they are now dynamic representations of geographic data with the ability to visualize data according to the user's demands. The integrated GIS system generates maps based on data returned from databases. As a consequence changes of the GIS data and tourism data are automatically considered and valid after the generation of a new map.

Actual geographic information systems work with two fundamentally different types of geographic models - the raster model and the vector model. With the raster model objects are represented as a matrix of cells in continuous space. A point is one cell, a line is a continuous concatenation of cells and an area is represented as a field of contiguous cells. Today most of existing Web-based information systems use raster graphics for the representation of maps. In this case a satisfying user interaction (e.g., zooming) is not possible. In contrast to raster maps vector based graphics use x, y coordinates to define the position of graphical objects. This allows graphic transformations, like zooming, at the client without losses of quality or makes textual search mechanisms available. In addition a dynamic generation of maps is possible, which enables the presentation of up-to-date information to tourists.

Until now, there have been several initiatives to establish a vector standard [11], e.g., SVF (simple vector format) [12], which was the first attempt for vector representation in the World Wide Web, or Flash [13] which is the most common vector format, but it has never been admitted to be an official standard. Furthermore, Microsoft specified his own proprietary vector format VML (vector markup language) [14], which was an appropriate attempt but limited to Microsoft platforms.

VML was one of the basics to define the more generalized and advanced vector format SVG.

The SVG grammar comprises, generalizes and advances the preceding attempts to build an open standard recommended and developed from the World Wide Web Consortium [7] to describe two-dimensional vector graphics in XML. SVG enables the integration of three types of graphic objects: vector graphic shapes, text and images. Graphical objects can be grouped, styled, transformed and combined with other SVG objects. SVG offers the features of embedded interactivity (vector zoom, move operation, , etc.), animation, embedded fonts, XML, CSS and supports scripting languages with access to the DOM (document object model) to obtain full HTML (hypertext markup language) compatibility.

Even though SVG is an upcoming standard, there already exist several SVG tools like editors, viewers, converters and generators. But until now Web browsers do not have an embedded support for SVG, which is promised for the next generation of Web browser. In the meantime Adobe's SVG plug-in [16] allows to process SVG-based graphics within Web browsers.

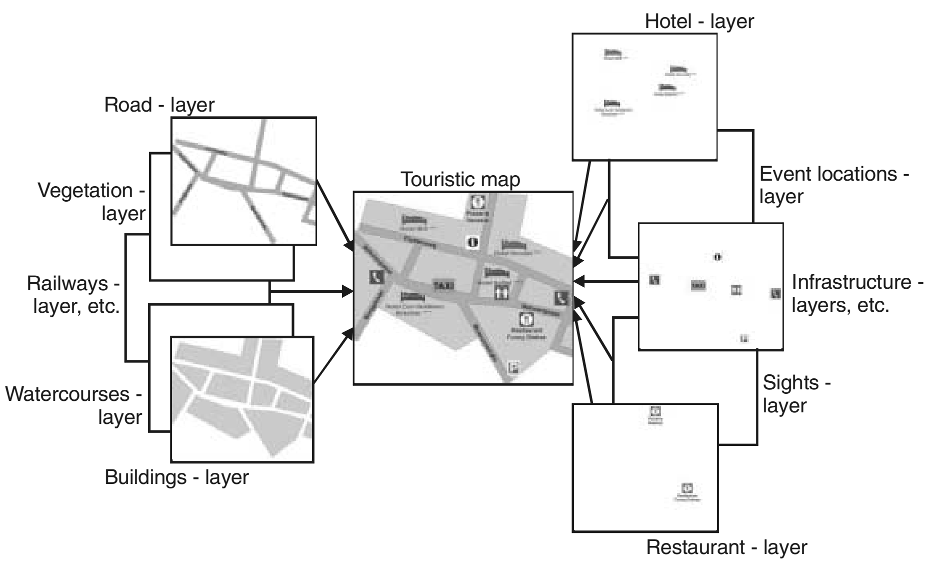

Fig. 1. Integration of themed layers.

The integration of geographical data, tourism information, textual data, images, and links are supported by SVG. Using vector graphics allows to interact, analyze, and to use screen-related functions, such as zooming and panning. It is easy to select features to be displayed, while ignoring unwanted data. Each kind of information within the map is located on a themed layer, which can be anything that contains similar features like roads, buildings, watercourses, hotels or sights. According to their needs users turn these layers on or off. To integrate GIS data and tourism data the traditional layer model (roads, buildings, vegetation, watercourses, railways, etc.) is enriched with additional tourism layers like hotels, restaurants, sights, event locations and further infrastructure layers (Figure 1).

SVG supports themed layers and the generation of user individual maps. In addition it is possible to integrate the following types of information and graphical symbols:

- object symbols representing the type of a touristic object,
- alphanumerical object descriptions (e.g., name and category of a hotel),
- colors, e.g., to visualize the availability of an hotel
- links to a homepage represented by a graphical object (e.g., a hotel's homepage).

In this approach tourism data stored in a tourism database still remains separated from GIS data stored in a GIS database. The XML-based integration of data takes place at the information visualization level. For the visualization of touristic objects it is necessary to add its geographical coordinates (latitude and longitude) which are then stored within the GIS database. The determination of the geographical coordinates of touristic objects is part of the administration expenses within the TIScover Extranet.

There are two different ways to determine object coordinates:

- *Manual positioning*. Authorized tourism information providers (e.g., hotel manager, tourist association) are responsible for the administration of geographical coordinates of those objects which are under their responsibility. The localization is done by marking the position of the object on the map. The corresponding geographic coordinates will be calculated and stored in the GIS database.
- *Automatic positioning*. Based on the postal address of touristic objects the geographic coordinates are calculated automatically [18].

Both approaches have advantages and disadvantages. The automatic approach will be faster, but it is less accurate and error prone because it is a prerequisite that the postal address of the touristic object is complete and that the address data within the GIS system matches to the data of the tourism system to enable accurate mapping. The manual approach is more precise because it is an WYSIWYG (what you see is what you get) approach since the administration as well as the map generation is based on the same map data. The problem of this alternative is a great administration effort.

After the geographic coordinates of touristic objects are determined, the objects can be visualized on touristic maps. Additionally, the geographic information system offers an advanced geographic search feature which allows to combine touristic search attributes like, object type, object name, category of hotels, etc. with geographic criterions like nearness, distance, location (city or province) or objects located inside a marked rectangular map region. Figure 2 describes the workflow of the data integration process.

The user starts a request, which is transmitted to the integrated GIS system. The GIS system submits the query to the spatial database to get the map data and to create a list of objects which are located in the queried area. Each object returned will be completed with tourism data like object name, URL of the object's homepage, category of the hotel, availability of hotel rooms, etc. Afterwards the transformation module converts the characteristics (layer extent, coordinate reference, shape, points, etc.) into SVG representations (polygon, line, path, etc.) and integrates the GIS data with the tourism data, identifies the layers, and defines the representation of the touristic objects (symbols, visualization of layers, object linking etc.). The data of the SVG map is provided in standard XML format and fulfills a DTD, which means that it is well-formed. Finally the touristic map is delivered to the client. On the client only queried layers are turned visible, but the users can turn layers on or off according to their needs.

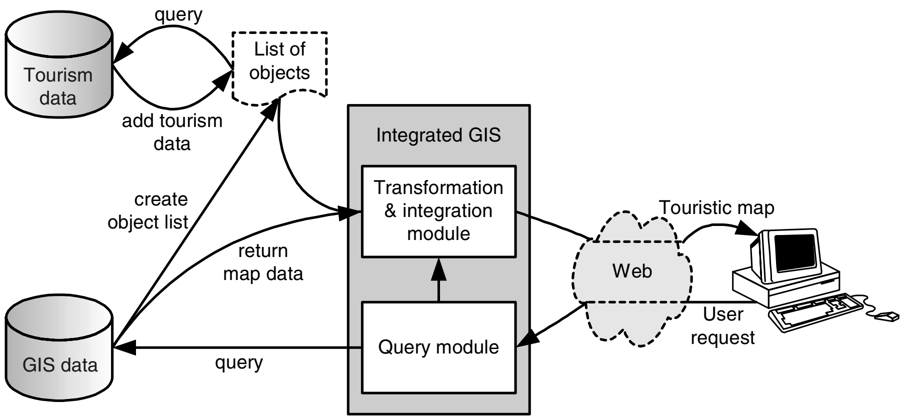

Fig. 2. GIS data integration workflow.

The results of the integration of GIS data and tourism data are touristic maps including the information where touristic objects are, how they can be reached, and which objects are located nearby. If the map representation of a touristic object is linked with the homepage of this object users acquire more detailed information about the object and / or in case of hotels features like online booking are provided.

A first prototype of the SVG based data integration and map representation is already implemented.

3. XML and Meta Data Based Information Integration

To fulfill the tourists request for an extensive data collection it is inevitable to make data from different sources accessible. In a first step towards a comprehensive integration of tourism data TIScover is extended with a flexible data interchange adapter which allows interchange of structured data with other tourism information systems (Figure 3).

Beside the main purpose of a global information integration this approach also focuses on data interchange (e.g., invoice data, order data, personal tourist data) between tourism information providers and other business partners without involving the TIScover system. For these data interchange requirements the existing functions of the TIScover Extranet are not sufficient or not adequate (Figure 4).

Especially the second data interchange scenario requires data interchange facilities which cope with the following demands:

- support of data interchange with a large number of business partners,
- high flexibility,

- lean administration and
- low costs.

In general, there are two contrary strategies to implement data interchange applications:

- Implementation of one data interface per EDI (electronic data interchange) communication partner. In the worst case *n* EDI partners require *n* different interfaces.
- Definition of a *common standard* which covers nearly all possible data interchange scenarios.

In the first case the implementation and maintenance of the whole set of interfaces is very time and cost intensive. In the second case a common standard results in a complex interface specification with a more or less large overhead. The question is: Why put up with the whole overhead if, for example, only a very simple data structure has to be transferred?

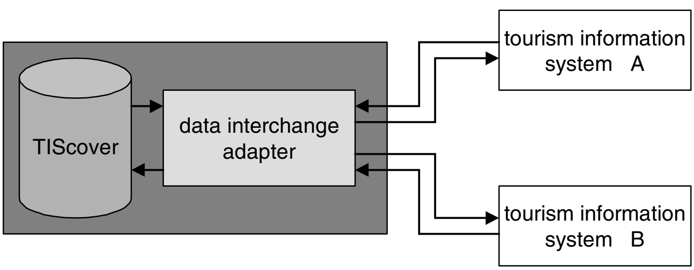

Fig. 3. Global information integration.

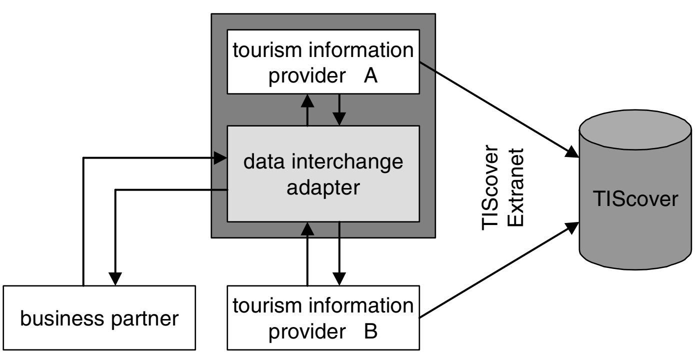

Fig. 4. Information interchange with other business partners.

The approach presented in this paper supports both strategies in order to combine the advantages of each of them. Knowledge about data structures and data formats is separated from the process of generation of destination files and electronic transmission. This separation achieves the purpose to make the data interchange application configurable. If any changes or extensions of the data interchange specification are necessary, it is sufficient to update the knowledge base. The implementation of the data interchange application remains unchanged. The required knowledge base is realized as a meta data structure.

3.1 Architecture of the Data Interchange Adapter

The general architecture of the data interchange adapter consists of three components (Figure 5):

- A *Knowledge base (meta data)* about data structures and data formats, represented by XML DTDs stored within a central database.
- The *mapping knowledge editor* is necessary for the administration of meta data information which is the representation of an individual data interface to an EDI communication partner.
- The *data interchange processor* in the first step receives input data which has to be transformed into the required destination format. In the next step the processor scans the meta data information corresponding to the required destination format. Based on this meta data the destination data file is generated dynamically. Data interchange is possible in both directions to and from a TIS application.

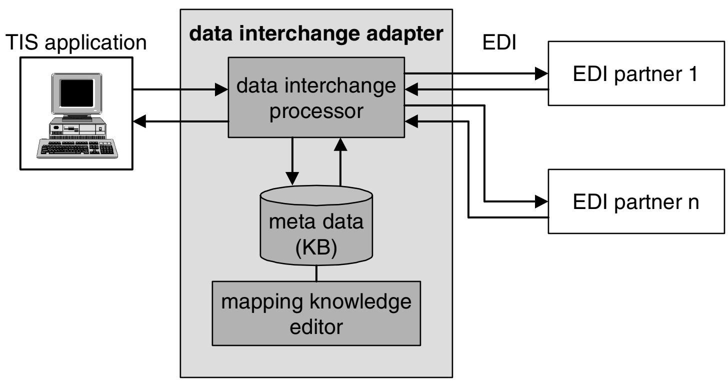

Fig. 5. General architecture of the data interchange adapter.

Beside several other key characteristics XML is a new and flexible concept for the specification of an EDI message. In contrast to HTML (hyperText markup language) which aims at the *presentation* of information XML focuses on *structuring* of information which is also the most important process of an EDI application. Hence, for example, it is possible to model existing EDIFACT message types [17] as XML messages.

Taking these considerations into account the latest implementation of the introduced data interchange adapter is based on XML DTDs (Figure 6). Within the XML based architecture source and destination files are in ASCII- (due to high compatibility) or XML-format. Meta data information for the specification of the transformation of a data source into its destination format is represented by XML DTDs.

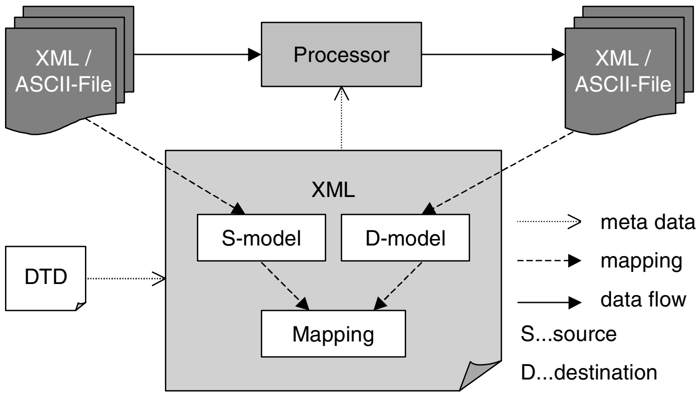

Fig. 6. XML based architecture.

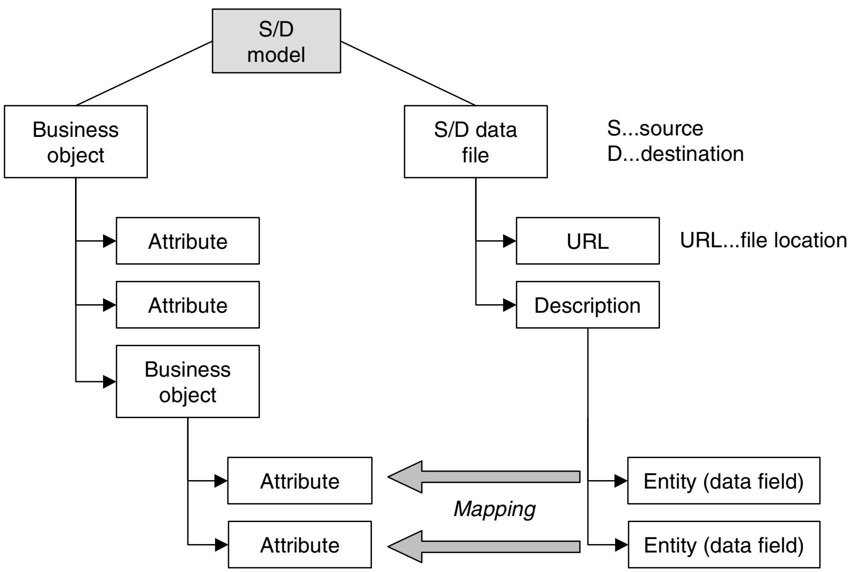

Fig. 7. Business object mapping.

A DTD also describes the mapping between the source model and the destination model each representing the data structures of the corresponding source/destination file. A model itself consists of business objects and references the corresponding source or destination file (Figure 7). Business objects are representations of the data structures within the source or the destination file. A business object is designed as a tree structure and therefore may itself consist of business objects and/or simple attributes (Figure 7). Each attribute of a source model knows its corresponding data field within the source file and each attribute of a destination model references the corresponding destination file, data field and data type.

The maintenance of the meta data information is done by using a mapping knowledge editor which allows to specify or update DTDs for source and destination models as well as the mapping between the represented data structures. Due to the central database system changes of the transformation specifications (meta data) are immediately valid for each interacting software system.

A further key aspect of the XML based architecture is the possibility to specify layout information of an XML document separated from the XML document itself. Hence, it is possible to generate different outputs of one XML document using different layout styles which are defined with XSL (XML stylesheet language) [8]. In this case it does not take great effort to provide different outputs of the same EDI process for different EDI partners.

3.2 The XML-KM Approach

The main advantages of XML based EDI are high flexibility and low implementation costs. In combination with XSL XML is important for structuring information *and* for its presentation.

A prototype of the XML and meta data based data interchange adapter is already finished. Towards a global and comprehensive integration of tourism information the introduced data interchange adapter is a first step, which makes flexible data interchange possible. Currently, XML is confined to syntactic exchange and thus can be regarded only as promising basis of semantically useful shared knowledge management [19], [20].

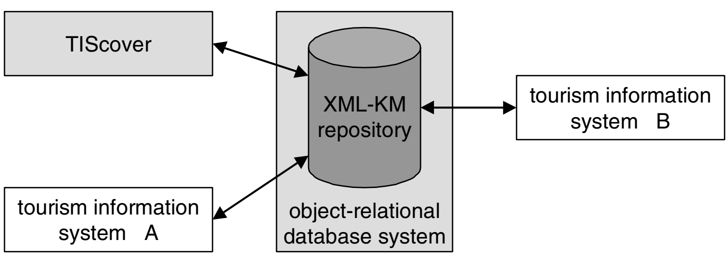

Fig. 8. Integration of tourism data in the XML-KM repository.

The main technical innovative points of the XML-KM project overcome some of these limitations. First, by adopting a semantic approach to the deployment of XML, combined with techniques to extract knowledge and to create conceptual views. Second, by developing concepts for the collection, fusion and dissemination of irregular, evolving information sources as opposed to the bilateral exchange. XML-KM extends XML DTDs with means to relate element-definitions and data types in ontologies using modeling means which will be harmonized with the efforts of the W3C working group on XML schemas [15]. This provides the basic framework to meaningfully collect information from diverse sources, integrate it into unified, domain-dependent schemas and to disseminate it according to the individual needs of diverse user groups or information systems.

Various data sources are integrated in the XML-KM repository (Figure 8), which provides facilities for querying integrated views of remote data and updating the repository. The XML-KM repository is built on top of an object-relational database system (Oracle 8).

The data to be integrated in the XML-KM repository can be structured or unstructured and can be managed by use of different technologies (databases, HTML pages, XML pages and text files).

Beside integration of tourism data in the XML-KM repository and integration of geographical data with tourism data the touristic part of the XML-KM project also focuses

- location based services in WAP (wireless application protocol) and
- dissemination of user obtained tourism newsletters on predefined schedules.

4 Conclusions

Due to the tremendous growth of the Web to search for and to locate interesting information distributed over various Web sites takes a great effort and is time consuming. This is especially true for tourism information systems. To fulfill the tourists request for an extensive data collection it is inevitable to make accumulated data from different sources accessible.

As discussed in Section 1 the following requirements have great impact on the quality of a tourism information system focusing the accessibility to the users:

- integration of geographical data with tourism data and
- integration of distributed data sources.

Actual tourism information systems are still characterized by a lack of integrated GIS data. The approach described in this paper integrates tourism and GIS data by using SVG the XML-based standard for vector and mixed raster-vector graphics for the generation of touristic maps. Dynamically generated vector maps based on SVG offer a wide spectrum of powerful functionalities like high performance zooming and printing as well as a flexible and open representation of interactive touristic maps in combination with the possibility to integrate other XML-related technologies. This kind of touristic maps gives tourist information providers the opportunity to present their tourism information to potential tourists in a clear, fast and powerful way. The flexibility of the presented concept and its implied technologies significantly improve the integration of tourism and GIS data to build dynamic interactive touristic maps.

Beside the integration of GIS data federation of tourism information sources is a primary goal. Thus, flexible and powerful data interchange mechanisms are required. TIScover is extended with a data interchange adapter which allows interchange of structured data with other tourism information systems. In contrast to many existing systems the presented approach separates knowledge about data structures and data formats from the process of generation of destination files and electronic transmission. This knowledge is transformed into a meta data structure represented by XML DTDs. The main advantage of this concept is that if changes of the data interchange specification to other tourism information systems are necessary, it is sufficient to update the corresponding meta data information within the XML DTDs. The implementation of the data interchange adapter remains unchanged.

Further work will be done primarily in the course of the IST project XML-KM. The goal of XML-KM which is strongly based on XML is to improve the data integration process in order to be able to collect and disseminate *knowledge* instead of just data. Through a rule-based XML-wrapper, information from corporate databases, HTML pages and office applications will be collected in the XML-KM repository. Through XML-based query tools, users will be able to receive personalized information in an appropriate format on various devices including computers, mobile phones based on WAP services and faxes.

References

1. Pröll, B., Retschitzegger, W., Wagner, R., Ebner, A.: Beyond Traditional Tourism Information Systems – TIScover. Journal of Information Technology in Tourism (ITT), Vol. 1, Inaugural Volume, Cognizant Corp., USA (1998)
2. Pröll, B., Retschitzegger, W., Wagner, R: TIScover – A Tourism Information System Based on Extranet and Intranet Technology. Proc. of the 4th Americas Conference on Information Systems (AIS 1998), Baltimore, Maryland (1998)
3. TIS Innsbruck, FAW Hagenberg: Homepage of TIScover, http://www.tiscover.com (2000)
4. Ebner, A., Haller, M., Plankensteiner, K., Pröll, B., Pühretmair, F., Starzacher, P., Tjoa, A. M.: TIS Business Workflow Specification (D5-1.0), XML-KM (IST 12030) (June 2000)
5. Bezares, J-L., Gardarin, G., Huck, G., Laude, H., Munoz, J-M., Pühretmair, F.: General Architecture Specification (D11-1.4), XML-KM (IST 12030) (April 2000)
6. The European Comission: Homepage of the IST-1999-12030 project: XML-based Mediator for Knowledge Extraction and Brokering, http://www.cordis.lu/ist/projects/99-12030.htm (2000)
7. W3C - The World Wide Web Consortium: Scalable Vector Graphics (SVG) 1.0 Specification, Candidate Recommendation, http://www.w3c.org/TR/SVG/ (2000)
8. Behme, H., Mintert, S.: XML in der Praxis. Professionelles Web-Publishing mit der Extensiblen Markup Language. Addison-Wesley Publishing, BRD (1998)
9. Bradley, N.: The XML companion. Addison-Wesley Publishing, Great Britain (1998)
10. Christodoulakis, S., Anastasiadis, M., Margazas, T., Moumoutzis, N., Kontogiannis, P., Terezakis, G., Tsinaraki, C.: A Modular Approach to Support GIS functionality in Tourism Applications, Proc. of the Int. Conf. on Information and Communication Technologies in Tourism (ENTER'98), pp. 63-72, Springer Verlag, Istanbul (1998)
11. Neumann, A., Winter, A.: Vector-based Web Cartography: Enabler SVG, Carto.net - Cartographers on the net, http://www.carto.net/papers/svg/index_e.html (2000)
12. Softsource: SVF (Simple Vector Format), http://www.softsource.com/svf/ (1996)
13. Macromedia: Macromedia Flash - Create animated vector-based Web sites, http://www.macromedia.com/software/flash/ (2000)
14. W3C - The World Wide Web Consortium: Vector Markup Language (VML), http://www.w3.org/TR/NOTE-VML.html (1998)
15. The World Wide Web Consortium (W3C). http://www.w3.org/XML/ (2000)
16. Adobe Systems Incorporated: Adobe SVG Viewer 2.0 (beta), http://www.adobe.com/svg/viewer/install/main.html (2000)
17. Schmoll, Thomas: Handelsverkehr elektronisch, weltweit: Nachrichtenaustausch mit EDI/EDIFACT. Markt & Technik Verlag, München (1994)
18. O'Neill, W., Harper, E.: Linear Location Translation within GIS. Proceedings of the ESRI User Conference (1997)

19. Huck, G., Fankhauser, P., Aberer, K., Neuhold, E.: Jedi: Extracting and Synthesizing Information from the Web. Proceedings of the COOPIS 1998 Conference, New York, IEEE Computer Society Press (1998)
20. Petrou, C., Hadjiefthymiades, S., Martakos, D.: An XML-based, 3-tier Scheme for Integrating Heterogeneous Information Sources to the WWW. Proceedings of the 10th International Workshop on Database and Expert Systems Applications (DEXA '99), IEEE Computer Society (1999)