

A GEOGRAPHICAL METADATA MANAGEMENT SYSTEM FOR SITE PLANNING

SYSTEM ZARZĄDZANIA METADANYMI GEOGRAFICZNYMI DLA MIEJSCOWEGO PLANOWANIA PRZESTRZENNEGO

Artur Krawczyk¹, Tobias Limbach², Gabriela Surowiec³

¹Department of Mining Surveying & Environmental Engineering, University
of Science & Technology, Poland

²Department of Computer Science & Information Engineering, University of Konstanz, Germany

³Department of Geodesy, Agriculture University of Cracow, Poland

Keywords: metadata, information system, geographical information

Słowa kluczowe: metadane, system informacyjny, informacja geograficzna

Introduction

Metadata about geographical data is one of the more important elements of Spatial Data Infrastructure (Gaździcki, 2003). Many various users other than the geographical data producer use geographical data. It is often produced on demand by one organization and used by another one. Proper documentation such as metadata should provide other organizations with a its better understanding, and enable them to use it effectively (Cichocinski, 1997). As geographical data producers are creating more and more data they try to find viable solutions for selling their data. One of the most popular solutions to do it is creating a Web service which presents textual information about geographical data offer. Some producers put samples of geographical data into their WebGIS services. It helps ordinary users to find data by using traditional web browsers like Google, Vivisimo, etc. But, this way the user receives a lot of various, non-standardized information about the data. Proper, standardized documentation would provide the user with much better knowledge about data and would allow him better managing data production, storage, updating, and reuse of geographical data. Site planning process is an example of the situation when users need of a lot of geographical data. A basic problem in the site planning process is search for current and essential data and analysing it. In particular, spatial data is needed for various analyses concerning the planning process. A possible solution to such a problem was created in the EU funded project – INVISIP: **I**nformation **V**isualization in **S**ite **P**lanning, IST-2000-29640. INVISIP provides a technical platform, which facilitates access to metadata information and data handling. It contains tools, which support metadata gathering, and different visual tools for data retrieval and data analyses.

At the beginning of the project we came across the problem of naming metadata. For the site planning process a lot of quite different documents are necessary. Besides the spatial data there are many non-spatial documents, e.g. laws, guidelines and reports, and both types of them are significant for the user. Therefore, metadata used in this project was divided into two different types. On one hand, there is metadata for geographical data, the so called metadata for geodata (G-MD) and on the other hand, there is metadata for non-geographical data, the so called metadata for documents (D-MD). For both types of metadata different software tools were created. During the project we found that the term “geo-metadata” for describing metadata for geodata is quite useful. In this paper we present problems concerning creating and using geo-metadata.

Geo-metadata creation

The geo-metadata creation process is effective when it takes place in the same organization which produces the geodata. Also, geo-metadata updating process is easier when it is done at the same time and by the same organization. On the other hand, retrieval process for any information is much easier when it is done in one centralized database system, rather than in different places in Internet. Using warehouse database architecture solved those two inconsistent requirements. A small database system for creating geo-metadata may be installed in different organizations at any time. Organizations may input geo-metadata and manage them in their local standardised databases. Geo-metadata may be exported from the local database to a central metadatabase warehouse. On the basis of the metadatabase we may introduce interfaces and tools for the user to make use of geo-metadata. For this task the Metadata InGeo EntryTool was developed. This application was written in VisualBasic by GIS-tec company and handles metadata based on ISO 19115.

This tool was created for the input and management of geo-metadata by the geodata owner (Krawczyk 2003). Using an intuitive GUI, the user may select metadata elements which he would like to input. The application can be configured so as to use it with five different languages. The integrated metadata concerning maps are stored in a local Access-Database and can be exported to the XML format, which is a base format for metadata exchange. This kind of files can be easily imported to the warehouse metadatabase and then integrated into GIS Web based systems. The Oracle Warehouse Database was created and tools for importing XML files into it were also tested. On the basis of warehouse technique a database with geo-metadata was established.

Geo-metadata browsing

The next stage is searching the metadatabase for eligible information. To realize this task a tool called Metadata Browser was developed. It is a visual seeking system which allows to select groups of metadata records by given keywords. The data found is presented in a result table with a variable level of detail. Based on the first selection, the user has the possibility to

make next, more detailed selection by using additional visualization techniques such as scatterplot, focus and context or drill-down functionality. The user can sort or filter metadata subset according to different attributes and different additional keywords. Records of metadata which meet the user's criteria can be stored in the project and used in order to get original geodata.

The metadata browser facilitates user interaction through a query preview, visualizes the search results in an intuitive, transparent way and is flexible enough to adapt to special user needs at different stages of the site planning process and at different stages of information retrieval as well. Based on redesign of the business-intelligence system INSYDER (Reiterer H., et al 2000) (Internet Systeme de Recherche, IST Project No. 29232.), we developed a geo-metadata browser with different strategies to avoid the cognitively demanding traditional representation of geo-metadata result sets, which usually follows metadata standard such as FGDC's or ISO 19115 and results in a long list of document attributes. We assumed that users would experience difficulties or at least a slowdown in productivity, if they had to work with a traditional representation of the result set, i.e. with the list based representation. Though these lists are very popular in search engines on the web, a higher level of abstraction and visualization for representing a complex result set from the geodata or geo-metadata domain can be expected. Our questionnaires accompanying various usability tests strengthened this opinion: users just do not want to scroll through endless result lists with limited possibilities to filter or sort them, as they are forced to do when using most search engines. To quote Chen and Dumais: *...users are able to find documents more efficiently when search results are organized into topical categories than when they are presented with a standard ranked list...* (Chen, Dumais 2000). One way to support the users' thinking in categories is to present meaningful visualizations. The geo-metadata browser aims at supporting the users need to work with topical categories.

The root concept of the geo-metadata browser is based on empirical evaluation of the business-intelligence system INSYDER (Mann 2002),(Reiterer, et al 2000) mentioned before. Formerly separated visualizations such as the Result Table, bar graphs, and SegmentView were put into one visualization called "SuperTable", combined with the Scatterplot. In a single window, users were offered different brushing and linking techniques of two components SuperTable and Scatterplot. The Scatterplot gives a quick overview and defines own views with zooming, selecting and filtering to reduce the amount of hits while the SuperTable follows the concept of a distortion-based table with BarGraph and SegmentView for analyzing documents in high detail.

Here the SuperTable is combined with the Scatterplot and gives an overview of the result set only with multicolored bars (Fig.1). Each row represents the metadata of a search on the INVISIP database. The columns describe all characteristics concerning this document, e.g. title, country, date, size and so on. The main feature of the initial view is that no text at all is seen in the table, but the user is able to get a graphical overview of the relevance of the search results. This relevance is displayed in the metadata browser in four changeable levels of detail, where every level has a different focus on aspects of the documents. The first level is only to give an overview; the following level can be used to explore various aspects of the metadata set. The metadata displayed may be configured by each user according to his personal likings or to the context of the task the user has to work with. For example, one level can be used to display only survey information of the geodata, while next level could be used to display pricing and ordering information.

Fig. 2 shows the last and most detailed level of the metadata browser. Only the most valuable textual contents of a document should be displayed, such as the title and abstract. Both are

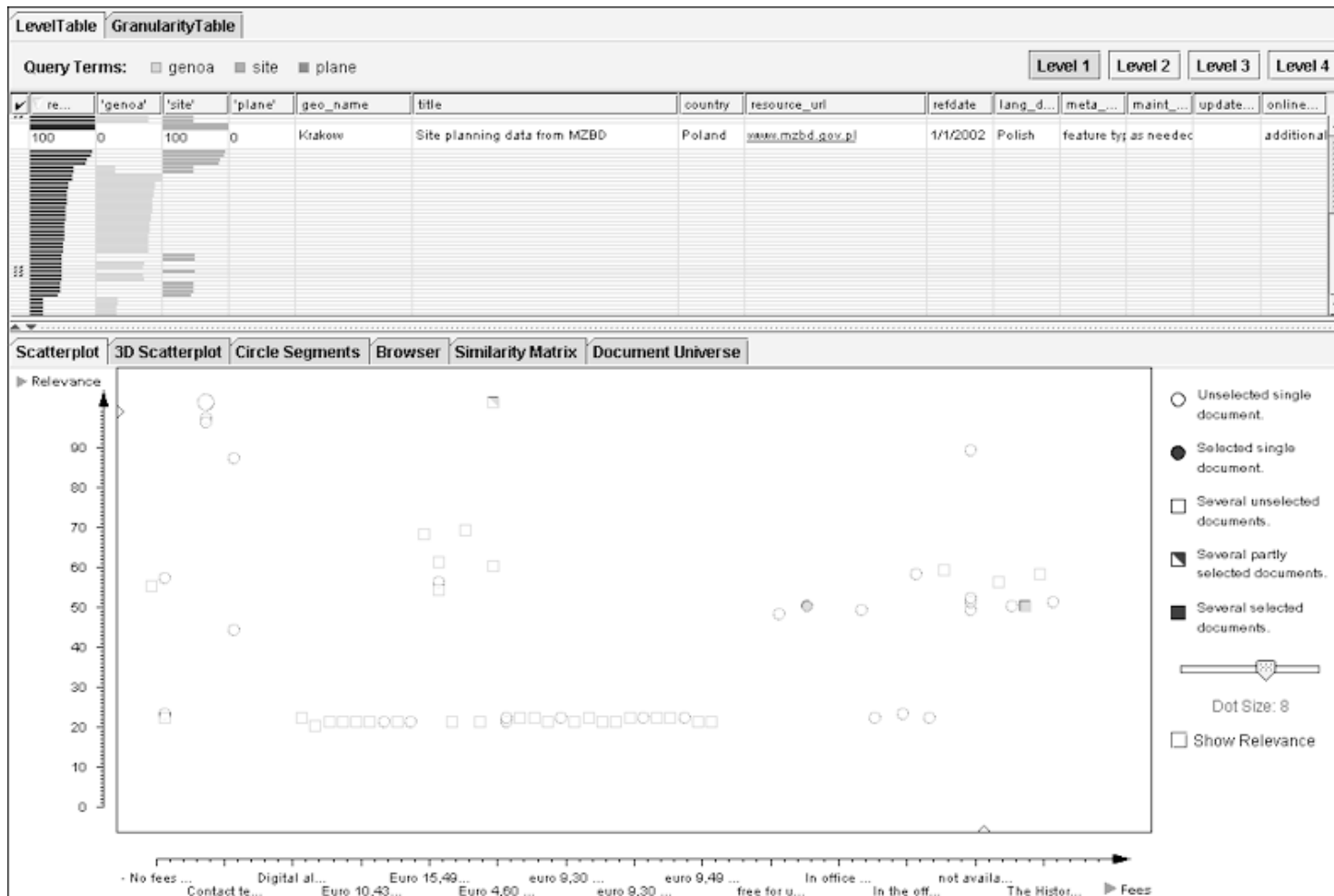


Fig. 1. Level 1 of the SuperTable with Scatterplot.

LevelTable GranularityTable

Query Terms: genoa site map

Level 1 Level 2 Level 3 Level 4

| <input checked="" type="checkbox"/> | Detailed Relevance Curve | geo_name | keywords | address | format... | order_instruc | turnar... |
|-------------------------------------|--------------------------|--------------|---|-----------------|-----------|---------------------------------------|-----------|
| <input checked="" type="checkbox"/> | | Genoa center | System of horizontal and vertical signs, traffic lights and the flow regulation | Via Garibaldi 9 | nff | request to the administration offices | 30 days |
| | | Genoa center | System of horizontal and vertical signs, traffic lights and the flow regulation | Via Garibaldi 9 | nff | request to the administration offices | 30 days |

Scatterplot 3D Scatterplot Circle Segments Browser Similarity Matrix Document Universe

Title: Traffic Urban Plain
Abstract: The main topic of the Traffic Urban Plain is the city traffic decrease within relocation of the crossing traffic by the city centre to the away or to sea bypass. The important objective is environmental retraining to give back vividly the city centre sites. Besides the project foresees the optimization of the public transport to obtain an integrated urban structure. The art. N° 36 of the Dlgs. N° 285, 30/04/92, forces all the municipality, that they have more 30.000 citizens, to adopt Traffic Urban Plain
Access_constraints: Copyright
Address: Via Garibaldi 9
Browse_graphic: PUT.jpg
City: Genoa
Country: Italy
Email: vmcontursi@comune.genova.it
Fax: ++39 010 5577818
Fea_cat_cit:
Fees: euro 8 for visualization only, euro 25 for hardcopy map; euro 26 for digital map
Format_name: nff
Format_version:

Fig 2. Segment View and, in the lower part, the browser view with highlighted segments

shown in the visualization called SegmentView, which uses stacked columns. To visualize this, the textual information is artificially divided into segments, and each segment is assigned a certain value corresponding to its relevance in the query. Segments, which have a high semantic significance for the query, are ranked higher and displayed with a longer column; segments with lower relevance get a shorter column. The stacked columns are colored according to the colored search terms in the former levels. So it is easier, for example, to discover segments where all search terms can be found, because the column is stacked with multiple colors. It is easy to spot segments which include only one or two search terms and discern them from parts including all terms. Per default, instead of the Scatterplot a preview browser is offered showing the selected document or per default the first document of the SuperTable. This offers the users the greatest level of detail showing them the textual information of the metadata set.

Through iterative redesigns in the user centered design process, it became more and more clear that interaction possibilities should be added. The first designs focused on the level concept and partially left out the manipulation through the Scatterplot, which was in the subsequent design enhanced with additional visual filtering mechanisms. A so called magic lens (Fishkin, Stone, 1995) was added to support the users' need to visually filter according to specified filter criteria. The magic lens is basically a semi-transparent rectangle in the Scatterplot, which lets only the relevant data dots to be visible through it. If the user is faced with many data dots, the magic lens with a specified filter criteria significantly reduces the amount of data that the user has to browse through. Multiple magic lenses can be added and combined with boolean operators.

The magic lenses in Fig 3 filter out any datasets, which do not match their specified criteria. Only relevant dots can be seen through the lens and are additionally highlighted with the lenses color in the SuperTable. Multiple lenses can be connected by AND or OR expression and so can visually filter numerous criteria. Fig. 3 for example demonstrates the red lens, which only shows documents where original geodata are stored in nff format, the blue lens shows only documents with country Italy (Limbach, et al, 2003).

A possible usage scenario would start in the first level of the SuperTable (see Fig. 1). The user has provided some query terms and now sees an aggregated result set of the query. Through the visual coding of bars, it should be immediately obvious, which query terms provided the best results. The user can now apply different sorting or marking functions to the set (for example sorting by category, or filtering by a magic lens as in Fig. 3), until he is sure to go to the next level of detail, where more information is provided, and less documents are shown. This process can go on until the user is satisfied and sees the documents, which are now left, according to his criteria, in the last level of detail (Fig. 2). From there (or from any other point during the query, as required) the results can be stored into the INVISIP session.

Two qualitative usability studies with mockups in different stages and a quantitative usability test with the working prototype during the user centered design process assured a high level of usability for INVISIP.

The usability tests were all held at the University of Konstanz, Germany. During two qualitative tests, 16 participants (mostly experts in site planning) performed various tasks derived from typical site planning scenarios. The quantitative usability test was done with 32 participants. Especially the qualitative user tests with participants from the geodata domain from the very beginning provided a lot of valuable feedback.

The comparative usability study focused on the Metadata Browser only. The test between the table based browser and a list based result set with 32 participants showed that with three

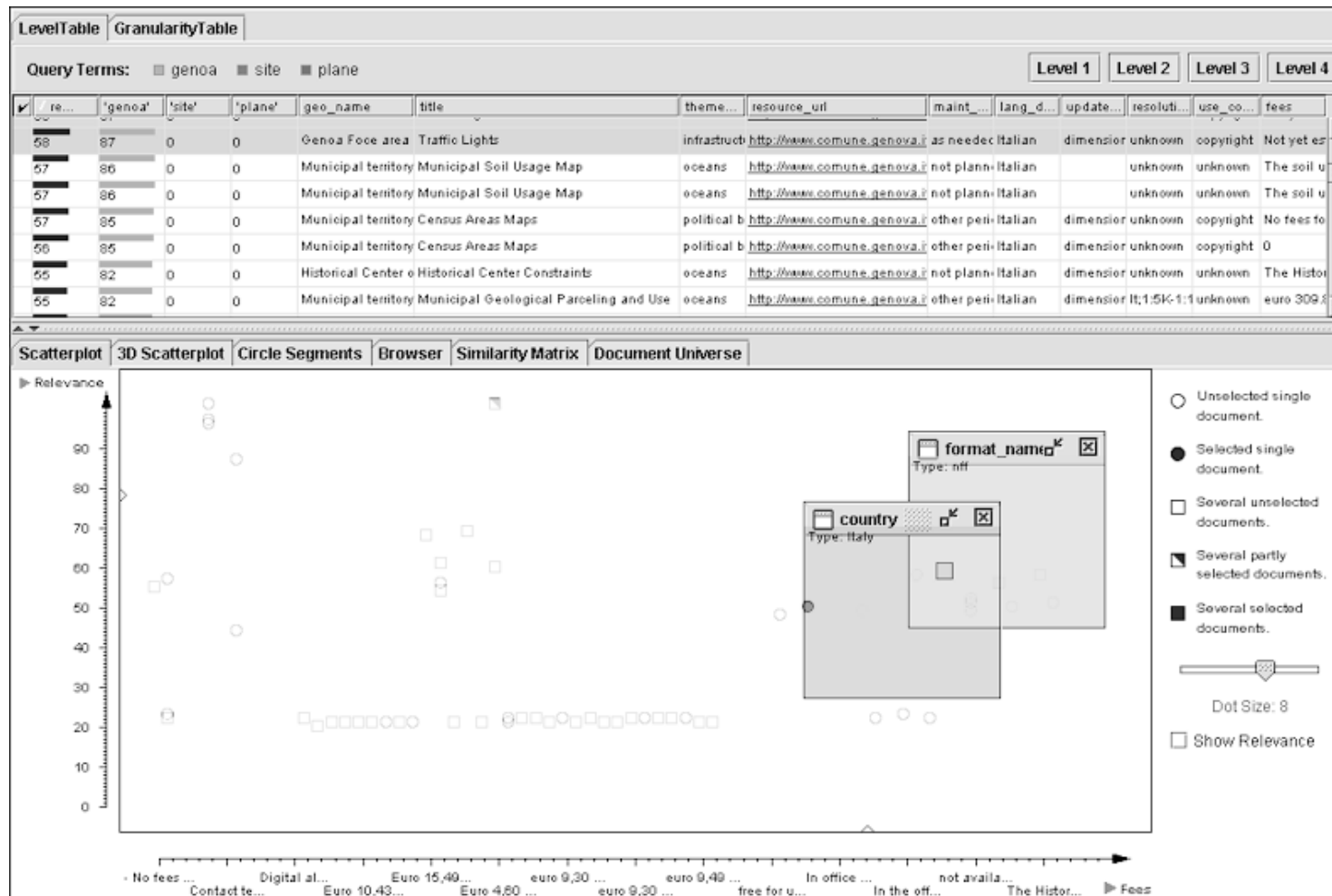


Fig. 3. Two Magic Lenses visually filter out data sets in the Scatterplot. This operation is linked to the SuperTable (Level 2) as well

categories of queries the geo-metadata browser using a table based representation of the results is significantly more efficient with two of them (Gerken, 2004). This supports the quote from Chen and Dumais mentioned before, that users think and work in categories and visualizations such as the geo-metadata browser in INVISIP can contribute to usable and user centered software in complex decision supporting systems.

Geo-metadata as context information of WebGIS

MetaData Browser can work through the Internet in Web Browser environment (eg. Internet Explorer, Opera etc.). MDB as a result of searching process gives the user a metadata record in textual form. Usually, geo-metadata in data provider contact category contains only information such as address, phone number or e-mail of the data provider. It means that user has to use other tools than Web Browser to contact with data provider. Of course, some geodata providers have their web pages with information about their offer. Link to their web services can also be added to the geo-metadata fields in data provider category. It is a useful feature but it means that the user is redirected from one record of geo-metadata (very detailed information) to data provider start webpage where he needs to start browsing the web again.

Some data providers put samples of geographic data into their WebGIS services. In such a case it is strongly recommended to put into geo-metadata also the information about http-link to the map or layer presented on the webpage. In this case, based on geo-metadata, the user can access corresponding map or sample available in the Web based GIS Viewer. This tool gives the user an additional possibility to preview original geodata that he is looking for. This feature gives the user a possibility "to see" the data before buying them.

In this case, the user can read geo-metadata found in Metadata Browser and, afterwards, he can open layer in the Web based GIS-Viewer. Of course, the user may come back to the Metadata Browser and browse next geo-metadata records. But he can also start browsing in WebGIS-Viewer other layers which belong to particular geodata provider. In such a situation, the user loses possibility to read standard geo-metadata information about other maps (layers) on data providers WebPage. Usually, data providers do not put any meta information about the layers presented. This situation causes problems with incomplete set of information about other geodata presented on the Web.

Publishing XML files with geo-metadata information as linked HTML web pages solved this problem. These XML files are the same files which were send by geodata provider to Metadatabase. To publish XML files with geo-metadata we need to make 5 steps:

1. Create an XSLT file (eg. Invisip.xslt) which allow us to present XML file as a nicely formatted HTML file.
2. To each XML file we need insert new one line of code with reference to Invisip.xslt: `<?xml-stylesheet type="text/xsl" href="Invisip.xslt"?>`
3. Copy XML files with one XSLT file into WebServer folder publish place (webpace – place where external http link to file can be established)
4. For each layer of table we need to add one column – eg. with name „Metadata” and fill it with proper HTML link to geo-metadata XML file
5. Based on hotlink feature in GIS-Viewer the user can load HTML formatted file with textual geo-metadata

Next important feature regarding GIS Viewer is the possibility to retrieve information from the layer of WebGIS. Usually, a tool called Info realizes it. With this tool, the user can indicate an object on the map and retrieve textual information about the object from geodatabase. In our case, column “Metadata” with URL link may contain object attributes . For making this URL address usable we implemented the function called HotLink. This function allows the user pointing and clicking on the URL link – in this case it opens default web browser with the webpage related to the map with geo-metadata information (Fig. 4).

Based on Info tool and hyper link the user can easily search ready maps (layers) published on WebGIS and retrieve metadata information about each layer.

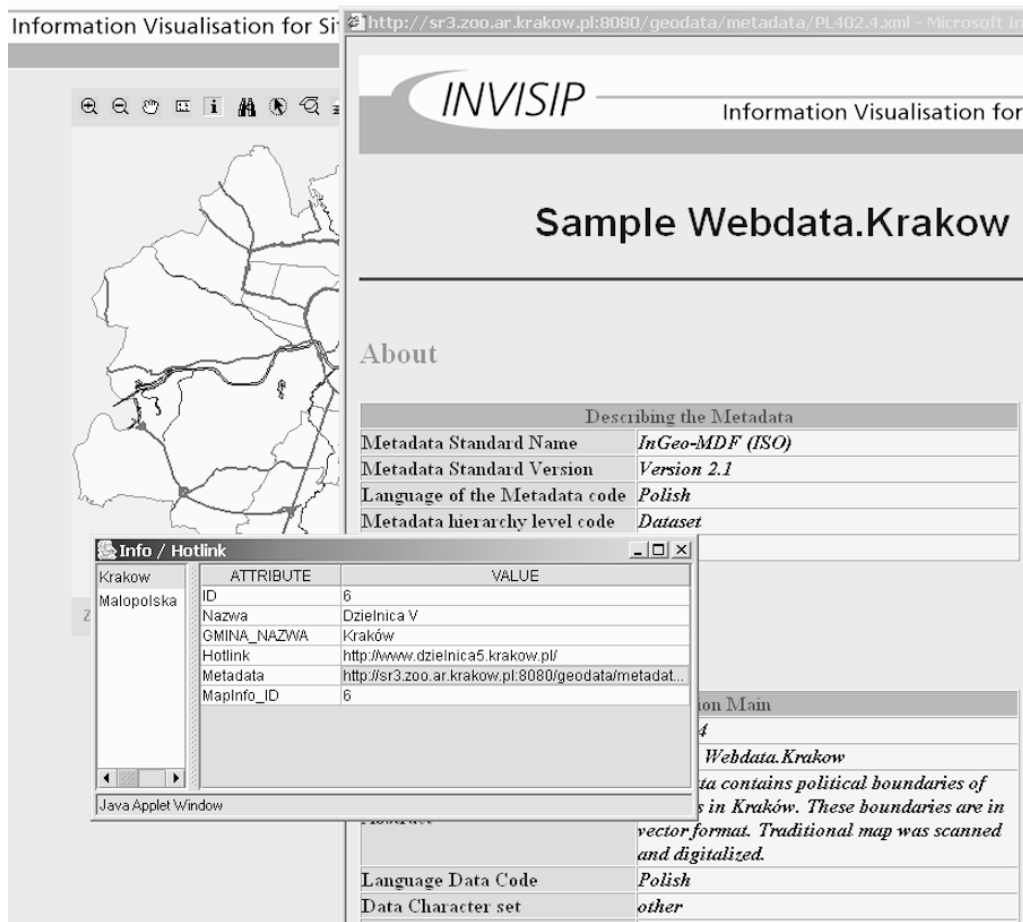


Fig. 4. Geo-metadata as formatted XML file in WebGIS

Summary

In INVISIP system two different ways of using metadata were created. The first method is based on metadatabase server with large sets of metadata, and in the second one, geo-metadata is accessible through the WebBased GIS services. In such a situation, creating metadata by geodata provider becomes much more effective and he has more reasons to take effort to produce metadata in proper standard.

INVISIP was mostly focused on investing in a metadata-based searchable repository in order to make data access and reuse as easy as possible. The usability tests showed that the concept of INVISIP, i.e. using standardized geo-metadata to search for geodata, fits well into typical site planning scenarios. Thus, when using tools such as the Metadata Browser, the users can work more effectively than with comparable list based tools such as search engines.

References

- Chen H., Dumais S., 2000: *Bringing order to the Web: Automatically Categorizing Search Results* Proceedings of the CHI: Conference on Human Factors in Computing Systems, pages 145-152.
- Cichociński P., 1997: *Metadane – sposób na dokumentowanie cyfrowych danych przestrzennych* Polskie Towarzystwo Informatyki. VII Konferencja Naukowo-Techniczna Systemy Informatyki Przestrzennej. Warszawa, 4-5 czerwca 1997 r.
- Gaździcki J., 2003: *Kompendium infrastruktur danych przestrzennych* Magazyn Geoinformacyjny Geodeta, Warszawa nr 3 pages 29-32
- Gerken J., 2004: *LevelTable vs. ListView: ein quantitativer Usability Vergleich*. Bachelor of Information Engineering Final Thesis, University of Konstanz.
- Kenneth P. Fishkin and Maureen C. Stone, 1995: *Enhanced dynamic queries via movable filters*. In CHI, pages 415-420.
- Krawczyk A., 2003: *Metodyka gromadzenia geometadanych w projekcie INVISIP*, Magazyn Geoinformacyjny Geodeta, Warszawa nr 7 pages. 35-36, 38
- Limbach T., Reiterer H., Müller F., Klein P., 2003: *Introductory Video to VisMeB (Video + Short Paper)*. In: Rauterberg M., Menozzi, M., Wesson, J. (eds.) Human Computer Interaction INTERACT '03, Zürich, Switzerland, pages 993-996 plus Video CD.
- Mann T., 2002: *Visualization of Search Results from the World Wide Web*. PhD thesis, University of Konstanz.
- Reiterer H., Mussler G., Mann T., Handschuh S., 2000: *Insyder – an information assistant for business intelligence*. In Proceedings of the 23 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 112-119. ACM press.

STRESZCZENIE

Metadane o danych geograficznych stanowią jeden z istotniejszych elementów infrastruktury danych przestrzennych. Wynika to z faktu, że dane geograficzne są wykorzystywane przez wielu innych użytkowników niż dostawcy danych przestrzennych. Wzrost ilości dostępnych danych powoduje konieczność usprawnienia dostępu do danych. Jednym z bardziej popularnych rozwiązań jest umieszczanie informacji tekstowych o danych na stronach www. Część dostawców danych dodatkowo publikuje je w postaci internetowego serwisu GIS – WebGIS. Dane udostępniane są tam w formie fragmentów warstw lub całych warstw tematycznych o ograniczonym dostępie i jakości. Sytuacja ta nie ułatwia wyszukiwania informacji. Jako przykład można podać planowanie przestrzenne. Podstawowym problemem podczas realizacji zadań związanych z planowaniem przestrzennym jest wyszukanie aktualnych i przydatnych w procesie planowania danych. Możliwe

rozwiązanie tego problemu zostało zaproponowane przez zespół realizujący Projekt Europejski – *INVISIP (Information Visualization in Site Planning, IST-2000-29640 – Wizualizacja informacji w procesie planowania miejscowego)*. System *INVISIP* został zaprojektowany dla wspomagania użytkowników na każdym etapie miejscowego planowania przestrzennego. Proces wspomagania decyzji bazuje na systemie informacji o metadanych (*MIS – Metadata Information System*) zawierającym przede wszystkim metadane o danych przestrzennych (*geo-metadane*) a realizowany jest w oparciu o zastosowanie zaawansowanych technik wizualizacji informacji. *INVISIP* dostarcza technicznej platformy, która ułatwia dostęp do metadanych oraz zapewnia przechowywanie danych użytkownika. System zawiera narzędzia wspomagające zbieranie metadanych oraz grupy wielu różnych narzędzi przeznaczonych do tworzenia, przeglądania i analizowania zasobu metadanych. Niniejszy referat koncentruje się na opisie kolejnych etapów tworzenia i użytkowania *geo-metadanych*. W ramach realizacji projektu opracowane zostało narzędzie *InGeo EntryTool*, które przeznaczone jest do tworzenia *geo-metadanych*. Jest to aplikacja napisana w *Visual Basicu*, która dane przechowuje w pliku bazy danych aplikacji *MS Access*. Sam format danych gromadzonych w bazie jest wzorowany na standardzie *ISO 19115*. Program ten może być instalowany u dostawców informacji przestrzennych jako lokalna baza danych. Po wprowadzeniu do bazy kompletu informacji dotyczących *geo-metadanej* można ją wyeksportować. Eksport danych dokonywany jest do pliku *XML*. Dodatkowo tworzony jest plik *DTD*, który zawiera definicje formatu danych. Format *XML* jest bazowym formatem wymiany danych wykorzystywanym w projekcie. Ten format danych może być łatwo zaimportowany do systemu bazy danych systemu *MIS* jak również może zostać łatwo wykorzystany do publikacji informacji w Internecie. Dzięki temu dostawcy danych mogą szybko i w prosty sposób przygotować *geo-metadane* i następnie wysłać je do integratora metadanych – np. systemu *INVISIP*.

Podstawowym zadaniem programu *InGeo EntryTool* jest formatowanie kompletnego zestawu informacji o *geo-metadanej* i poprzez export przygotowywanie danych do wysyłki. Następnym etapem jest import plików *XML* do hurtowni metadanych (*Oracle Warehouse Database*). Po dokonaniu importu danych można przystąpić do wykorzystania narzędzi przeszukujących zbior *geo-metadanych*.

Podstawowym narzędziem służącym do przeszukiwania jest przeglądarka metadanych (*MDB MetaData Browser*). Jest to narzędzie wizualnego wspomagania przeszukiwania danych, które pozwala na selekcję z całego zasobu wybranych grup *geo-metadanych* poprzez użycie słów kluczowych (*preselekcja*). Grupa *geo-metadanych* jest następnie wizualizowana w postaci tabeli posiadającej możliwości wyświetlania *geo-metadanych* na wielu poziomach szczegółowości. Bazując na wybranej grupie *geo-metadanych* można dokonywać następujących analiz mających na celu dokładniejszą selekcję danych poprzez użycie dodatkowych narzędzi wizualizujących: wykresów 2D i 3D, przesłon oraz innych funkcji. Użytkownik może filtrować grupę *geo-metadanych* poprzez zadanie nowych słów kluczowych. Wybrane *geo-metadane* przez użytkownika mogą następnie zostać zachowane w systemie *INVISIP*. Na podstawie wybranych rekordów użytkownik dokonuje decyzji pozyskania danych oryginalnych. Ma on wtedy do dyspozycji typowe informacje adresowe. Tymczasem wykorzystanie linku *http* do danych prezentowanych przez np. *GIS-Viewer* na stronach *www* użytkownik może wykorzystać do uzyskania dostępu do danych oryginalnych lub prezentacyjnych. Cecha ta wzmacnia pewność podejmowania decyzji o pozyskaniu danych przez użytkownika.

W przypadku wykorzystania przez użytkownika konkretnego *GIS-Viewera* dostawcy danych możliwe jest zaprezentowanie przez tego dostawcę kolejnych warstw informacyjnych. Niestety warstwy te obecnie nie posiadają żadnych informacji tekstowych przypominających zestaw metainformacji objętych standardem *ISO 19115*. A przecież dostawca przygotował i wysłał już metadane o tych warstwach do systemu *MIS*. W takim przypadku w ramach projektu *INVISIP* przygotowano proste rozwiązanie, które pozwala na efektywne wykorzystanie plików *XML* z metadanymi. Rozwiązanie to polega na wykonaniu plików *XSLT* formatujących pliki *XML* do wyświetlania w przeglądarce

internetowej. Następnie pliki z geo-metadanymi (XML) są umieszczane w udostępnionych katalogach serwisu WebGIS serwera. Do tego katalogu należy wkopiować pliki DTD oraz XSLT. Od tego momentu do każdego pliku XML można określić link http. Następnie w geobazie danych prezentowanych przez GIS-Viewera do każdej warstwy należy dopisać nową kolumnę – np. o nazwie „metadane” i do niej wkopiować adres URL do właściwego dla danej warstwy pliku z metadanymi. W ten sposób uzyskuje się możliwości przeglądania standaryzowanych geo-metadanych z poziomu przeglądarki WebGIS.

Podkreślić należy uniwersalność i jakość opracowanego systemu zarządzania metadanymi. Narzędzie Metadata Browser zostało opracowane w oparciu o szereg testów przeprowadzonych zarówno na etapie projektowania, prototypu jak i na etapie produktu finalnego. Opracowany system zarządzania metadanymi umożliwia podwójne wykorzystanie metadanych. Podstawowym kierunkiem wykorzystania metadanych jest importowanie ich do Metadatabase Warehouse, gdzie następnie mogą być wyszukiwane spośród dużej liczby innych metadanych. Drugą możliwością jest wykorzystanie geo-metadanych do publikowania informacji przez dostawcę danych na własnych stronach WebGIS.

Artur Krawczyk
artkraw@uci.agh.edu.pl

Tobias Limbach
tobias.limbach@operamail.com

Gabriela Surowiec
rmkoster@cyf-kr.edu.pl