

METATÉR

Implementation of a nation-wide metadata server

Gábor Uhrin

Scriptum Inc.
6771 Szeged, Mályva u. 34.
e-mail: scriptum@mail.matav.hu
<http://www.scriptum.hu>

Abstract. The rational management of the continually broadening data resources, available in the public administration, making the data resources useful and accessible, is an essential interest of the public administration organization operating the governmental structure. With the development and operation of data servicing systems, which are based on effective information fundamentals, these data become real values for the public and business parties too.

The metadata service gives a practical solution to the access of the available data. In the metadata service, the users obtain information about the available data (access location, data about the data host organization, e.t.c).

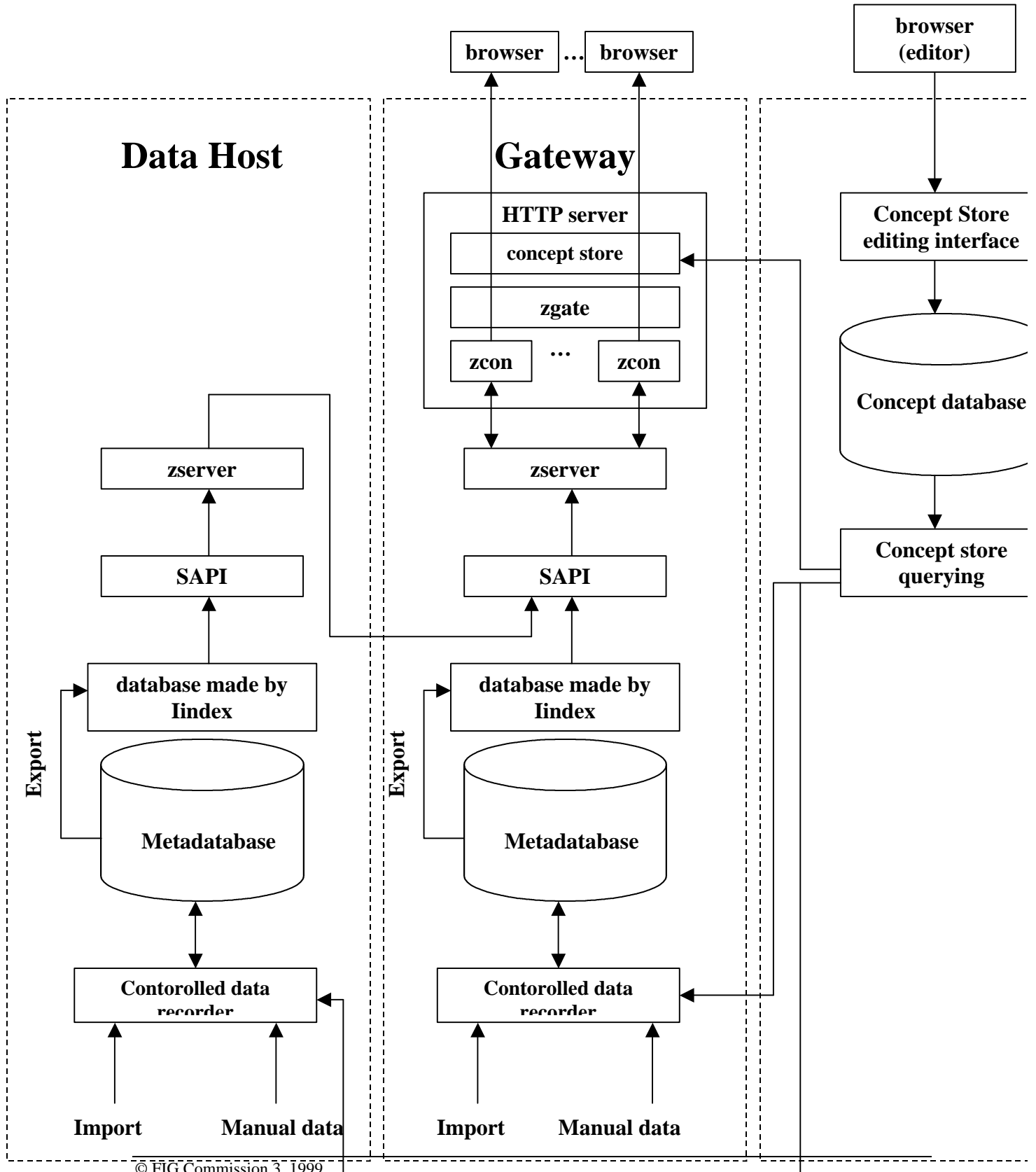
By recognizing the *raison d'être* and significance of such services, the Prime Minister's Office launched a call for bids for the implementation of the METATÉR spatial information metadata servicing system, and the Scriptum Inc. won the call.

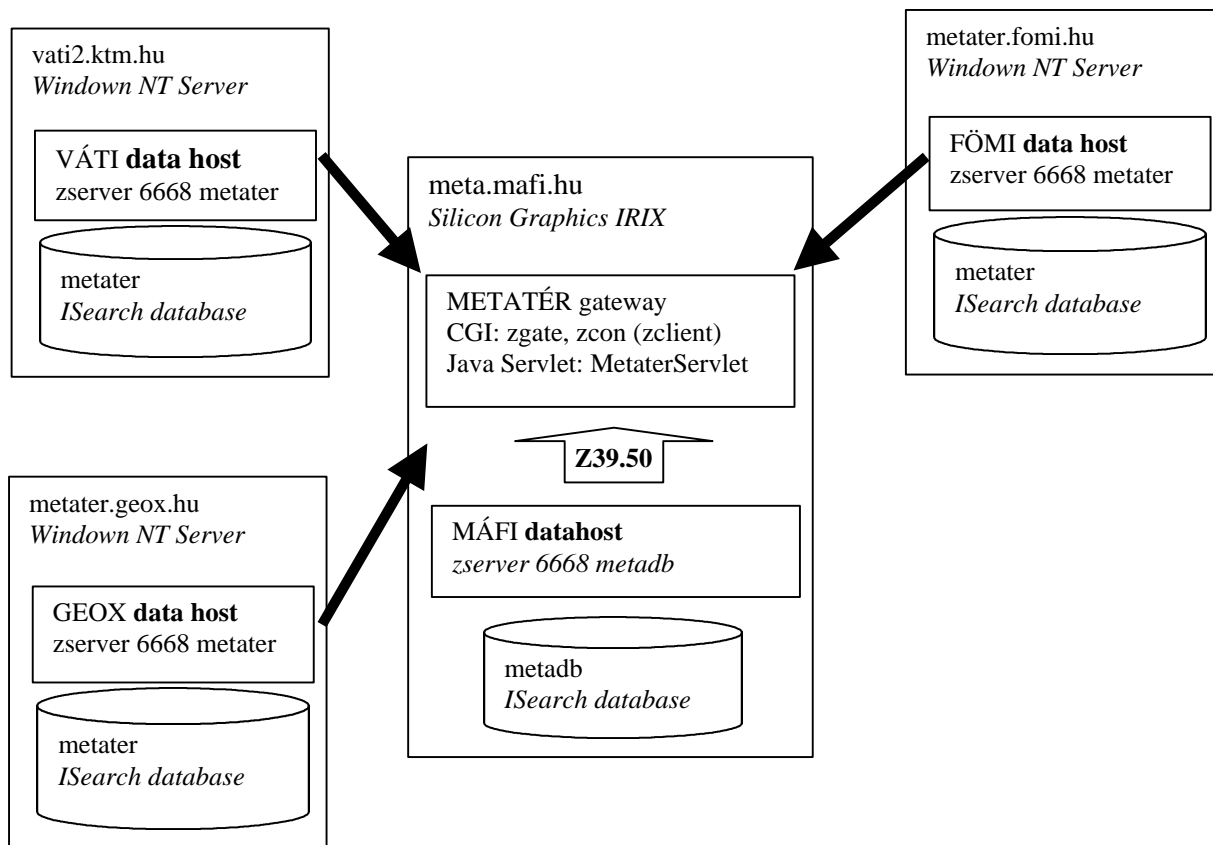
It is worth examining the structure and functioning of the system, as it can serve as a basis for further metadata-servicing initiatives, and means a window to other international data servicing projects, through which Hungary and other parties can mutually benefit of learning and using well-functioning systems.

1. The functional structure of the METATÉR pilot system

The structure of the pilot system is based on the unity of the data host – gateway – concept store according to the following figure:

METATÉR structural figure with the associated institutes





The figure introduces the organizational background of the operation of the pilot system representing the levels and technological background by which the (in time) primary data hosts connect to the system.

2. Hardware and software tools necessary for the operation of the system

Gateway

The following pieces of software are necessary for the operation of the system: UNIX operation system, Netscape Fasttrack Server 3.01, Java Runtime Environment 1.1.5 or later version, JRun Servlet Engine 2.3. The hardware components necessary for the operation of the gateway: computer with one or two processors, minimum of 256 MB RAM, 1 GB free hard disk capacity, and with a 256 kbit/s bandwidth permanent Internet connection.

Data Server

Pieces of software necessary for the operation of the data server: Windows NT or Unix operation system, Java Runtime Environment 1.1.5 (under Windows NT the Microsoft Java VM 5.00.3176 is necessary). The hardware components necessary for the operation of the data server: computer with a minimum of 64 MB RAM, 30 MB free hard disk capacity, and with a 64 kbit/s bandwidth permanent Internet connection.

Data owner / Data host: Windows 9x or Windows NT operation system, IBM PC compatible P133 or later model.

Web client

All sorts of instrument, capable of running a web browser, are suitable for the web client. The Microsoft Internet Explorer 4.0, Netscape Navigator 4.06 or Hot Java Browser 3.0 is needed for using the Java applet.

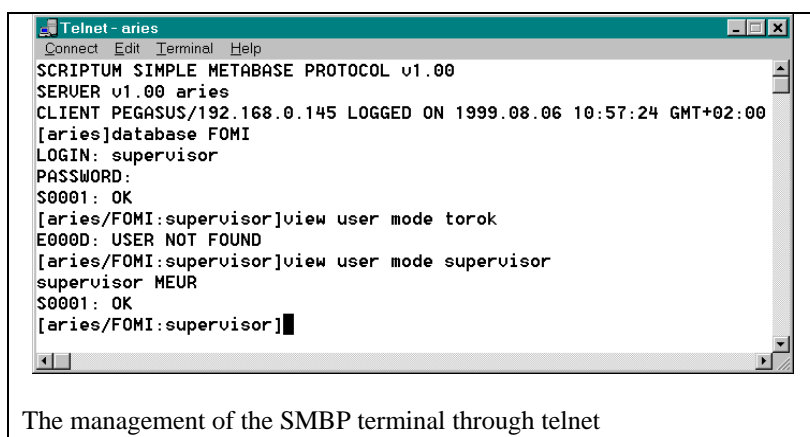
3. The user interfaces of the METATÉR system (Java GUI, Web, telnet, command line)

The user interfaces of the systems of METATÉR are not uniform. The reason of it can be traced back to the different architectures of the individual components.

The interfaces, designed for the end-user, are generally Java-based graphical applications. The metadata recording interface, the mapping applet and the editing interface of the concept store are like this. The typical character of these interfaces is the easy and intuitive usage.

Certain components have an HTML-based interface, usable from a browser. The metadata search has a combined Java and HTML interface and the user interface of the concept store is also made up of HTML pages dynamically generated on the server.

The tools of the system administration are less spectacular as they primarily serve the expert system administrators. These tools are more or less puritan in appearance, however they provide the rapid and efficient system maintenance. A typical example for this is the terminal of the SMBP (Simple Metabase Protocol) servers, which is accessible by a standardized telnet tools.



The management of the SMBP terminal through telnet

4. The options of the modular connection (Z39.50, RMI, SMBP)

An institution can connect to the METATÉR project through the Information Inter-Portfolio Committee (ITB). Through the process the institute obtains the install CD-ROM, to install the necessary software components.

There is nothing to prevent that the institutes develop specific modules according to their own needs, which modules can use the infrastructure of METATÉR, since the components of METATÉR communicate with each other by an open protocol. This means that the description of the protocols is public, therefore any module, which enhances the given protocol, is able to connect to the METATÉR. The overview of the protocols is given in the following sections.

The servers of the data hosts communicate to the system administrators and the data-recording client through the SMBP of Scriptum. The SMBP is a simple protocol and is directly built on the TCP/IP, and in its functionality this protocol resembles very much to one of the most prevailing protocol of the Internet, to the FTP. Compared to the FTP, the most significant surplus is the regulation of access, supporting group work, and its database-centered approach. Clients may be developed to the

SMBP, but even the accomplishment of an alternative SMBP server is possible if the original implementation, which makes part of the distribution of the METATÉR, is not satisfactory.

The data, stored on the servers of the data hosts, can be queried by the standardized Z39.50 protocol of ANSI/NISO, thus any Z39.50 client can make use of these data. Since the gateway itself represents the distributed database as a uniform database behind a Z39.50 interface, then the whole METATÉR can be connected to any database system based upon the Z39.50, e.g. to other international spatial information metadata services.

The Concept Store carries out all of its communication through the protocol of the Sun Microsystems Java RMI (which is also a protocol built on a TCP/IP). If a client program can manage the concept store interfaces visible through the RMI, then it can query the concept store without any problems. With the help of a universal gateway (e.g. ObjectSpace Voyager ORB) the RMI interfaces of the Concept Store can be immediately available for clients communicating with CORBA, DCOM and Voyager protocols.

5. Query Interface

The gateway provides a common, easily accessible and simply usable query interface for the metadata of the data servicing parties within the boundaries of the METATÉR.

The gateway communicates to the data hosts with the help of the Z39.50 protocol. In favor of the wide-scale query options, the query interface is available in an HTML form and in a Java applet.

5.1 *The more modest solution*

The usage of the HTML form is simple and practical but is not reviewable enough when setting up more complex queries. In case of querying according to geographical co-ordinates one needs to know the numeric values of the co-ordinates.

The Web interface uses Java Servlet API and CGI for communicating with the user.

For the implementation, Scriptum took the cgi gateway of the ISite as a basis, which was purposely modified on several points. We put a Java Servlet in front, which communicates to the concept store as well.

5.2 *The Java-based solution*

The Java applet can easily be overviewed and managed. It also has connection to the concept store, and the key words put in can be monitored any time. It provides a map for the query by geographical co-ordinates, on which the queried regions can be allocated.

The Java applet communicates through a HTTP protocol with the Java Servlet of the gateway, as well as to the CGI interface through the servlet. It uses a separate protocol (built onto the HTTP) for map visualization and for the query in the Geographical Name Repository.

6. How and what can be given in a query

The query can be carried out according to three different aspects:

- by geographical expansion,
- by temporal expansion,
- or by fields.

At least one of these aspects shall be given or all of them can be given. If the user gives more than one query aspect, then he will obtain the records, which covers the conditions of all given aspects.

6.1 *Query by field*

Queries can be carried out in three fields simultaneously. The user has to give that in which field he would like to submit the query, that the field shall contain the given text or not, he has to give the queried text, and that what logical connection the next row shall associate with the current row.

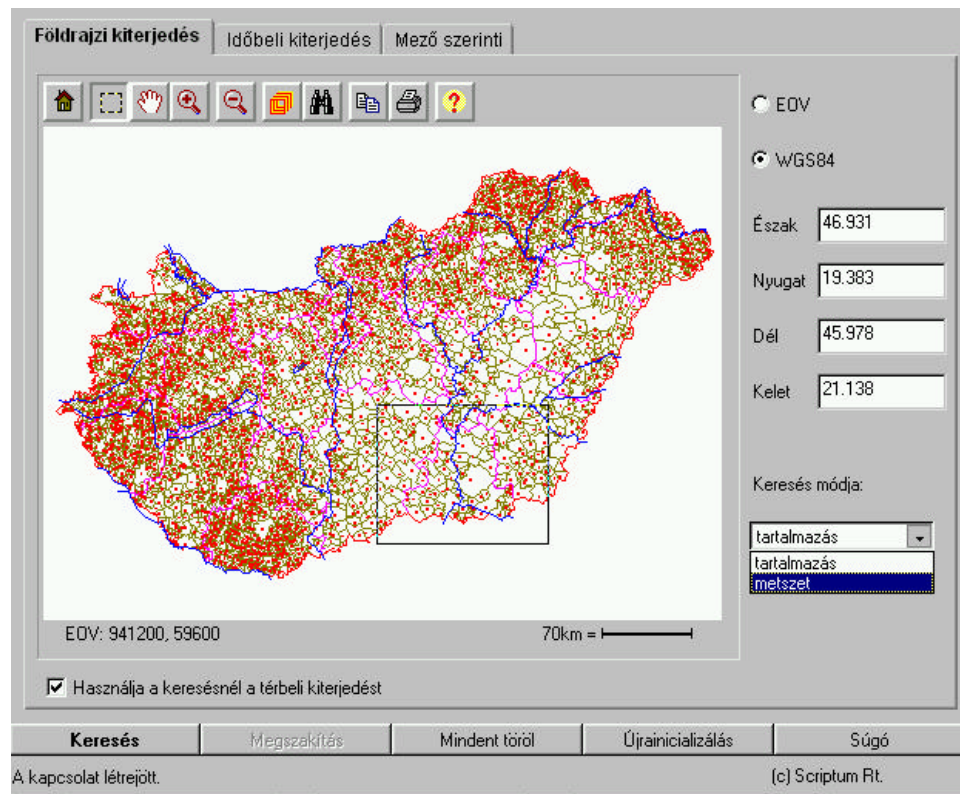
Below a field can be found which a query can be submitted to, and the attributes of these fields according to the Z39.50 standard too:


Nam of the field	Z39.50 attributes
Full text	bib1/1035
Name	bib1/4
Name of the series	bib1/5
Product	bib1/3805
Summary	bib1/62
Status	bib1/3108
Name of the organization	bib1/2024
Theme key word	bib1/2002
Layer key word	bib1/3130
Time key word	bib1/3133
Other key word	bib1/6903

The text given in the first two fields can be automatically queried in the concept store too (Concept Store validity).

6.2 *Query by geographical expansion*

In case of a query by geographical expansion, the queried rectangle (the expansion of the given area) can be given numerically according to the EOJ or WGS84 projection system, common on the HTML form. It is possible to change between the EOJ and the WGS84 co-ordinate systems. The EOJ data shall be given in meters while the WGS84 data shall be provided in degrees.



On the map, these data can be also given by drawing a rectangular ( allocation).

Afterwards queries may be submitted to data:

- which contain the given rectangle in their spatial expansion, or
- whose spatial expansion has a common section with the given rectangle.

Another interesting query option is the usage of the Geographical Name Repository, when metadata, belonging to particular locations, geographic names (county, settlements, regions, important geographical objects) are accessible. This query method can be constricted in a way that a settlement name can be dedicated to a geographic name with the purpose of in which settlement the user wants to find the given geographical object.

Within the boundaries of METATÉR, the Geographical Name Repository -- previously mad by FÖMI -- was complemented by more than 3400 data (county names, settlement names, region names), to which not only the interpolating points but the whole encasing rectangles are available. In case of such a hit the map, in the possible maximum magnification, locates and allocates the area found, placing it on the center of the screen. If only a single interpolating point can be found in the database, then it is placed onto the center on the map and a 4 km by 4 km square is allocated around.

6.3 Query by Temporal Expansion

The query by temporal extension can be carried out according to the temporal extension or to the publishing date of the data set. After defining it a temporal (prior to a given date, within a temporal

interval, after a given date) or relative (how early the data set is to a given day) query can be accomplished.

7. Map Management Options of METATÉR

The map management options of the METAÉR pilot project mean a certain, task-oriented implementation of the Internet-based map and photograph servicing system of Scriptum Inc.

The navigation on the map has three operational modes:

- allocating mode: information may be obtained from the map segment allocated by the user,
- pan mode: the user can freely move the map in order to reach the area examined,
- magnifying mode: the system can magnify a map square assigned by the user, in order to obtain more detailed information.

In the meantime, the system enables the separation and independent examination of the different map layers (settlement border, county border, country border, waters).

8. Concept Store – The Other Interesting Part of the Querying

The Scriptum Concept Store is a novel software system, which cannot be classified into any existing software category. If one would like to characterize it shortly, he could describe it as a group-work-oriented bilingual thesaurus. However, the concept store means something more behind.

- It is a centralized terminology dictionary, accessible on the Internet, which advances the terminology usage of experts working in the same subject area.
- It provides a productive group-work for even in the case of geographically separated editorial teams, using the Internet as a carrier medium.
- It provides the formation of an editorial hierarchy, which uses strictly controlled authorization levels. The system stores the responsibility and superiority relations. By using these relations the system provides simple work-flow management, as a matter of fact it defines permissions.

Physically the Concept Store is implemented through a program, namely through the Scriptum Concept Store Server. Its operation requires a computer with a continuous Internet access, running Microsoft Windows NT 4.0 operation system.

The rich grammatical and semantic knowledge stored in the Concept Store is used by more components of the METATÉR:

The key words can be selected from the Concept Store in the metadata-editing program. Thus it is assured that the key words are obtained from a controlled, finite set, and this greatly increases the accuracy of the retrieval.

The metadata editor also utilizes the bilinguality of the Concept Store, which provides the automatic, rough translation of the Hungarian version of the metadata record into English (or vice versa).

The Java-based query interface of the METATÉR also communicates with the Concept Store: the meaning of the key words, selected for query, can be controlled from the interface. On the other hand the Concept Store makes the query intelligent: if the query is not successful with a key word set, the query system automatically searches the synonyms of the key words from the Concept Store, and with them repeats the query.

The Concept Store has an open, HTML-based interface with which the conceptual databases, stored in the server, can be queried. In turn, a separate, Java-based program is available for editing the Concept Store, freely given to the editors. The editing program, on a graphical interface, provides the option for editing, contribution, mailing and the maintenance of the authorization

system on an intuitive manner. With its built-in HTML display the look of the edited concept on the open interface can immediately be seen.

