Advances in Environmental Research 5 (2001) 415–424

Advances in Environmental Research

www.elsevier.com/locate/aer

Visualization techniques in metadata information systems for geospatial data

Stefan Göbel[1,*], Uwe Jasnoch

Fraunhofer Institute for Computer Graphics, GIS department, Rundeturmstr. 6, D-64283 Darmstadt, Germany

Accepted 2 January 2001

Abstract

This paper describes MIS in general and presents — by means of an information system for residual waste — several approaches to overcome the lack of usability in these information systems for environmental and geospatial data. MetaViz uses visualization techniques and spatial metaphors to generate effective and intuitive representations of the metadata. InfoCrystal is settled in the field of mathematics and tries to facilitate the query formulation and query refinement process by using abstract Venn diagrams instead of traditional fill-in forms. VisDB and TileBars are two approaches which have originated in the field of visual data mining/exploration, aimed at visualizing as much information as possible. Finally, some suggestions for further research in that field are described. The more information and data[2] that are produced in the actual information society, the more important mechanisms and systems which organize the data and include information on where to find these data become. Most popular peculiarities of such information systems are digital libraries, metadata[3] information systems (MIS) and catalogue systems (CS). Recent initiatives to geospatial libraries and environmental MIS/CS provide access to a wealth of distributed data, but offer only basic levels of interactivity and user assistance. This includes all steps of the information retrieval process: query formulation, query modification, comparison of (metadata) result sets and detailed presentation with respect to the visualization of result sets. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Metadata; Environmental information systems; Residual waste; Landfills; Geospatial data; Visualization; Information retrieval; Visual data mining; User assistance

* Corresponding author. Tel.: +49-6151-155-418; fax: +49-6151-155-444.
E-mail addresses: stefan.goebel@igd.fhg.de (S. Göbel), uwe.jasnoch@igd.fhg.de (U. Jasnoch).
[1] Internet: (http://www.igd.fhg.de/igd-a5/).
[2] Between 80 and 85% of all data do have a spatial reference, e.g. coordinates, geographic entities or postal addresses.
[3] Metadata or 'data about data' describe the contents, quality, origin, conditions and other characteristics of data.

PII: S1093-0191(01)00093-4

1. Introduction

Digital libraries help users to search for, evaluate, and retrieve appropriate information. Metadata is indispensable in this process of locating information and evaluating its fitness for specific tasks. Beard and Sharma (1998) suggest a multi-level organization and presentation of metadata for spatial information in digital libraries. Corresponding to the three stages of resource discovery (overview, search and details), they identify the following three functions of metadata:

- to provide an overview of digital library content (collection level);
- to enable comparison of multiple information items (search result level); and
- to provide detailed description of individual items (metadata for individual spatial information objects).

Nevertheless, apart from the steps of query formulation and modification, all three levels of metadata require appropriate visualization techniques in order to improve usability. The following chapters present several visualization techniques and analyze its concepts referring to the functional roles of metadata and the usage inside MIS for geospatial data in general. These approaches cover not only typical information or scientific visualization techniques, but also techniques originated in the fields of information retrieval, visual data mining or abstract mathematics.

As a representative MIS for environmental/geospatial data in this paper, an information system for residual waste is considered. In that scenario, traditional search criteria with respect to typical data characteristics or metadata attributes include dates and time periods (notification date of landfills, temporal extent of residual waste sanitation), the location/spatial reference and category of existing facilities and landfills such as municipal solid waste landfills or composting, or concentration of pollutants.

1.1. Metadata information systems

The more information and data produced by the actual information society, the more important the mechanisms and systems which organize the data and include information on where to find these data become. Most popular peculiarities of such information systems are digital libraries, metadata information systems (MIS) and catalogue systems (CS).

The main difference between these systems is the extent of the information spectrum covered by these systems. Whilst, e.g. catalogues for department stores or phone directories/books are typical representatives of CS covering a relatively clearly outlined information spectrum (address, phone number, article no., price, etc.), other systems cover a quite larger spectrum. Examples of this are environmental MIS, such as the UDK (German: Umweltdatenkatalog; English: Environmental Data Catalogue) on a national basis, or GELOS (Global Environment Information Locator Service) and EIONET (European Environment Information and Observation Network) on a European basis, GEIXS as European MIS for scientific geospatial data (e.g. geology, geography) or GCMD (Global Change Master Directory) and EOSDIS (Earth Observation System Data and Information System) as representatives of MIS in the EO community.

Fig. 1 shows an overview of existing MIS and institutions/organizations involved in MIS applications and projects in Germany, Europe and the United States. The geospatial market, with respect to all kind of geospatial data are subdivided into environmental data, geoscience (geology, geography) and earth observation data. From the technical point of view, the most important parts of these systems are the different metadata formats and protocols for data exchange. Whilst in USA, the FGDC format is widespread and used in the NSDI clearinghouse (coupled with the protocol Z 39.50 and the GEO profile), in Germany and Europe especially the situation on the geospatial market and MIS market for geospatial data is very heterogeneous.

Therefore, with regard to the harmonization and integration of existing MIS/CS into networked distributed information systems, especially on an international basis, the necessity of standardized data protocols and metadata formats becomes obvious and results in various activities and projects. Typical examples which underpin that tendency are WebCDS (web-based version of the multilingual catalogue of data sources), INFEO (information system for earth observation data), ESMI (European spatial metadata infrastructure) or InGeo-MIS (MIS of the InGeoForum,[4] realized as a broker system which integrates metadata from various application areas such as environment, traffic, earth observation or geology and offers a public gateway to locate and compare geospatial data).

2. Visualization techniques

This chapter presents several visualization techniques and analyses its concepts referring to the several stages of resource discovery in MIS for geospatial data. The techniques are subdivided into groups which

[4] InGeoForum: Information and cooperation forum for geospatial data of the ZGDV e.V., Darmstadt.

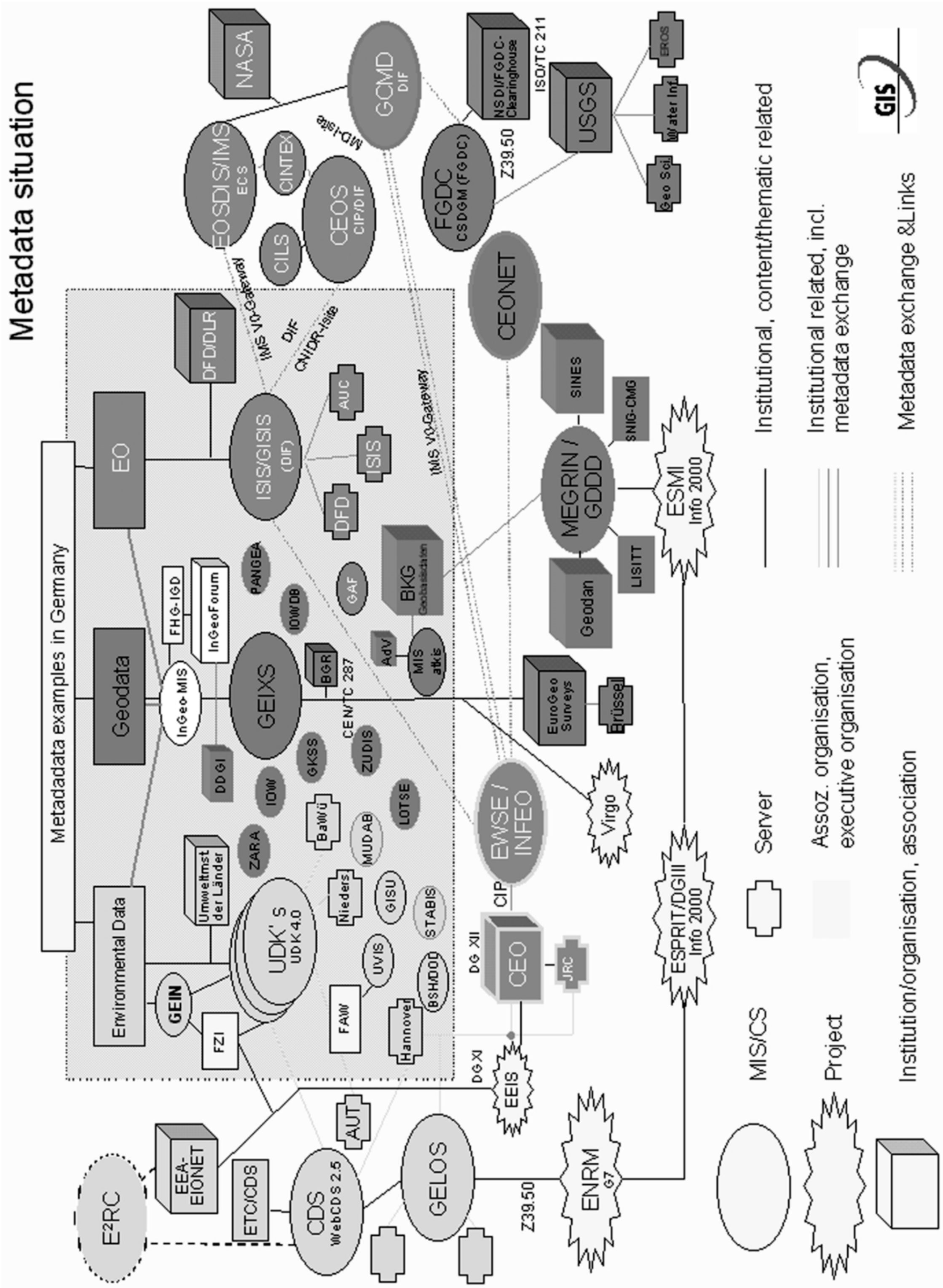

Fig. 1. Situation of metadata information systems.

Table 1
Mapping metadata attributes in visual attributes in MetaViz

Data characteristics	Metadata attributes	Mapping/Visualization	Literature
Geographic space	Extent	*xy* plane of result scene	Bertin, 1983
Temporal attributes	Publication date, time period	*z* plane of result scene	
Quantitative attributes	Sice, scale	color intensity and/or hue	Ware, 1988
Categorical attributes	Data set formats, update frequency	color hues	

are based on its originating application area: traditional visualization, information retrieval and visual data mining.

2.1. Visualization

The two most important peculiarities of computer visualization in the context of MIS for geospatial data are 'computer-aided cartography' and 'information visualization'. Computer-aided cartography arose in the early 1970s and is still the commonly used visualization form in GIS. Information visualization is a relatively new discipline, which is founded on the rapid growing of the internet and other networked database systems. It dynamically generates visual representations of database structures and data characteristics in order to facilitate the process of locating/searching suitable information.

2.1.1. MetaViz

The MetaViz project (Jung, 1999) addresses two of the major barriers preventing the extensive use of digital libraries: lack of usability and information overload. The research is focused on geospatial data, making it possible to develop effective visualization and interaction techniques that are based on familiar spatial metaphors like maps, landscapes and city models. Hence, both computer-aided cartography and information visualization are used.

The visualization techniques developed transform result sets of GDL (geospatial digital library) queries into a visual form: so-called result scenes. In this transformation process, a number of mappings from metadata attributes to visual attributes of the (3D) result scene are performed. These mappings make sure that metadata attributes can be processed effectively and correctly by the human visual system (Jung, 1995; Mackinlay, 1986). Table 1 summarizes the mapping techniques used in MetaViz.

As visualization techniques for the MetaViz browser, i.e. the visual presentation of result sets in the result scene, surface plots, box piles, temporal box plots, space time plots, glyph plots and aggregation plots are used. Fig. 2 shows an example of a space–time plot: Contrary to temporal box plots, the boxes are no longer uniform in height; instead, the height corresponds to a duration. That time period is expressed by the vertical box position and size. Additionally, in this example, the box color is used to map the starting date of the time period, but it could also map another (quantitative) attribute, such as the file size of a complete metadata set.

A shortcoming of both temporal box plots and space–time plots is that the boxes can overlap. Some boxes may be obstructed or barely visible. It can be especially hard to spot the numerousness of result sets (single metadata sets) that have identical bounding boxes and time periods. Glyph plots avoid this shortcoming by mapping a compact glyph near the centers of the result set bounding boxes. The area of the bounding boxes can be mapped to the glyph size and/or color. Fig. 3 shows a glyph plot of a result scene where glyph size and color correspond to the geographical

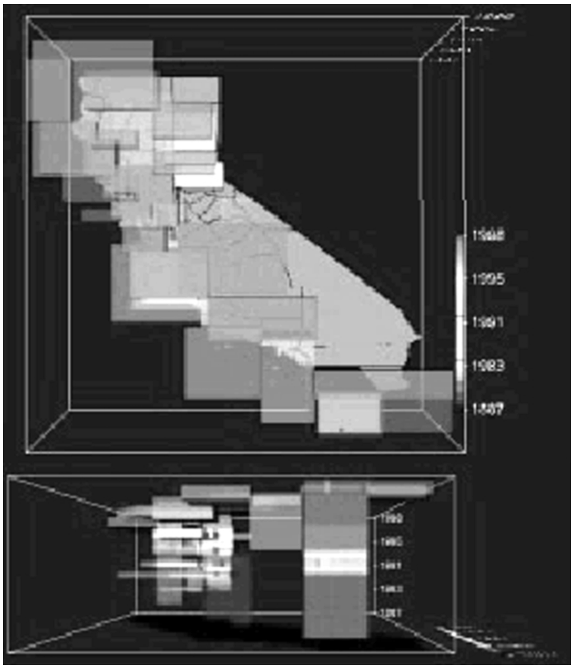

Fig. 2. MetaViz — space–time plot.

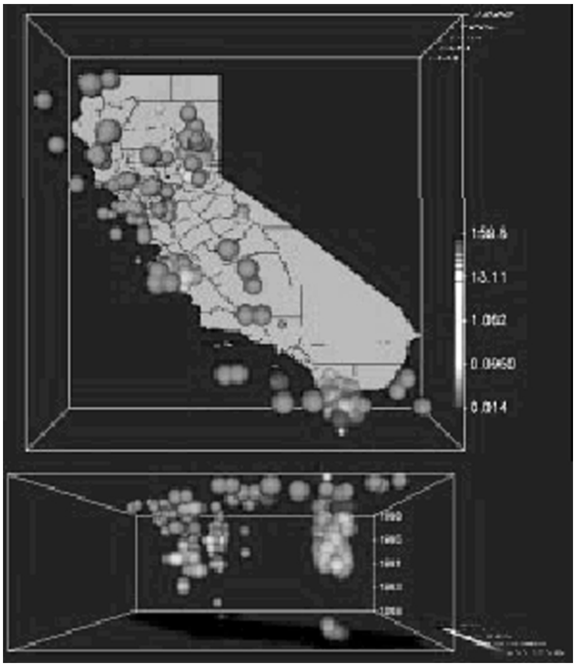

Fig. 3. MetaViz — glyph plot.

area covered by the result sets and the glyph position corresponds to the end of the time period covered. Additionally, the glyph color shows an additional attribute, the (file) size of the result set.

The advantage of MetaViz is that it uses spatial metaphors and landscapes which are very easy to understand. Otherwise, interpreting 3D result scenes needs some visualization knowledge.

2.2. Information retrieval

Typical examples for information retrieval systems are query languages for databases (e.g. SQL) or search engines for the internet (www). The most important disadvantage of these systems is that there is no visual feedback from either query formulation or query modification. Thus, users often neither know how to enhance their queries nor notice any effects in the result set referring to query modifications. Therefore, several concepts and methodologies have been developed in order to improve that situation. One example for this tendency is InfoCrystal (Spoerri, 1993), presented in the next section.

2.2.1. InfoCrystal

InfoCrystal is a visual tool for information retrieval in which concepts are settled in the field of mathematics (algebra). First, each search parameter as part of a query is mapped to a graphical symbol/shape, then the different symbols are combined in the form of Venn diagrams.

Fig. 4 shows an InfoCrystal using Venn diagrams with four attributes, e.g. the appearance of plumb/lead, copper, slag or chlorinated hydrocarbon in the context of an environmental information system for residual wastes. The colored circles at the corners of the InfoCrystal indicate the number of landfills matching exactly one individual search criteria. The inner symbols represent landfills matching a combination of various search criteria (two criteria: rectangles; three: triangles; four: rhombus). Thus, users easily get an overview of landfills matching their search requirements and get an idea about the distribution of various search parameters and potential effects referring to query refinements.

Furthermore, it is possible to include relevance weights and threshold factors. Also, nested queries can be visualized as multi-level InfoCrystals. With regard to the usage of MIS for geospatial data, the most important disadvantage of that technique is the lack of spatial metaphors, the limited number of search criteria and the determination of Boolean type search criteria (is there an appearance of pollutant x at landfill A or pollutant y at landfill B).

2.3. Visual data mining

Data mining usually aims at recognizing semantic relationships in large databases. Often, this task is closely related to knowledge discovery. Due to the complex and high-dimensional characteristics of geospatial data, concepts of visual data mining should be taken into account apart from traditional visualization techniques based on computer-aided cartography.

2.3.1. VisDB

The global aim of VisDB (Keim and Kriegel, 1994) is to (interactively) explore large databases using high-dimensional visualization techniques. Each dimension, with respect to the search criterion of a query, is visually presented for all corresponding datasets in a browser and users can dynamically modify the conditions of each query parameter using sliders. Thus, similar data sets are recognizable by similar patterns (color mapping techniques) in the result scene (2D) and users get an impression (a) about the matching of individual data set attributes; and (b) about the level of correlation between queries and datasets in general. In order to get an overall rank for matching data sets, standardized scales for sliders are necessary. For visualizing individual dimensions, this is not necessary.

The main advantage of VisDB is the possibility of visualizing very large data archives and high-dimen-

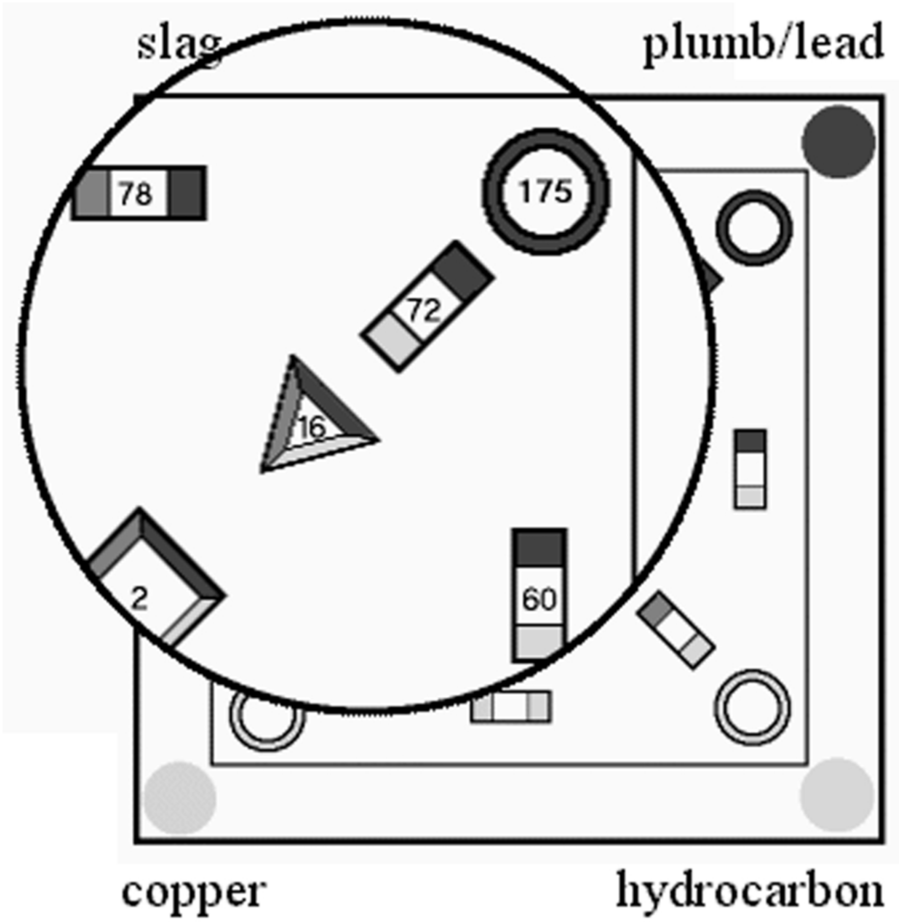

Fig. 4. InfoCrystal.

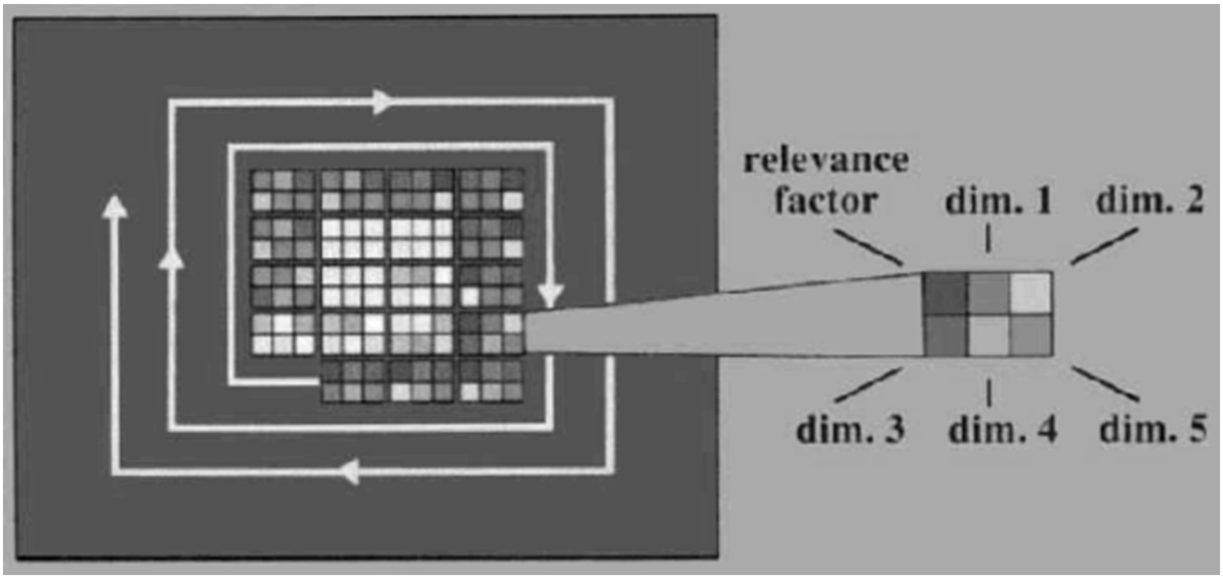

Fig. 5. VisDB — grouping technique.

sional data sets. On the other hand, a lot of experience in reading/understanding result scenes — which are composed by hundreds or thousands of multicolored pixels corresponding to a lot of dimensions with respect to search parameters — and interpreting visual patterns — as results of axes, spiral or grouping techniques — is necessary (Fig. 5). In the case of MIS for geospatial data in the World-wide web (www; e.g. environmental information systems for residual wastes) this assumption cannot automatically be applied.

2.3.2. TileBars

Another example, which aims at presenting as many attributes of data sets and their relationship to queries as possible is TileBars (Hearst, 1995). This approach is funded in the field of information retrieval for text documents, but it uses similar (data mining) visualization techniques such as VisDB; therefore, it is mentioned in this section.

TileBars use a text tiling algorithm to structure documents in segments and simultaneously and compactly indicate relative document length, query term frequency and query term distribution in text segments. Fig. 6 shows an example of TileBars: rectangles correspond to documents (data sets with respect to result sets 1 and 2), each row corresponds to individual search criteria (search parameters, such as a keyword term) and each part of a row refers to a text segment, the color darkness indicates the frequency of terms which match the search term. A grayscale is used instead of color hues (spectra), because varying shades of gray do

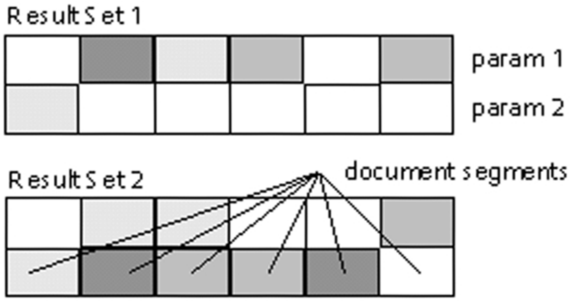

Fig. 6. Tile bars.

have a natural visual hierarchy and show varying quantities better than color (Tufte, 1983).

Referring to the context of MIS, TileBars are useful for searches in long description fields of metadata formats (Göbel and Lutze, 1998), e.g. an abstract, lineage or quality narrative report about a dataset for geospatial data or a general description about existing facilities and landfills including hints for special sanitation issues, which are not associated with individual fields of a metadata format or elements of an environmental information system for residual wastes.

3. Comparison

In order to analyze and evaluate the different approaches referring to its usage in MIS for geospatial

Table 2
Usability of visualization techniques in metadata information systems

Approach visualization technique	Comparison criterion	Data types			Input interactivity	Output/result		Prerequisites	
		metric	ordinal	Boolean		number of data sets (clarity)	Number of params	users: experts	casual users
Residual waste	2D	+ +	+	+	+	+ +	+	+	+ +
MetaViz	Surface plot	O	not relevant		O	+	O	+	O
	Box pile	+			O	+ +	O	+	o
	Temporal box plot	+			O	O	+	+	O
	Space–time plot	+			O	−	O	+ +	O
	Glyph plot	+			O	O	+	+ +	+
	Aggregation plot	+			+	+	+	+ +	+
3D-Techniken	Perspective wall	+ +	+	+ +	+ +	+ +	+ +	o	+ +
	Cone tree	−	−	+ +	−	+ +	O	+ +	−
Information retrieval and visual data mining	Info crystal	+	−	+ +	+	+ +	+	+ +	+
	Tile bars	o	+	+ +	+	+ +	+ +	+	+ +
	VisDB	+ +	o	+	+	+ +	+ +	+	−

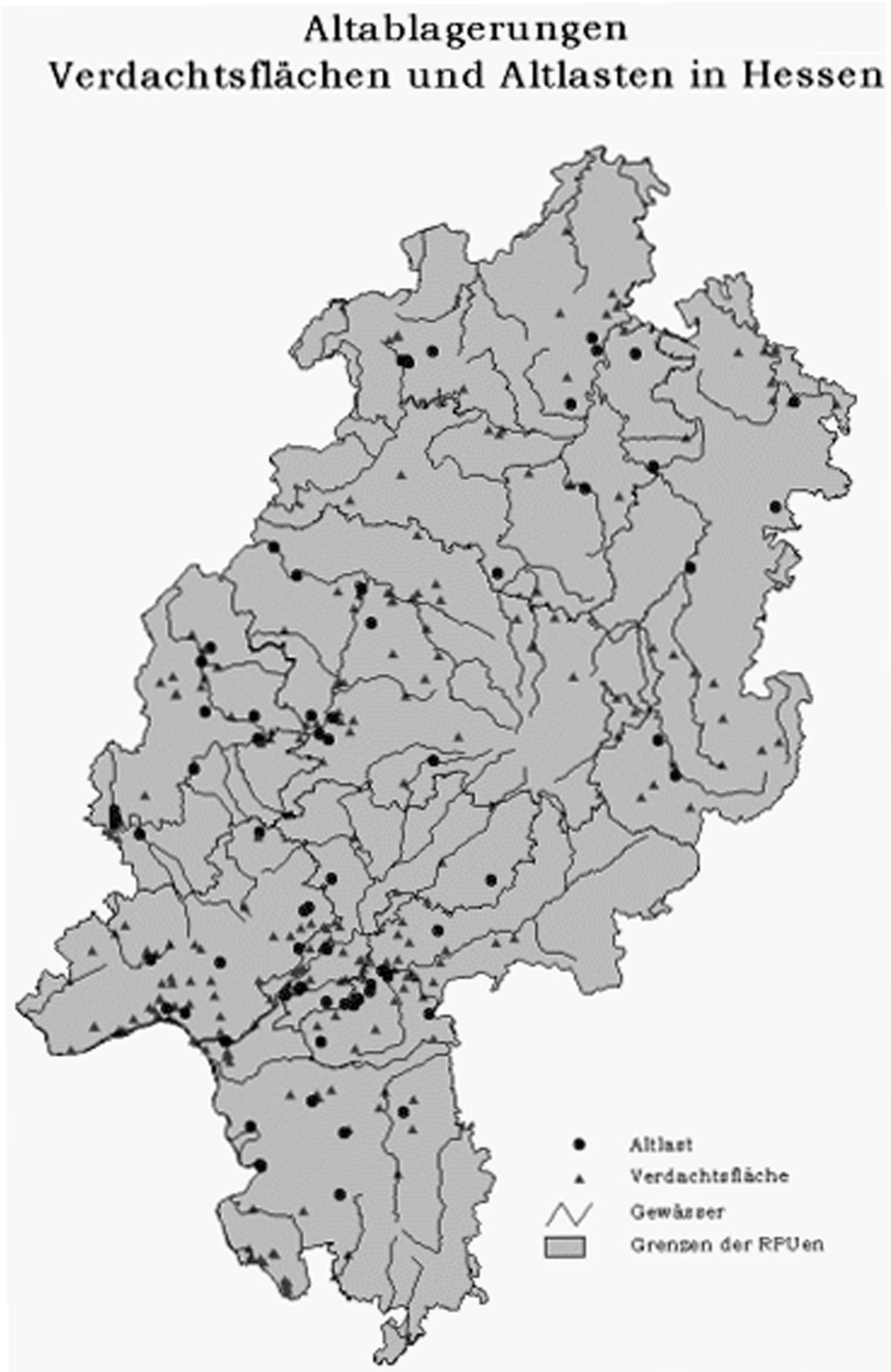

Fig. 7. Environmental information system for residual wastes.

and environmental data, the following criteria for comparison have been taken into account:

- *Which data types can be visualized?* Possible data types are metric (quantitative, e.g. scale, size, number of datasets or distances between geographic entities such as landfills and schools), ordinal (e.g. a category, domain values such as data formats, peculiarities of landfills: composting, dump, municipal solid waste), Boolean (for instance ISO (ISO, 2000) compatibility of metadata sets or appearance of various pollutants at landfills), or non-metric. Examples for the latter are common description fields of metadata such as abstracts, lineage statements or quality narrative reports or general descriptions of landfills.
- *Graphic interactivity for user assistance.* Which kind of interactivity do the approaches offer referring to query formulation and refinement? Are these methods intuitive and are they integrated in browsers: graphical user interface consisting of search form and result scene.
- *How many result sets and parameters can be visualized in a clear way?* This means the number of matching documents with respect to datasets or hotels and the number of dataset attributes (metadata fields or characteristics of landfills).
- *Which prerequisites are necessary to use / understand included visualization technique(s)?* Some visualization techniques are intuitively usable for casual users (such as spatial metaphors and landscapes), others need some kind of expert knowledge, such as pattern matching/detection in VisDB or interpreting Venn diagrams in InfoCrystal.

Table 2 shows the evaluation results relating to the criteria listed above. In addition to the visualization techniques and approaches presented in this paper, some other 3D techniques [Lyber World (Hemmje et al., 1994), Perspective Wall (Mackinlay et al., 1991) and Cone Tree (Robertson et al., 1991)] are taken into account.

4. Summary and conclusion

This paper presents several visualization techniques and their usage in MIS for geospatial data. All approaches have their strengths and weaknesses, as outlined in Section 3. Referring to the functional roles of metadata suggested by Beard and Sharma (1998) and the different stages in the process of resource discovery, the presented techniques primarily aim at the relationship between search forms, query refinement and result set presentation, especially in relation to the comparison level of metadata. Some of the techniques are very scientifically oriented and pre-suppose some expert level in visual structure interpreting (VisDB) or mathematics (understanding Venn diagrams in InfoCrystal). Others are limited to special data types (e.g. TileBars for non-metric data, text documents; respectively, metadata fields with type String, free text; see ISO (2000) or MetaViz primarily for metric, quantitative data). In contrast to the visualization capabilities, casual users in particular rely on intuitive methods and metaphors such as spatial metaphors, landscapes or ordinary lists as a presentation form for result sets. Fig. 7 shows an example of an environmental information system for residual wastes/landfills of the state Hesse (ALTIS, 1999), where 2D maps are used as spatial metaphors and visualization technique for the presentation of result sets. Landfills matching queries are symbolized on the map. Alternatively, 3D boxes could be used combined with transparency which tells the user the overall rank (computed by a distance function over all search parameters and relating characteristics of the landfills, see Keim and Kriegel, 1994) of a landfill corresponding to a query: a distance of zero would result in a landfill (box without transparency) which matches all query parameters, increasing distances would result in higher transparency.

In future, other more abstract approaches such as InfoCrystal or Lyber World shall be integrated into MIS as visual information retrieval techniques, while other 3D techniques like MetaViz shall be tested as visualization techniques to present result sets in so-called result scenes instead of/in addition to traditional 2D spatial metaphors such as maps. Further research will focus on the evaluation of current approaches and the creation of suitable test scenarios, including several of the approaches discussed in this paper. In addition to the technical aspects, a lot of time shall be investigated to determine representative user groups in the environmental sector and criteria for comparison of the different techniques.

References

Beard, K., Sharma, V., 1998. Multilevel and Graphical Views of Metadata. Proc. IEEE Advances in Digital Libraries, Santa Barbara, CA.

Bertin, J., 1983. Semiology of Graphics. The University of Wisconsin Press, Madison.

Göbel, S., Lutze, K., 1998. Development of Meta Databases for Geospatial Data in the WWW. Proc. ACM GIS'98, Washington, DC.

Hearst, M., 1995. TileBars: Visualization of Term Distribution Information in Full Text Information Access. Proc. ACM SIGCHI'95, Denver, CO, 59–66.

Hemmje, M., Kunkel, C., Willet, A., 1994. Lyber World — A

User Interface Supporting Fulltext Retrieval. Proceedings of SIGIR'94. ACM Press, Dublin.

Altlasten-Informationssystem des Landes Hessen, Broschüre des Hessischen Landesamtes für Umwelt, Dezernat Altlasten, Wiesbaden, Mai 1999.

ISO. Draft International Standard on Metadata for Geographic Information, ISO 19115-Metadata (status: 2nd CD, first notation was ISO 15046-15), see http://www.statkart.no/isotc211/.

Jung, V., 1995. Knowledge-Based Visualization Design for Geographic Information Systems. Proc. ACM GIS'95, Baltimore, MD. ACM Press, New York, pp. 101–108.

Jung, V., 1999. MetaViz: Visual Interaction with Geospatial Digital Libraries. Technical report TR-99-017. ICSI Berkeley, CA.

Keim, D.A., Kriegel, H.-P., 1994. VisDB: Database Explorata-tion Using Multidimensional Visualization. IEEE Computer Graphics and Applications, 40–49.

Mackinlay, J., 1986. Automatic Design of Graphical Presentations. PhD thesis. Stanford University, Computer Science Dept, Stanford CA.

Mackinlay, J.D., Robertson, G.G., Card, S.K., 1991. The Perspective Wall: Detail and Context Smoothly Integrated. Proc. CHI'91, ACM Conference on Human Factors in Computing Systems, New York, 173–180.

Robertson, G.G., Mackinlay, J.D., Card, S.K., 1991. Cone Trees: Animated 3D Visualizations of Hierarchical Information. Proc. SIGCHI'91, 189–194.

Spoerri, A., 1993. InfoCrystal: A Visual Tool for Information Retrieval. Proc. IEEE Visualization'93, 150–157.

Tufte, E.R., 1983. The Visual Display of Quantitative Information. Graphics Press, Chesire CT.

Ware, C., 1988. Color sequences for univariate maps: theory, experiments, and principles. IEEE Comput. Graph. Appl. 9, 41–49.

Vitae

Stefan Göbel has studied Computer Science at the Technical University of Darmstadt and is a researcher at the Fraunhofer Institute for Computer Graphics since August, 1997. His main interests are in the field of web-based information systems for geospatial data and information visualization. The topic of his graduation thesis deals with graphic-interactive user guidance to geospatial data archives.

Dr Uwe Jasnoch leads the GIS group at the Fraunhofer Institute for Computer Graphics in Darmstadt since 1998. He has previously worked as a researcher at the department for industrial applications. He graduated from the Technical University of Darmstadt in the field of CAD systems to support concurrent engineering.