iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Graphical User Interface Design for Climate Impact Research Data Retrieval

M. Wrobel [a], **T. Nocke**[a], **M. Flechsig**[a] **and A. Glauer**[a]

[a]*Potsdam Institute for Climate Impact Research(PIK), Telegrafenberg, 14473 Potsdam, Germany (wrobel/nocke/flechsig/glauer@pik-potsdam.de)*

Abstract: Complex, heterogeneous environmental data sets are produced and stored in many scientific disciplines. For managing complex structured data, database and data warehouse technologies have been established, providing access via standardized interfaces. Nevertheless, there are some obstacles for researchers in applying the full functionality of sophisticated data management systems. Here, graphical user interfaces (GUI) are essential for bridging the gap between such systems and scientific users. However, designing / developing user interfaces for complex, environmental data sets is not trivial, facing user diversity and a variety of tasks. Consequently, for the example of climate impact research, this paper introduces an approach to support data retrieval. We propose tailored solutions for metadata navigation and filtering as well as for visualization and visualization design, and present three tools in combination with lessons learned for GUI design.

Keywords: Human Computer Interaction, Graphical User Interfaces, Data Retrieval, Climate Impact Research, Visualization

1 INTRODUCTION

Easy-to-use access to information from large heterogeneous environmental data sets is not straightforward. In particular in the multi-disciplinary field of climate impact research, providing appropriate tools for data retrieval is hampered by various obstacles, resulting from the heterogeneity of users as well as data and a dynamically developing research environment. Utilization of these data sets in an integrated manner requires the following two steps: (1) data integration and (2) provision of an appropriate design of human computer interaction (HCI) for data retrieval. In this paper, we focus on the second step. Therefore, we identify challenges that arise for graphical user interfaces (GUI) in climate impact research (sec. 2), outline related work (sec. 3), present our approach for data retrieval in our specific background (sec. 4) and resulting solutions (sec. 5). Finally, we discuss these solutions and the developed tools (sec. 6) and conclude with a summary and challenges for future work (sec. 7).

2 PROBLEM ANALYSIS

Designing intuitive access mechanisms to data in our context faces a variety of challenges, above all (i) data heterogeneity and complexity, (ii) the gap between required functionality and intuitive access to this functionality, (iii) user diversity and the variety of user tasks and (iv) the demand for smooth support of and integration into the users' ongoing research processes.

Data heterogeneity and complexity. First of all, data required for climate impact research refers to a variety of disciplines, has to be compiled from various sources, and can – due to the data complexity (e. g., different kinds of spatial grids, resolutions, formats) – not be managed using a single database scheme. Thus, comfortable integrated access is severely hampered.

Gap between required functionality and intuitive access to this functionality. Efficient access to data in our context may not be provided using a single standard software tool, but requires utilization and combination of functionality, partially provided by database management systems, geographical information systems, visualization systems and even programming languages. Since only very few users can be expected to be experts in applying all of these, there is a gap to be bridged between the required functionality, the intuitive access to it and the functionality provided by standard technologies and tools.

User diversity and the variety of user tasks. The HCI design in our context must be adaptable to heterogeneous users groups, including users with different skills and qualification grades (from students to senior scientists), with different retrieval goals (scientists, potentially also stakeholders and the general public) and from different disciplines (e. g. meteorology, hydrology, socio-economy). Moreover, these users are confronted with a variety of tasks, including gaining an overview of the available data sets, metadata and data navigation, filtering of subsets of interest, retrieval of derived / aggregated data and data exploration.

Demand for smooth support of and integration into the users' ongoing research processes. Identifying the required functionality needs input from many climate impact researchers. Their knowledge has to be incorporated into the design of data retrieval tools and appropriate interfaces without hampering scientists' ongoing, genuine research tasks. Here, a specific situation in our working context arises due to the fact that data sets required for specific research tasks are - for practical reasons - compiled and managed by individual scientists. This results in a locally available and quality-assured data landscape, which is still heterogeneous and steadily growing. Solutions for integrated access to these data resources have to achieve acceptance both from scientists that provide individually managed data, and those who need to access it. Here, well-designed functional graphical user interfaces are a driving argument to shift towards integrated data sources, competing with established work-flows based on proprietary data structures.

3 Related Work

Evolving information technologies are not easiliy adaptable to scientific requirements. For instance, relational databases that are the de-facto standard for data management in commercial applications, still have a low acceptance for everyday scientific data management processes (see Gray et al. [2005]). To reduce obstacles for managing heterogeneous data, material data integration based on data warehouses has been developed, facilitating flexible access to previously heterogeneous data (see e. g. Inmon [1996]). However, the resulting metadata and data landscape are still complex. Thus, concepts in human computer interaction concerning improved interface design (see e. g. Norman [1988], Shneiderman [1998]) are essential for reducing the gap between information technologies managing data and domain researchers.

In addition, visualization techniques have been established to filter, evaluate and explore data (see e. g. Bürger and Hauser [2007]). Besides a multitude of methods for spatial and temporal data representation (e. g. Aigner et al. [2008]), visual exploration techniques for multi-variate data (e. g. Bürger and Hauser [2007]) and metadata (e. g. Yang et al. [2005]) have been introduced.

4 Approach for data retrieval in climate impact research

Meeting the challenges outlined in section 2, our approach is to provide retrieval interfaces for integrated, data warehouse based data management structures, enabling intuitive access to metadata as well as to measured and simulated spatio-temporal environmental data sets. The approach is pragmatically-driven, balancing between user freedom and the task-related and computational restrictions. Our procedure is

1. **to develop the software iteratively in tight interplay with the users**: In our case, complete a priori specifications are usually not available or tend to be risky and failure prone. Instead, continuous and mutual learning of users and developers is required. Thus, we follow an incremental process model for software development and apply prototypes to involve users in this process.
2. **to identify required and attractive functionality and make it available intuitively:** we combine know-how from different information technologies and select/adapt suitable techniques for the specific requirements in our context (DBMS, GIS, HCI and visualization).
3. **to provide individual paths based on a modular interface design with recurrent interface elements:** Taking into account the given multitude of user abilities, preferences and tasks, we strive to support individual user approaches instead of forcing individuals to complete a rigid number of interaction steps. Additionally, due to the advantages of decomposing complex user tasks into simpler ones [cf. the OAI-Model, Shneiderman, 1998], we modularize our GUIs into subtask-specific elements that are easier to understand and handle. Moreover, recurrent usage of these elements – where appropriate – contributes to

both, increased consistency and reduced learning overhead.

4. **to choose interaction metaphors the users are familiar with:** although coming from different scientific disciplines, many users in our context are familiar with interaction paradigms such as form filling for data filtering or adjusting color maps for visualization. We build on such metaphors, enhance them if necessary and provide support for novice users.

As a result, we design portable, efficient, extensible object-oriented software (in Java and C++), providing easy-to-use and uniform access to climate-related data, supporting an accelerated, quality-ensured research process. Furthermore, as added value for data retrieval of simulation model output, we incorporate data analysis / exploration functionality.

5 Solutions for Climate Data Retrieval

In this section, we propose solutions for making the heterogeneous climate impact data sets easily accessible. We provide appropiate mechanisms to navigate through the available databased on metadata, and to filter and represent it. Therefore, we distinguish the following steps: (a) navigation through the data landscape, (b) filtering of data subsets, (c) selection of data representation technique, and (d) interactive visualization.

5.1 Navigation through the data landscape

To simplify access to the various available data sources we investigated specific GUI strategies providing overview and navigating through heterogeneous data sources. Here, besides the dialog-based, textual display of metadata, we use visual metaphors as a guide to improve orientation, recognition and comprehension of the provided data.

To access DBMS-integrated station measurement data compiled from world-wide sources (xDat tool), we built **a hierarchy of dialogs** that enables users to navigate through the different station databases and to review relevant metadata (e. g., type of database, content overview, see Fig. 1a,b). Color coding of the database types turned out to be a good orientation hint.

For a file based visual data exploration scenario (SimEnvVis tool), we developed a **tabbed-dialog based metadata representation**. Important types of metadata such as dependent variable ("results"), global, spatial and dataset-related metadata (cf. Nocke and Schumann [2002]) can be represented for several data subsets compactly (see Fig. 1i,j,k). Additionally, small icons are used to improve the comprehension of data/metadata. This includes grouping of dependent variables by subsets (color: subset, shape: the subset's dimensionality, Fig. 1k) and the illustration of subset specific metadata (grid and cell types, Fig. 1j).

Furthermore, to limit the amount of metadata to be displayed for the current task, we apply **task-dependent metadata presentation** for the retrieval of gridded climate data (PixDat tool, see Fig. 2). This greatly reduces the amount of data / metadata to be presented to the user, and thus significantly increases usability. Figure 2a represents the metadata required for displaying a single variable on a colored map, Figure 2b depicts the metadata relevant for visually comparing two variables (from one or two different data sets) and Figure 2d limits the represented metadata sets to data sets applicable for climate diagrams.

5.2 Filtering of data subsets

Filtering of data subsets to be retrieved is tightly linked with the navigation step described above. For instance, the data subset and variable selection in the file based data visualization tool SimEnvVis can be directly executed in the metadata view (see Fig. 1k).

However, for the retrieval of complex database managed data, both flexible and easy-to-use filtering is a severe challenge. To avoid user overload and to reduce complexity, we designed an extensible collection of configurable filter dialogs, each allowing the definition of specific query subtasks. A main filter dialog provides an overview of the filters available for the currently selected database, each filter's status (applied or not) and the currently selected filter criteria (Fig. 1c). Filters are presented on demand; users may freely choose their preferred filters in any order to define individual search criteria. This concept allows to provide GUI metaphors tailored for individual subquery definition tasks. The user may use interactive maps to define spatial restrictions intuitively (Fig. 1d) or hierarchical dialogs to select station types, depicted by both names and

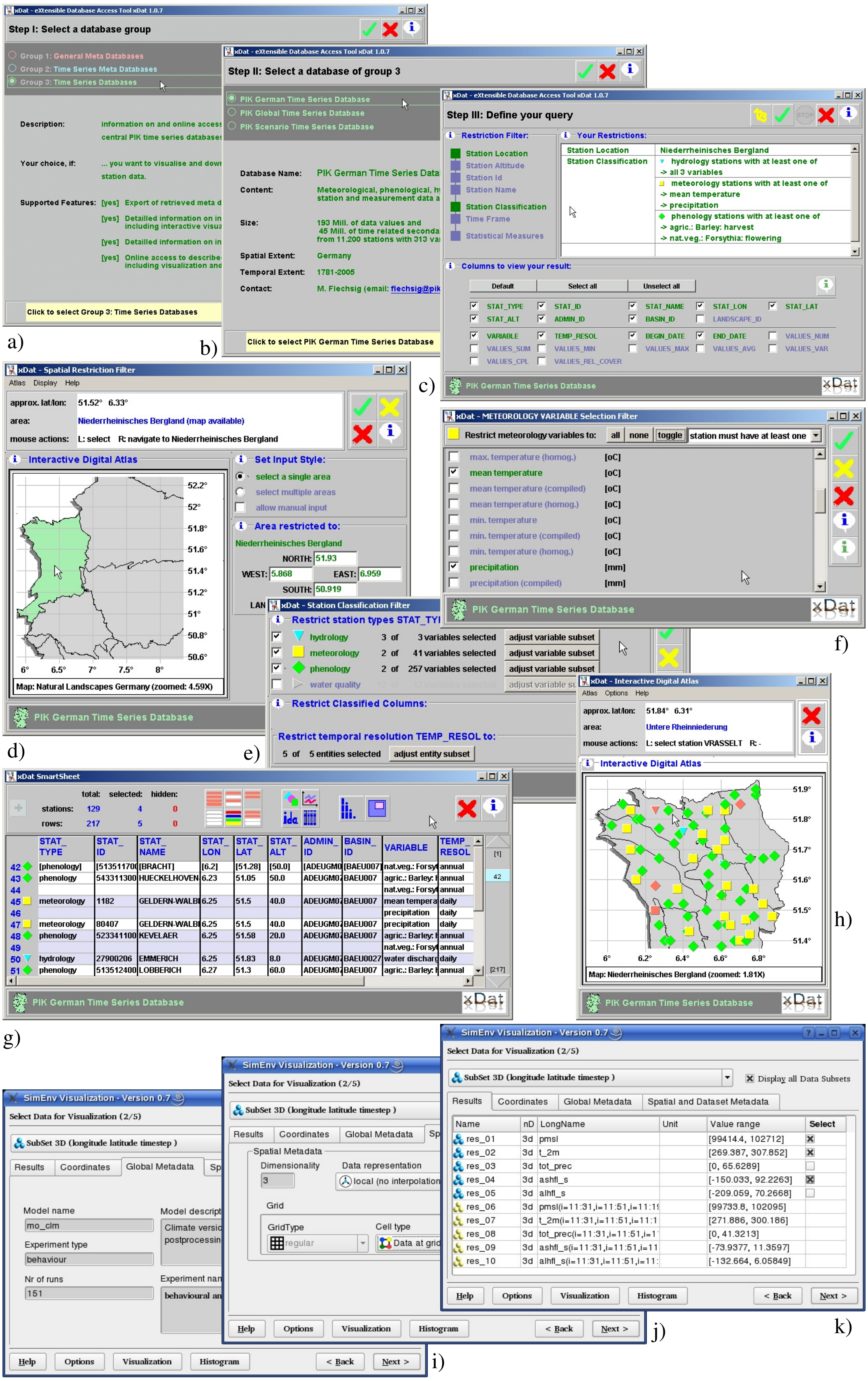

Figure 1: Metadata navigation (a,b,i-k) and data filtering (c-h) with the tools xDat (a-h) and SimEnvVis (i-k)

Figure 2: Task-based data filtering and interactive data visualization with the PixDat tool: a) single data set filtering; b) double data set filtering; c) map visualization; d) climate diagrams

icons, and according variables (Fig. 1e,f). To further reduce complexity, subsets defined using coarse filtering can be further filtered subsequently, e. g. using picking functionality on interlinked map visualization and tabular representation (Fig. 1g,h). Different structures / semantics of databases require flexible GUI adaptation. Thus, we provide configuration mechanisms assigning filters to specific databases and database attributes. For instance, picking an area on a map (see Fig. 1d) will transparently lead to different SQL statements depending on the stored geographical reference of the queried data (e. g. station locations or bounding boxes).

5.3 Selection of data representation

Considering the variety of visualization techniques available to represent the filtered data subsets, user support for selecting and parameterizing these techniques is benefical (e. g. Jiawei et al. [2004]). Thus – for the visual exploration tool SimEnvVis – a rule-based mechanism has been designed, which ranks techniques and supports users to launch them with appropriate starting parameters based on metadata, task specification, user preferences and available resources (Nocke [2007]; Nocke et al. [2007]). Figure 3 top-left illustrates a GUI list representation of automatically ranked visualizations available for the filtered data set, retrieving technique description and feedback about limitations of the technique for the current problem. Based on this information, the user can interactively (re-)select techniques.

5.4 Interactive Visualization of / on a data subset

Data retrieval strongly benefits from visualization techniques supporting different tasks. For intuitive retrieval of geo-spatial data it is essential to provide flexible, intuitive data selection and a first orientation about selected data using interactive maps (e. g. xDat and PixDat). Such maps provide a variety of interaction mechanisms (picking, rubber band selection, zooming), facilitating spatial filtering as well as intuitive comprehension of meta data / data in a geo-referenced manner (see Fig. 1d,h). In PixDat, where maps are used to represent measured or modelled data, we additionally provide a variety of color map definition mechanisms for filtering and data classification

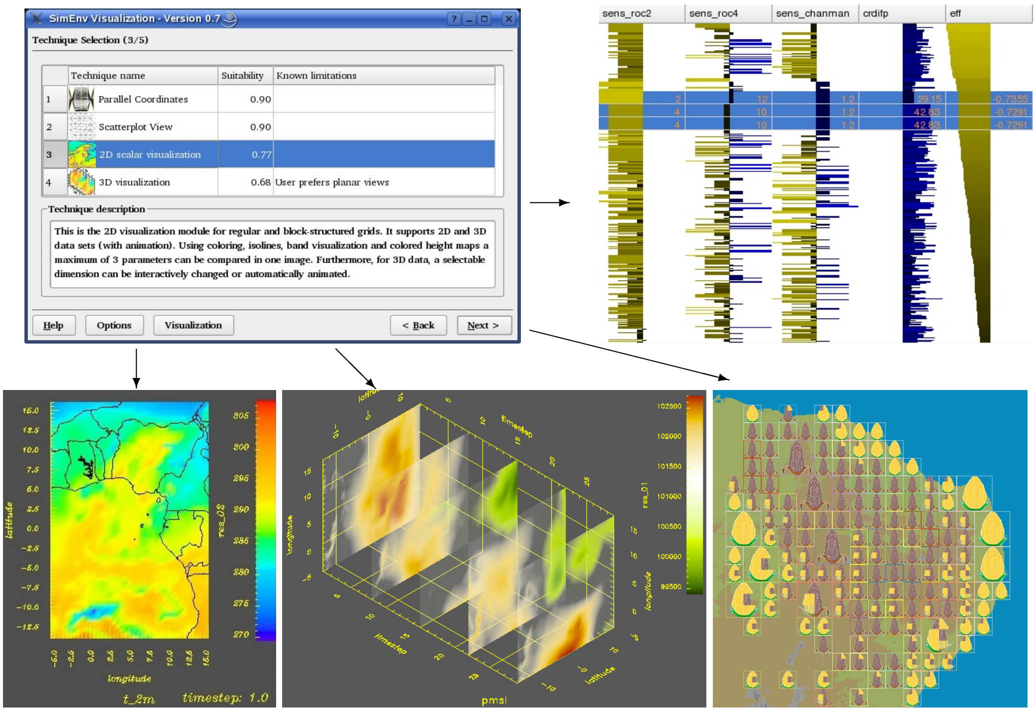

Figure 3: Semi-automatic visualization technique selection in SimEnvVis; top-left: list selection dialog; bottom-left: colored map; bottom-center: transparent slices tracking data extremes over space and time; bottom-right: icon based multi-variate visualization of maize cultivation conditions; top-right: graphical table

(see Fig. 2c) and a tailored visualization of aggregated annual climate conditions for a selected position (see Fig. 2d).

Otherwise, if the focus is on data analysis, a variety of interactive visualization methods are required, going beyond presentation oriented map-based techniques. Therefore, in SimEnvVis, visualization techniques for different data classes (1D, 2D, 3D, nD scalar data) have been integrated (see Nocke et al. [2007]). Besides providing standard techniques in this field (1D graph plots, 2D maps, . . .), alternative visual metaphors for interactive multi-variate data exploration have been elaborated and integrated.

Due to the fact that scientific users appreciate direct feedback about the data values, we additionally provide tabular representations of metadata (see Fig. 1g) and data (see Fig. 3 top-right). Going beyond the presentation of textual values, we added a graphical mode, which maps the data values to bars inside a table cell. To explore multi-variate data with such a table, we provide cell shrinking, interactive re-sorting of rows and columns and displaying details for selected rows of interest. Thus, a large number of data records can be analyzed at a glance and be intercompared (see Fig. 3 top-right).

Furthermore, in the context of climate simulation data exploration, tailored spatio-temporal representations are required for different analysis tasks. Thus, besides the standard (animated) map visualization with color mapping (Fig. 3 bottom-left) and / or isolines, techniques comparing different time slices at a glance (Fig. 3 bottom-center) or comparing variables using metaphor-based multi-variate icons on a map (Fig. 3 bottom-center) have been successfully applied for interactive climate data analysis (see Nocke et al. [2007]).

6 Discussion

It is not trivial to decide whether certain sets of functionality should be integrated into one tool or rather be kept separate by designing many, less complex tools. Decisions should be made carefully, since the first approach leads to increased homogeneity as well as to increased application complexity and learning effort; the latter can lead to easier understandable, less complex applications, however, also to increased application heterogeneity. Avoiding both extremes, three applications integrating the steps outlined in the previous section have been designed and developed. These address various aspects/tasks for data retrieval for climate impact research and assign different interaction schemes for the outlined steps.

xDat (Wrobel [2004]) was designed to provide fast, integrated and direct access to a set of databases containing general and station specific metadata and a large variety of locally compiled time series, gathered from world wide stations and managed using data warehouse methods. The underlying interaction scheme we applied is abd (cf. sec. 5): The user starts choosing a metadatabase of interest (a) and is afterwards provided with flexible and intuitive filtering mechanisms for this database (b). Then, visualization (d), which is implicitly determined by the type of metadata to be retrieved, is used to represent and to perform further filtering on station metadata and to provide quicklooks for selected time series. The applied filter dialog concept (see sec. 5.2) is flexible enough to fulfil various retrieval tasks both for general and station specific metadata and sufficiently intuitive for novice users. The result representations provided – form based for general metadata, tabular interlinked with maps for station metadata – are adequate for typical result evaluation tasks occurring in this context.

PixDat, still under development, is the extension of xDat to access spatio-temporal data on structured grids with different resolutions, managed using a central database (Wrobel et al. [2005]). We apply the interaction scheme $cabd$: the user starts with selecting a task, which directly determines a visualization technique (c), resulting in a simplified navigation process (a) and subsequent subset filter criteria definition (b); finally, visualization (d) is used to present and evaluate the retrieved data. Due to the lower complexity of grid-based climate impact data sets compared to station data, here we apply compact, adaptable dialogs providing all relevant criteria at a glance, immediately adjusting to user selections. For instance, selecting a dataset influences the variables and temporal aggregations offered in the GUI. This immediate feedback supports orientation towards possible interaction options, allows for quick task definition and reduces interaction errors.

SimEnvVis is a visualization environment for the visual exploration of gridded, file-based climate data (see Nocke et al. [2007]). Using a wizard metaphor (interaction scheme $abcd$), it represents metadata of a given data source (a) and reduces the set of possible visualizations by interactive filtering (b) as well as by semi-automatically selecting available visualization methods and parameterizing them (c). Accordingly, a selected visualization technique can be started (d). At any time, users can short-cut this process by starting a visualization based on the current decisions, using standard values for later steps within the wizard. Thus, a flexible support mechanism generating entry points into complex visualization implementation environments (OpenDX, Ferret, OpenGL, Qt) has been developed, greatly simplifying the user's effort in generating images for a certain problem context. Furthermore, user profiling (e. g., distinguishing between novice and advanced users and incorporating preferences for certain visualization systems or metaphors) allows adjustment of the degree of information offered to the user and to individually steer the visualization design process.

For all these tools, our experience is that acceptance is increased if users do not perceive them as a competition to established expert tools they are familiar with. Rather, new applications should complement existing tools by combining a carefully chosen set of relevant 'best of different worlds'-functionality in both an easy to use and efficient manner. For instance, to support flexible query definition and a first intuitive result evaluation not requiring any GIS or SQL knowledge, xDat combines form filling with map interaction. Nevertheless, the user is free to export selected metadata and data directly to (further) explore it using his / her favorite expert tools. In a similar manner, SimEnvVis does not require any detailed knowledge about visualization networks or scripts, however, enables exporting them for later reuse by expert users.

To achieve location independent access to software functionality, *internet* technology has been established as de-facto-standard. Against this background, evolving *web* technologies play an important role for comfortable client deployment. GUI development in this context is still confronted with the well-known drawback between ease of deployment (thin clients) and rich functionality (rich clients). Thus, a retrieval system can benefit from supporting both rich and thin clients. For instance, PixDat can be accessed via a highly responsive Java Swing client (Fig. 2), while a subset of the functionality is also accessible via servlets.

7 Conclusion

We outlined our approach for interactive data retrieval, incorporating strategies for accessing heterogeneous databases intuitively as well as for evaluating and exploring extracted data. Therefore, we investigated GUI strategies to support metadata navigation and data filtering mechanisms accompanied by interactive visualization techniques. Thus, a variety of challenging tasks in the context of climate impact data retrieval have been essentially facilitated. In particular, we contribute a detailed problem analysis for our context, propose a tailored approach for these problems, describe its systematic application to the relevant interaction steps and discuss the resulting data retrieval and exploration tools.

However, there are still challenges for future work. First of all, this includes extending the proposed retrieval mechanisms, e. g. adapting them to highly heterogeneous socio-economic data. A further challenge is to provide a management mechanism for the rather complex filter and visualization parameterizations, improving later reuse and facilitating reproducibility of scientific results. Furthermore, it is necessary to investigate possible synergies of the GUI mechanisms introduced and of the related tools. Moreover, future work needs to investigate the opportunities arising from sophisticated metadata visualization (see e. g. Yang et al. [2005]) for the navigation and filtering of complex data sets.

ACKNOWLEDGMENTS

The authors wish to thank all the scientists at PIK that are testing and using the developed tools for their valuable support and hints. Furthermore, we thank Stefan Petri for LaTeX support.

REFERENCES

Aigner, W. et al. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1), 2008.

Bürger, R. and H. Hauser. Visualization of multi-variate scientific data (state-of-the-art report). In *Proceedings Eurographics'07*, 2007.

Gray, J., D. Liu, M. Nieto-Santisteban, A. Szalay, D. DeWitt, and G. Heber. Scientific data management in the coming decade. Technical Report MSR-TR-2005-10, Microsoft Research, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, January 2005.

Inmon, W. *Building the data warehouse*. John Wiley & Sons, Inc., New York, 2nd edition, 1996.

Jiawei, H., A. Bailey, and A. Sutcliffe. Visualization Design Reuse. In *Information Visualization (IV'04), London*, 2004.

Nocke, T. *Visuelles Data Mining und Visualisierungsdesign für die Klimaforschung (Visual Data Mining and Visualization Design for Climate Research), in German*. PhD thesis, University of Rostock, Faculty of Computer Science and Electrical Engineering, 2007.

Nocke, T., M. Flechsig, and U. Böhm. Visual exploration and evaluation of climate-related simulation data. In *Winter Simulation Conference (WSC'07), Washington D.C., USA*, 2007.

Nocke, T. and H. Schumann. Meta data for visual data mining. In *Proceedings Computer Graphics and Imaging (CGIM'02), Kaua'i, Hawaii, USA*, August 2002.

Norman, D. A. *The Psychology of Everyday Things*. Basic Books, 1988.

Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer-Interaction*. Addison-Wesley, Reading, 3rd edition, 1998.

Wrobel, M. *Multidimensionale, heterogene, visualisierbare Datenräume. Anforderungen, Entwurf und Implementierung einer adaptiven und interaktiven Schnittstelle für transdisziplinäre wissenschaftliche Daten im Kontext der Erdsystemanalyse (in German)*. PhD thesis, Freie Universität Berlin, 2004.

Wrobel, M., J. Glauer, M. Flechsig, and C. Rachimow. *xDat for regular spatial grids*. Status report, Potsdam Institute for Climate Impact Research, http://www.pik-potsdam.de/research/research-domains/transdisciplinary-concepts-and-methods/pixdat/xdat4rsg.pdf, 2005.

Yang, J., M. Ward, E. Rundensteiner, and W. Ribarsky. Value and relation display: A general framework for interactive exploration of high dimensional datasets. *IEEE Visualization Conference*, 2005.