HUDA
A Toolbox for Environmental Report Production

Klaus Tochtermann¹, Wolf-Fritz Riekert¹, Lilijan Kadric¹, Ralf Kramer², Rainer Schmidt², Werner Geiger³, Norbert Peter³, Matthias Reißfelder³, Ernst-Erich Doberkat⁴, Hans-Gerald Sobottka⁴, Andree Keitel⁵, Wolf Zitzmann⁶, Thomas Schütz⁶ and Jacqueline Burkhardt⁶

Abstract

HUDA (Hypermedia Techniques for Environmental Data) is a research and development project in the framework of the environmental information systems of the Federal Republic of Germany and the State of Baden-Württemberg. The objective of HUDA is the design and implementation of a toolbox for the largely automatic production of environmental reports on both printed and electronic media. Exploiting as much as possible the possibilities of electronic media, extending the target audience by providing better and easy access to up-to-date information of the environment and an increasing customer orientation of environmental agencies and ministries provide additional motivation for the HUDA effort. In particular, HUDA supports the environmental report production of the German Federal Environmental Agency and the Environment Protection Agency of the State of Baden-Württemberg.

¹ Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW) an der Universität Ulm, Postfach 2060, D-89081 Ulm, email: tochterm | riekert | kadric@faw.uni-ulm.de
² Forschungszentrum Informatik an der Universität Karlsruhe (FZI), Haid-und-Neu-Straße 10-14, D-76131 Karlsruhe, email: kramer | rschmidt@fzi.de
³ Forschungszentrum Karlsruhe - Technik und Umwelt (FZK) Institut für Angewandte Informatik (IAI), Postfach 3640, D-76021 Karlsruhe, email: geiger | peter | reissfel@iai.fzk.de
⁴ Lehrstuhl für Software-Technologie der Universität Dortmund, Universität Dortmund, D-44221 Dortmund, email: doberkat | sobottka@ls10.cs.uni-dortmund.de
⁵ Landesanstalt für Umweltschutz Baden-Württemberg, Postfach 210752, D-76157 Karlsruhe, email: Andree.Keitel | Wolf.Zitzmann@x400.lfuka.um.bwl.de
⁶ Umweltbundesamt, D-14193 Berlin, Bismarckplatz 1, email: Thomas.Schuetz | Jacqueline.Burkhardt@uba.de
1. Introduction

Reporting to and providing the public with information on the environment is becoming increasingly important for environmental governmental agencies at both the federal and state level (e.g., the German Federal Environmental Agency (Umweltbundesamt) at the federal level and the Environment Protection Agency of the State of Baden-Württemberg (Landesanstalt für Umweltschutz Baden-Württemberg) at the state level). In general, these agencies are required by law to provide this information. In the past the reports were made available as printed documents only. They are accessible in libraries and selected organizations and can be ordered at the respective agency. Due to the limited number of copies (about 6000) the public hardly has the free and unlimited access as intended by the environmental agencies that publish the reports. This situation can be improved with publishing environmental reports not only in printed form but also on electronic media, such as CD-ROM and the Internet. In addition, electronic media provide better means for illustrations (e.g., animations, video clips etc.) which in turn help the public to better understanding the correlations and interrelations between complex environmental topics. Also, with connecting an online report to online environmental databases the public can be provided with up-to-date information on the environment. Finally, the increasing customer orientation of these agencies provides additional motivation for this dissemination. It is worth mentioning that HUDA does neither deal with secret data related to individual persons nor with data which is protected by obligations to maintain secrecy.

Overcoming the limitations of print media, providing the public with easy access to up-to-date environmental information and increasing customer orientation were the three driving key factors for the German Federal Environmental Agency and the Ministry for Environment and Traffic of the State of Baden-Württemberg to launch an ambitious research and development project, referred to as HUDA (Hypermedia Techniques for Environmental Data) (Mayer-Föll et al. 1997; Keitel et al. 1998). The objective of HUDA is the design and development of a toolbox, with which the generation of environmental reports can be made largely automatic. The reports generated with the toolbox should simultaneously be effected in both printed and electronic (i.e., CD-ROM and World-Wide Web, WWW) form and should capitalize on previous results generated in the Environmental Information System of the State of Baden-Württemberg (UIS), part of the research and development effort GLOBUS (Global Environmental Data) (Mayer-Föll et al. 1997).

The remainder of the paper is set out as follows: Section 2 describes the working environment in which the HUDA toolbox is embedded and the four functional modules of the HUDA toolbox. Section 3 focuses on the design of the software architecture. Related work is discussed in section 4 before the paper closes with an outlook and a conclusion in section 5.
2. The HUDA Toolbox

The requirements the HUDA toolbox has to meet were identified by conducting a considered requirements and user analysis (Riekert et al. 1997; Keitel et al. 1998). The analysis also helped to identify the working environment with which editors of environmental reports are familiar. It also leads to the decision to divide the HUDA toolbox into four functional modules which offer appropriate solutions adapted to the problems of automatic report generation. The editors’ working environment and the functional modules are described in that chapter.

2.1 Working Environment

User acceptance is one of the main success criteria for HUDA, thus, users should become familiar with the toolbox after a short training period. Some tools are already available, albeit in different disguises and other contexts. To hide the complexity of the tools and their interaction, and to cater for the users’ needs, one of the major challenges of the project is constructing a uniform facade behind which all the tools reside which either have been in use already or which are developed specifically for the purpose of this project.

With a series of questionnaires (Riekert et al. 1997; Keitel et al. 1998) the working environment in which the HUDA toolbox is to be integrated could be identified. Currently, the editors work on PC platforms under a Windows operating system. Since the operating version of HUDA is supposed to be implemented by the end of 1998, Windows NT 4.0 and Microsoft Office 97 are chosen as implementation platform. The editor teams in the agencies at both the federal and the state level are using QuarkXPress as desktop publishing tool for the generation of reports. QuarkXPress can be extended by XTensions (Qext 1998) which can be plugged into QuarkXPress to provide additional functionality. In addition, standard MS Office tools (Office 1998), such as MS Word, MS Excel etc., and specific tools from outside the Office suite such as Harvard Graphics (graphic tool) and ArcView (desktop GIS) are used in their day-to-day work. At the federal environment agency, the Database „Basic Data for Environmental Report Production“ (Datenbank „Grunddaten zur Umweltberichterstattung“/DB Grunddaten) is already in use for storing report parts. Up to now this database has been filled with report parts once a report was published and has not yet been integrated in the overall process of environmental report production. Finally, the environmental data catalog (Umweltdatenkatalog UDK) (Kramer et al. 1997) was expected to be integrated in the HUDA toolbox.

2.2 Data Translation and Preparation

External institutions deliver many of the report parts (e.g., graphics, tables, maps etc.) to the report editor teams. This causes a high degree of heterogeneity in layout
and in formats which in turn requires many resources for unifying the report parts before they are integrated in a report. The data translation and preparation module assists the editor teams in these time-consuming tasks and provides semi-automatic solutions to enhance the process of layout and format unification. The following functionality is provided by that module:

1. Administration of pools storing all initial report parts (i.e., report parts delivered by external institutions) which can possibly be used in a section of the environmental report. This information serves as input to the preparation process.
2. Information Retrieval functionality to search for and retrieve report parts from the pools.
3. Translation functionality to support users in translating data from one format into another (e.g., translating a table in a Word document to a business graphic in Excel).

An important criterion for user acceptance is the seamless integration of these functionalities into the working environment which is already used for the preparation of the report parts. Therefore the architecture of the data translation and preparation module consists of three layers:

- The **presentation layer** is made up of tools which are familiar to the users of the HUDA toolbox (e.g., Microsoft Word or Harvard Graphics). These tools are extended through services from the service layer.
- The **service layer** provides services supporting the generation of a uniform layout for the tables, graphics and maps.
- The **tool layer** is built on customized tools which are used to automatically generate these elements. These tools are Microsoft Excel for the generation of tables and graphics, Harvard Graphics for the generation of user defined graphics and ArcView for the generation of maps.

### 2.3 Central Data Management System

The central data management system is the heart of the HUDA toolbox (Riekert et al. 1997). It is basically a repository for long-term storage of report parts (e.g., text, graphic, table) used in environmental reports. It has to fulfil the following requirements:

1. To reach a high user acceptance for HUDA the central data management system is supposed to be smoothly integrated in the process of report production.
2. On average, a report is made up of about 1000 report parts which are often used in more than one report. This calls for features supporting search and retrieval of report parts and for the provision of easy to use navigation features.
3. Writing a report is a highly dynamic and iterative process during which the report’s structure changes many times before the final structure is fixed. Hence the data management system has to support the creative process of step-wise designing the structure of a report.

The data management system capitalizes on the DB Grunddaten as an already existing database for report parts of the German Federal Environmental Agency. This database, however, is adapted to the specific needs of the HUDA toolbox. Of special value is an extension which allows the differentiation between reports for different media (i.e., print, CD-ROM and WWW). This extension includes an interface to the module described in Section 2.5 through which report parts generated for an electronic report (e.g., HTML documents) can easily be translated to and stored in the data management system. The seamless integration into the working environment of the end users is also taken into account. For example, for editing users can launch the respective application from within the data management system. Both, the redeployment of an existing tool as well as the full integration of that tool in the working environment lead to a relatively high user acceptance.

For accessing report parts in the data management system both searching and browsing are supported. Traditional information retrieval approaches are used to support searching. Users can browse through the data using a graphical interface displaying the structure of the reports stored in the data management system. The structure in which the report parts are stored in the data management system corresponds exactly to how the report parts are organized in the respective report. This makes browsing more efficient as users can gain insights into the structure of a report easing the access to its components.

A drag&drop mechanism provided at the user interface allows easy restructuring of reports. This is frequently required during the dynamic process of incrementally designing a report’s structure. Still, drag&drop can be used not only during the design process but also represents an intrinsic feature for generating new reports on the basis of already finalized reports. This feature is of particular importance for reusing those report parts of which the content does not change much over time.

### 2.4 Metadata Management System

One requirement is to integrate the UDK as a metadata system for environmental data objects in the HUDA toolbox. The UDK requires metadata of a fairly coarse granularity in order to make reports and their parts easily retrievable on the WWW.

The DB Grunddaten also captures what is called frame information for each report part. The frame information mainly serves the purpose of comparing the content of different report parts to retrieve those parts which address the same issue. This feature is used when several parts are merged and condensed to a new one. Compared to the metadata of the UDK, the frame information is of a much finer gru-
larity. Still, some of the data used in frame information overlaps with the data required for the UDK.

Hence, the metadata management system in HUDA is faced with the challenge to redeploy frame information from the DB Grunddaten for its use in the UDK. This leads to an aggregation strategy the main objective of which is to reuse as much frame information of fine granularity as possible to create high granulated metadata for the UDK. Related approaches are discussed in (Nikolai et al. 1997). The strategy consists of three different sub-strategies: automatic, semi-automatic, and manual aggregation, respectively.

- Automatic metadata creation can be used, if metadata of high granularity can be created without user interaction using metadata of low granularity. For example, if the keywords for a single chapter of a report are ordered according to their priority, the keywords for the whole report can be created by using the keywords of first priority from all chapters, omitting those of lower priority.

- In most cases, however, aggregation is more difficult to achieve, because contextual knowledge must be used to derive proper abstractions. Therefore, the available metadata of lower granularity is shown to the user. Then the user is asked to select or to create the metadata of higher granularity. An example are metadata covering the geographical origin of the data. If a specific environmental condition and, hence, the corresponding type of metadata is found in one particular region of a state only, an automatic algorithm would select only this part of the state as geographical region, although the state as a whole would probably be correct as well. Another example is the creation of keywords for a report that describe the report but do not appear in its text. For example, if there are chapters covering the appearance of lead and cadmium in the soil, a suitable keyword for the report could be „heavy metal“. Thesauri can be used to support this process.

- The manual sub-strategy is applied, if metadata cannot be created by user selection or by means of thesauri. In this case the user has to enter the metadata manually.

The HUDA toolbox supports all three aggregation mechanisms by providing a set of presentation and administration components.

In addition, the online connection of the UDK features several options to search the WWW. Among these are (1) multilingual query modes for searching for UDK objects and addresses, (2) on-line access to the actual data residing on the WWW.

2.5 WWW and CD-ROM Preparation

The current prototype of the HUDA toolbox places the emphasis on the preparation of reports for the WWW and the seamless integration of that functionality into QuarkXPress. The main reason is to gain first experiences in the complexity of electronic report production using reports which are available in QuarkXPress format
only. For the time being, a CD-ROM report can be generated using the static WWW pages.

The module for WWW preparation within the HUDA toolbox consists of two parts: functionality to generate WWW reports (e.g., HTML and GIF files) and the WWW access system.

The WWW variant of the report sections are generated on the basis of QuarkXPress documents using the HexWebXT XTension of QuarkXPress. The generation process can be divided into the following steps:

1. Preparation of the QuarkXPress document according to a given guideline,
2. Conversion of that document into the HTML format,
3. Post-processing of the conversion result.

In order to get a satisfactory conversion result with HexWebXT, the QuarkXPress document to be converted has to be provided in an adequate form. This means that the document has either been prepared in this form or it is translated to this form prior to conversion. The form required is specified in a guideline.

The conversion program then translates the QuarkXPress document into the HTML format. The program selects the document frames (i.e., parts that a QuarkXPress document is made up of) and converts the contents of the frames by using the HexWebXT XTension. Thereafter, the document is decomposed into individual HTML pages in accordance with its sectional structure. Also, a table of contents with hyperlinks pointing to the respective individual HTML pages is generated. Instead of the automatic conversion, a manual conversion using HexWebXT can be performed as well. Furthermore, HexWebXT allows to define links in the QuarkXPress document, which are then converted into the according hyperlinks of the HTML pages.

The post-processing programs provide the central data management system with the information required for storing the HTML files of a report in the DB Grund-daten. Furthermore, the index for full-text search is updated to include the new HTML documents. Finally, certain deficiencies of the conversion result of HexWeb are eliminated. A “final touch” can then be given by manual post-processing.

With browsing and searching facilities, the WWW access system of the HUDA prototype offers different ways for accessing information in WWW report:

- **Browsing:** Users can follow links and browse through the WWW report. A start page with a table of contents can be used as starting point for a navigation through the report. In addition to the table of contents, the start page provides links to information about the HUDA project and to a number of online resources that are related to environmental information.
- **Searching:** A full-text search engine makes information searchable and retrievable in the WWW report. This feature supports users if they are interested in a certain topic but do not know where it is addressed in the WWW report.
3. Software Design for HUDA

The main objective in the design of the software architecture for HUDA was to create an open and flexible toolbox. Openness is not only required to easily add new tools to the toolbox, but also to easily uncouple tools from the toolbox in order to make them available to users who are not mainly concerned with the environmental report production. A high degree of flexibility is required to easily replace existing tools by others or new versions. This section describes the conceptual decisions that were taken to meet these requirements.

3.1 Software Architecture

For the implementation of the software architecture including the four functional modules described in section 2, a prototype-based approach has been chosen. The most important advantage prototype-based methods offer consists in the constructive participation of the user of the HUDA toolbox during early phases of requirements analysis (Doberkat et al. 1989, Doberkat et al. 1997). Through early presentations of implementation results the users are given the opportunity to state and to refine the requirements for each module based on an executable model thereof.

To keep the HUDA toolbox as flexible and as open as possible, the software design differentiates between three functional layers (c.f., Figure 1), a presentation layer, an administration layer and a data access layer.
Users interact at the presentation layer with the HUDA toolbox. All the tools used in the working environment of the editor teams are captured in this layer, including QuarkXPress and the developed XTensions, the MS Office tools, and the HUDA Explorer (a graphic user interface to the DB Grunddaten).

The administration layer provides the functionality for the handling and transformation of report objects. This allows different components at the presentation layer (e.g., QuarkXPress and the graphical user interface of the central data management system) to use the same administrative functionality at the administration layer.

The third layer of the HUDA software architecture is the data access layer. Currently, it provides transparent access to the DB Grunddaten and the UDK but further database systems can easily integrated in later phases of the project.

With this architecture a high degree of openness is attained: New tools can be added at the presentation level without affecting other tools. There are two options to add administrative functionality at the administrative level. One of which is the reuse of already existing administrative functionality by connecting it to the corresponding new components at the presentation level. The other option is to add completely new administrative functionality. In addition, tools that are beneficial to users who are not mainly concerned with environmental report production can be provided with an interface that is not an integrated part of HUDA but that provides access to functionality at the administrative level. For example, web administrators can already access the post-processing programs developed for the HTML production of reports (c.f., section 2.5) from outside the HUDA toolbox. This enables them to use these programmes for their day-to-day tasks (e.g., splitting of a HTML document in several linked HTML documents). Another example is the new graphical user interface of the central data management system which is loosely coupled to the HUDA toolbox. This allows database users to access the DB Grunddaten from outside the HUDA context (c.f., section 2.3).

The openness of the HUDA software architecture is also the key to a better reuse of the HUDA software components. They can easily be used in other contexts and thereby help to achieve cost benefits by reducing the amount of time which is necessary to implement and test applications.

Furthermore, the separation of the technical aspects in the HUDA software architecture increases the evolution capabilities (Kiczales 1996). Components can easily be replaced by new ones, offering an improved or enhanced functionality. This flexibility is guaranteed through the distribution of the specific functionality among the three layers (i.e., user interfaces at the presentation layer, administrative functionality at the administration layer and database access functionality at the data access layer). For example, removing or replacing a tool at the presentation level does not necessarily affect the administrational functionality of that tool.
3.2 Technical Basis

Technically, the three layer architecture is implemented as follows: The Component Object Model (COM 1998), ActiveX (Chappell 1996) and OLE (Open Linking and Embedding) (Brockschmidt 1995) build the technical basis for the HUDA toolbox. Both, COM and OLE are widely used in the MS Office suite. Visual Basic provides a comfortable support for the creation of ActiveX Controls and ActiveX Servers. The presentation layer is implemented as a collection of ActiveX Controls (Chappell 1996). The administration layer is implemented as ActiveX Server with the respective components which can be accessed by OLE-automation. Finally, the components of the data access layer are implemented using Data Access Objects (DAOs 1998). They use an Open Database Connectivity (ODBC 1998) driver to access heterogeneous data.

4. Related Work

In this section, we relate our work to other approaches in the field of environment report systems and to other technically related approaches that we deem to be important.

4.1 EcoMediaToolBox

A project in the context of the toolbox HUDA is the EcoMediaToolBox which had been designed at the University of Dortmund (Döbmann at al. 1995) and which served as one of the conceptual starting points to the entire HUDA effort. This toolbox is a set of tools for the creation of possibly different issues of environmental reports which are published as paper reports, as a CD-ROM or in the WWW. Breaking the system down into subsystems reflects the distinction of the characteristics of information packages: the content, the connection to other packages and the presentation of the package to the different users of the information provided by the content.

The essential difference between the toolbox HUDA and the EcoMediaToolBox is the way of constructing the reports for the three types of Media: in the HUDA toolbox the report for the print media serves as the only input for the generation of the electronic reports. The tools integrated in the EcoMediaToolBox generate the report for the print media from a database which in turn acts as a repository and is the heart of the EcoMediaToolBox.

4.2 DEAH

Another project which is related to HUDA is the Digital Environmental Atlas of Hamburg (DEAH) (Greve et al. 1995; Blasius et al. 1997). The DEAH is a Point-Of-
Information System providing information on the state and quality of the environment in Hamburg. HUDA and DEAH are similar in many ways: both tools are flexible and extendable, they help editors to publish their data in electronic media, they support retrieval features and possibilities for continuing report production, and they provide the public with free access to information on the environment.

The main difference is on the technical side. Technically, DEAH uses Lotus Notes as document management system. The Lotus Domino Server provides the connection to the WWW. The analysis of document management systems carried out in early phases of HUDA, however, turned out, that Lotus Notes could not fulfil the requirements of seamless integration of the HUDA toolbox in the working environment of the end users. Lotus Notes was developed by Lotus. Many of the Lotus tools, however, are not compliant to the tools used in the working environment of the HUDA users. Another difference between Lotus and HUDA is the component-based architecture of HUDA which allows users to use specific components even outside the HUDA toolbox (c.f., section 3.1).

4.3 DEAB

The Digital Environmental Atlas of Berlin (DEAB 1998) also provides free access to environmental information on the WWW. The differences to HUDA are, (1) the WWW version of that atlas was capitalized on an existing CD-ROM version; print media is not taken into account and (2) a special emphasis is placed on maps and GIS functionalities. The latter, however, will be addressed in phase III of HUDA.

4.4 HyperWave

With HyperWave, a second generation Internet information system was examined in the course of HUDA. In HyperWave, documents can be organized hierarchically and can be stored along with meta-information in an object-oriented database. The search capabilities provided by HyperWave are more powerful than those offered in HUDA. By contrast to the central data management system of HUDA, a full text search is supported. With HyperWave it is also possible that the results of a search query can be defined to be the search space for further queries.

Among others, HyperWave dropped out for the following reasons: (1) It is not compliant to the working environment of the HUDA users; (2) unlike the frame information concept in DB Grunddaten and the meta-information concept in UDK, meta-information in HyperWave does not cover attributes which are required for describing environmental data (e.g., geographical references).
4.5 Other approaches

Finally, many of the experiences made in the GLOBUS effort had remarkable impact on HUDA (Mayer-Föll at al. 1997): Converting documents to HTML was already done in the AlfaWeb project (Geiger at al. 1997), in the UVIS project (formerly UFIS) (Henning at al. 1997, Wiest et al. 1997), environmental databases are made accessible through a WWW interface, one of the partners extended the UDK to the WWW-UDK and, thus, gained experiences in accessing environmental meta-information on the WWW (Kramer et al., 1997).

Although we could take advantage of the insights and experiences we gained in these projects, the focus and objective of HUDA required the design and implementation of new technical solutions specifically adapted to the requirements of HUDA users.

5. Outlook and Conclusions

The current prototype has been installed at the two environmental agencies so the editor teams can gain first experiences with the HUDA prototype. Their feedback will feed into a requirement analysis for phase III of HUDA. The goal of phase III is to improve the prototype to a fully operative system. Among others, the following features will be integrated in the operative system:

- Extension of the HUDA toolbox to a multi-user system so that several users can work on environmental reports at the same time.
- Tighter coupling of QuarkXPress and the central data management system (DB Grunddaten) for a better integration of the central data management system in the report production process.
- Sealing of reports protects them from changes after they were agreed and published.
- Unified user interfaces for all of the modules with an integrated context sensitive online help to reach a high user acceptance.
- Integration of more sophisticated viewers to improve the use of the electronic reports (e.g., tools for displaying maps or complex graphics).
- Provision of online access to environmental databases from the electronic reports.

As pointed out in section 3, a high degree of openness is a major success criteria for HUDA. For that reason, those functionalities of the HUDA toolbox that can be used outside of HUDA will be provided with an user interface which is not linked to the HUDA toolbox. This broadens the potential user community of the system since parts of it can even be used by users who are not involved in the environmental report production.
HUDA acts as an enabling technology for the economic exploitation and redeployment of existing data as well as for the information and communication supported environment report generation. Furthermore, HUDA supports the simultaneous generation of reports in three different types of media. The cost/benefit ratio for environmental report generation with HUDA is notably more favorable compared to the traditional environmental report generation for print media only. The electronic reports can be produced at a low cost which makes it possible to produce a larger number of reports without increasing the overall expenses.

The benefit for the public is that these reports become available more easily which in turn contributes to the building of a better environmental awareness among the public. Environmental awareness is an important contribution which the rapid transition to an information society can make to sustainable development.

The copyrights for HUDA are owned by the German Federal Environmental Agency and the Ministry for Environment and Traffic, Baden-Württemberg. After phase III of the HUDA effort, interested and new partners can get the right to use the HUDA toolbox if they are willing to provide an adequate contribution to further developments of the HUDA toolbox.

**Literature**


Band A - Abschlußbericht und Band B - Anlagen.


http://www.iai.fzk.de/~weideman/doc/globus4/.


