Customization in Environmental Information Systems

David L. Hicks[1], Klaus Tochtermann[2], Andreas Kussmaul[2] and Sandro Neils[3]

Abstract

The world of information today is an increasingly diverse, distributed, and fragmented one. As the amount of information available in digital form continues to increase, so does the number of systems that are employed to make that information available. Often environmental information systems provide functionality which is based on metadata, i.e. data describing data, in order to support users in searching for and retrieving data required to perform a specific task. Support for customization and personalisation of these metadata according to the personal needs of a user can further increase the potential benefit and effectiveness of environmental information systems. Against this background, the paper introduces a customization architecture which allows users to customize environmental information systems according to their personal needs. To better illustrate how such systems can benefit from customization, the paper also presents the status quo of our prototype implementation.

1 Introduction

There is an increasing need in companies and public administrations to store, provide and manage environmental information using modern information and communication technologies. Since the amount of environmental information to be managed is growing rapidly, environmental information systems are becoming an indispensable prerequisite for both companies and public administrations to manage efficiently their environmental information.

Today's environmental information systems provide users with easy access to a variety of environmental information, including satellite photos, cartographic information, the results of empirical observations, or environmental laws, reports and

---

[1] Aalborg University Esbjerg, Computer Science Department, Niels Bohrs Vej 8,
6700 Esbjerg, Denmark, email: hicks@cs.aue.auc.dk
2 Forschungsinstitut für anwendungsorientierte Wissensverarbeitung, Postfach 2060,
D-89081 Ulm, email: {tochterm | kussmaul}@faw.uni-ulm.de
3 Universität Magdeburg, email: neils@cs.uni-magdeburg.de

studies. In order to provide effective access to this type of information, often environmental information systems do not only store the environmental information itself but also information about the environmental information (metadata). Metadata can be used for describing many different facets of environmental information such as where a data item is located, the area covered by a satellite image, and the time period covered by a series of measurements. Metadata supports an access strategy in which the user can first retrieve the metadata for an information object, and then later retrieve the larger information object only if it is of interest, based on the metadata description.

However, all systems examined share one drawback: the lack of capabilities that allow users to customize the systems and their content to their personal needs. The next section (Section 2) underpins this observation by two examples of existing environmental information systems. Section 3 then proposes an architecture that supports customization of environmental information systems without affecting the systems themselves or the information stored therein. Section 4 closes the paper with a brief description of the current prototype implementation and an outlook on future work.

2 The need for support of customization

This section provides two examples to stress the needs to better support customization in environmental information systems and related digital catalogue systems.

2.1 Environmental report production systems

There is an increasing need to better support the largely automatic production of environmental reports for both printed and electronic media (Tochtermann et al. 1998, Lurk/Alber 1998). Many of the report parts (e.g. tables, graphics) of an environmental report are produced and delivered by different external institutions. This causes a high degree of heterogeneity in layout which requires many resources for unifying the report parts before they are integrated in a report. The solution provided by existing environmental report production systems is a layout translation tool which assists the editor teams to enhance the process of layout unification. Typically these tools comprise the following components: Repositories storing all original report parts (i.e. report parts delivered by external institutions) which can possibly be used in an environmental report; information retrieval functionality to search and retrieve original report parts from the repositories; translation functionality to convert data from one layout into another and repositories to store the converted report parts.

There are two disadvantages of this approach: Firstly, since the layout and the content of report parts are tightly connected to each other changes in layout always affect the original report parts; due to this environmental report production systems often store both the changed and the original report parts. Secondly, even slight changes in the layout which often arise in later phases during the report production require that all report parts must be adapted again. This situation could be improved if a) layout definitions are stored separately from the original report parts and b) the application of the layout definitions does not affect the original report parts.

2.2 Environmental digital catalogue systems

There exist many digital catalogs for environmental data. Some of these are independent of any environmental information system (UDK 1998, Tochtermann et al. 1997), others are an integral part of an environmental information system (DBG 1999). In either case catalogs are used to facilitate the identification, location, access, and use of items referenced in the catalogue. To assist users in the identification and selection of relevant information, each item in a digital catalogue is described by up to 40 different fields (DBG 1999) such as temporal coverage, geographical relationships, etc. Several strategies exist for cataloguing the environmental information, each of which impacts the result of a search operation. In addition, different users often have different expectations of a search result yielded by a specific catalogue query. In the case of cataloging geographic relationships, one strategy is to categorize resources into the most detailed geographic category possible. For example, if a resource is related to Munich, Munich is entered for the geographic relationship for it even though Germany or even Europe would have also been valid. When a user performs a search for environmental information related to Europe, the resources catalogued as related to Munich will not be part of the search results (unless a geographical information system was part of the digital catalogue system). For some users this might be surprising since Munich is part of Europe and would therefore be expected to be part of the search results. To overcome this situation users should be provided with functionality which allows them to define their own individual policies for assigning geographical relationships without the need to change the original metadata fields in the digital catalogue. In addition, often large number of metadata fields used to catalogue resources is relevant to the process of identifying related information in the digital catalogue. Experts would prefer to limit the metadata fields they work with to those that are helpful in finding relevant information. This leads to the requirement to allow users to customize the metadata records from the database so they include only the information they need to perform their job efficiently. Finally, users should have the possibility to further customize metadata records from the digital catalogue by adding new fields to the records to hold expert knowledge they have about the records but which was not available during the initial cataloguing of the resources.

3 Overview of the customization architecture

The ability for users to customize their information space, to personalize the information resources with which they work, is an important capability, one that is valuable in the performance of complex tasks (Van House 1995). In providing customization capabilities, system developers cannot simply allow users to modify a data object directly since this would change the data object for everyone. The resultant situation would be one of chaos rather than effective support for users of environmental information systems. We therefore suggest the use of a metadata based approach (Hicks et al. 1999). An important characteristic of metadata is that it can exist and be maintained for a data object completely independent of the data object itself. For example, consider a digital catalogue system. The information described in a digital catalogue might exist locally (with respect to the catalogue), it could be managed by a remote system of some kind, or indeed it might not even be available in digital form. In any case, it is possible for the items contained in the catalogue to be defined and maintained separately from the information they refer to. The customization architecture exploits this characteristic of metadata to enable the customization of data objects in a way that places no restrictions on where data objects are stored or by which system they are managed. Similarly, write or update access to the data objects being customized is not assumed or required.

To achieve this, a strategy is employed within the architecture in which metadata is used to represent customizations that are made to data objects. Instead of using metadata in its more traditional role, to create universal and widely applicable descriptions of objects, such as in digital catalogue systems, it is instead used to support much finer grained individualized descriptions of data objects. These adaptations to the conventional use of metadata enable the approach described here to support the individual user level customization and personalization of a data object, independent of the data object itself.

Another important characteristic of the approach described in this paper for supporting the data object customization process is that it is an architectural one. Because support for customization is important for all types of data objects, the customization capabilities are provided in a way that allows them to be utilized by a range of different applications, such as digital catalogue systems, environmental report production systems and environmental databases. The customization architecture is illustrated in figure 1.

Within the architecture, client applications can utilize the customization functionality provided by the customization metadata manager (CMDM) to support the data object customization process. The CMDM is a server process that, in response to client application requests, retrieves data objects and supports metadata-based customization operations on those objects. As illustrated in figure 1, the CMDM maintains a store of customization metadata that applies to the objects that have been

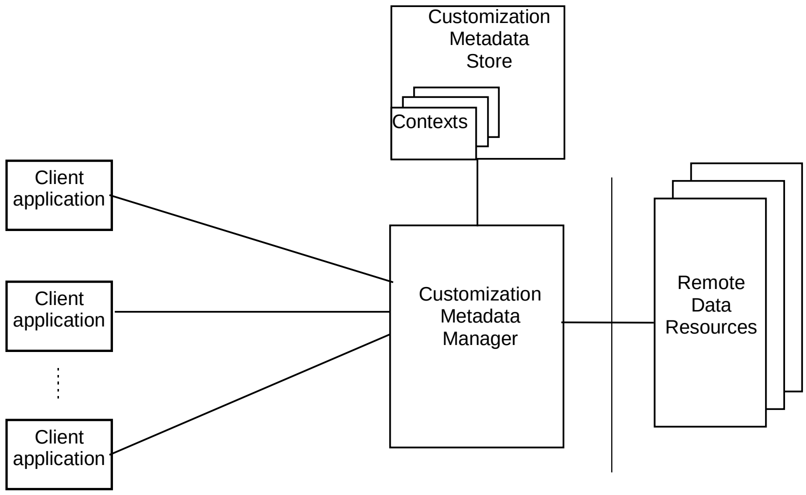

Figure 1
Customization Architecture

customized by client application users. It is important to note that the customization metadata store does not contain actual data objects themselves. The data objects continue to be managed by the system where they were originally located. Only the metadata information that represents changes that have been made to data objects as part of the customization process are contained in the metadata store.

The object customizations supported within the architecture can be organized into two major categories. The first category of customization applies to those objects that internally are organized into a series of fields or attributes. Examples of this type of object include a description item from a digital catalogue system, an HTML file in which tags are used to partition the internal contents of the file, or a database record consisting of one or more fields of information. This category of customization within the architecture provides a way for a metadata field to be defined for an object and assigned a value that will be used to override or replace a specific field, attribute value, or other part of the structured object whenever the object is requested from the CMDM. In order to provide this functionality, the CMDM must have knowledge about the structure of the particular type of the data object being customized. An example customization of this type could be a metadata item created for a digital catalogue item. This metadata item is used to override a particular filed (e.g., the geographical relationship) whenever it is retrieved by the CMDM. As described earlier, in either case the changes defined for the data object are stored in the customization metadata store; the data object itself is not updated.

The second category of customization is a more general one and is applicable to all object types, regardless of whether they are internally structured or not. Report parts, such as tables and graphics of environmental reports fall into this category. It corresponds to the case where a metadata field is created for an object, but the metadata field is not intended to replace or override any corresponding existing field or attribute value of the object to which it applies. Instead, the metadata field is being used to contain some auxiliary or supplemental information (e.g., a layout definition) about the object that will be used to support the customization process. No specific details or knowledge about the internal arrangement of the object are required to support this type of customization. A metadata field created for a graphic file might represent a customization of this type. The metadata field might be used to contain a layout definition that the user needs for the integration of the graphic in an environmental report. The layout definition doesn't replace or override any part of the graphic file it pertains to. Rather, it simply provides supplemental information about the data object.

Applications utilizing the functionality of the architecture are free to interact with the CMDM to create metadata records as necessary. No limits on or predefinitions for customization metadata records exist within the architecture. As many metadata records as necessary from either of the categories described above can be created to support the customization of data objects.

4 Prototype implementation and Outlook

A prototype implementation of the customization architecture is currently under development. The functionality provided by the CMDM prototype will be used and tested within the context of an environmental database (DBG 1999). The client's interface is implemented in HTML. The client connects with http to a standard web server which manages several Java Servlets each of which provides specific functionality (e.g., for connecting to the data management system or for connecting to the CMDM store). The CMDM store is implemented using Microsoft Access. The communication between the CMDM store and the Java Servlets are based on JDBC. The following three figures illustrate the prototype implementation of PADDLE.

The Field Manager (c.f. figure 2) allows the customization of the metadata fields for objects from the remote data resources. Figure 2 displays the metadata fields for the remote data object „Waste Management“. On the left several buttons are provided: The „Add Field“ button is used to add a new metadata field to a given remote data object. The „Remove Field“ button is used to remove a metadata field which was created by the „Add Field“ button. Note, metadata fields which are defined in the remote data resources cannot be removed, because this would affect the orginal system itself. Instead, users can hide „original“ metadata fields. For example, in figure 2 the user wants to hide the metadata field „Abstract“ which means that the value for this field will not be shown if the corresponding data object matches a query.

Hiding of metadata fields is an important instrument to reduce the information to a level which is sufficient for a user to perform his task effectively.

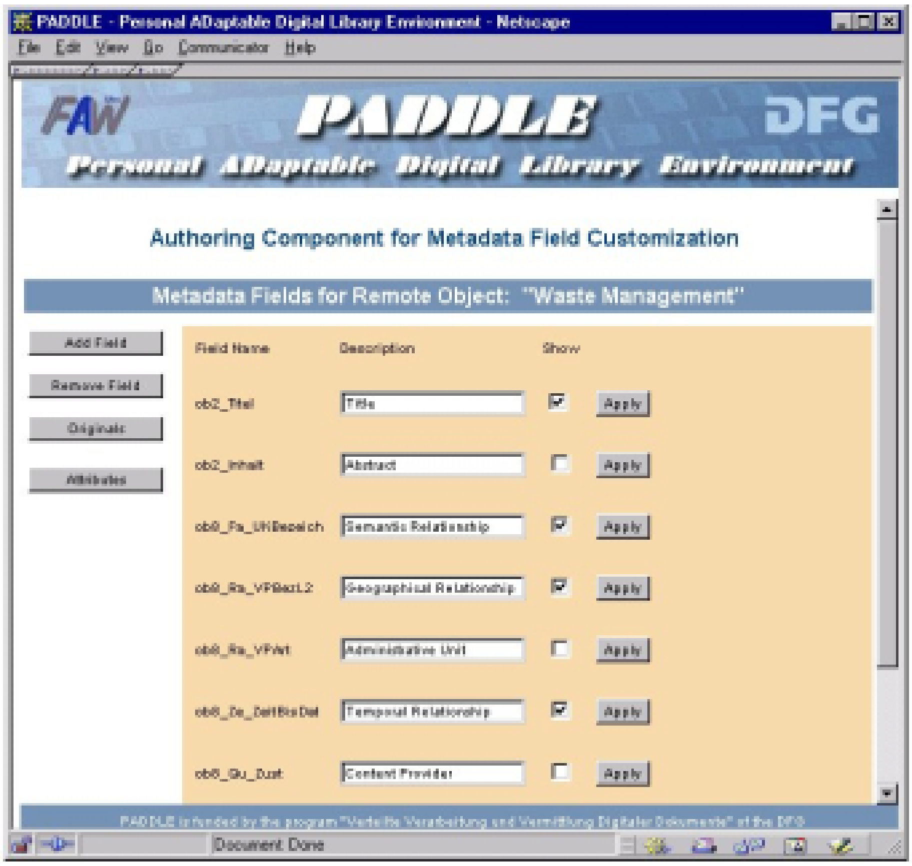

Figure 2
Field Manager

On the right in figure 2 one can see the field names as they are used in the remote data resource and the speaking names as they are defined by a user. For example the technical field name *ob2-Titel* is associated with the speaking name *Title*. This allows users to define their own names for the different metadata fields.

Similar to the form displayed in figure 2, another form exists which can be used to design query forms according to the metadata fields customized by the Field Manager.

Figure 3 displays the form which is used to customize the values of metadata fields for a given remote data object. Those values which are changed are indicated with a preceding *changed*. For example, the value of the field *Semantic Relationship* has been changed by a user. To change a value users can simply enter the new value and press the „Apply“ button. Also, metadata fields which were added using the „Field Manager“ are highlighted with a preceding *new* (e.g. *Temporal Relationship* is a new metadata field which does not exist in the original system).

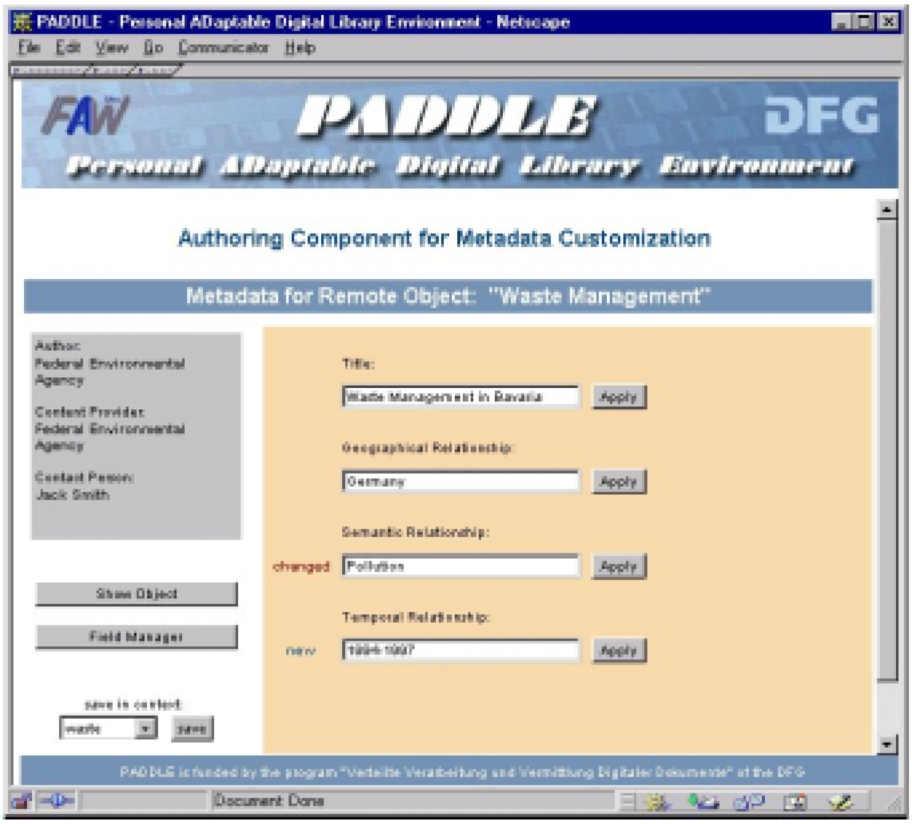

Figure 3
Authoring Component for Metadata Customization

All the changes made with the authoring component are to be stored in a context (see select box *save in context* on the left). This allows users to apply different customizations to the same data object. The upper part on the left displays those metadata fields which cannot be customized. For example, the A*uthor* or *Content Provider* of a remote data resource should always be the same.

Figure 4 displays the search component of the PADDLE prototype system. On the left, a user can choose a context using the „Context Menu“ button. The search will then be restricted to the customized metadata which belongs to the selected context.

On the right, the values for the metadata fields can be entered. The result is a list of links to documents which matched the query along with a brief abstract. The documents can be downloaded from the remote data resource by selecting the corresponding links.

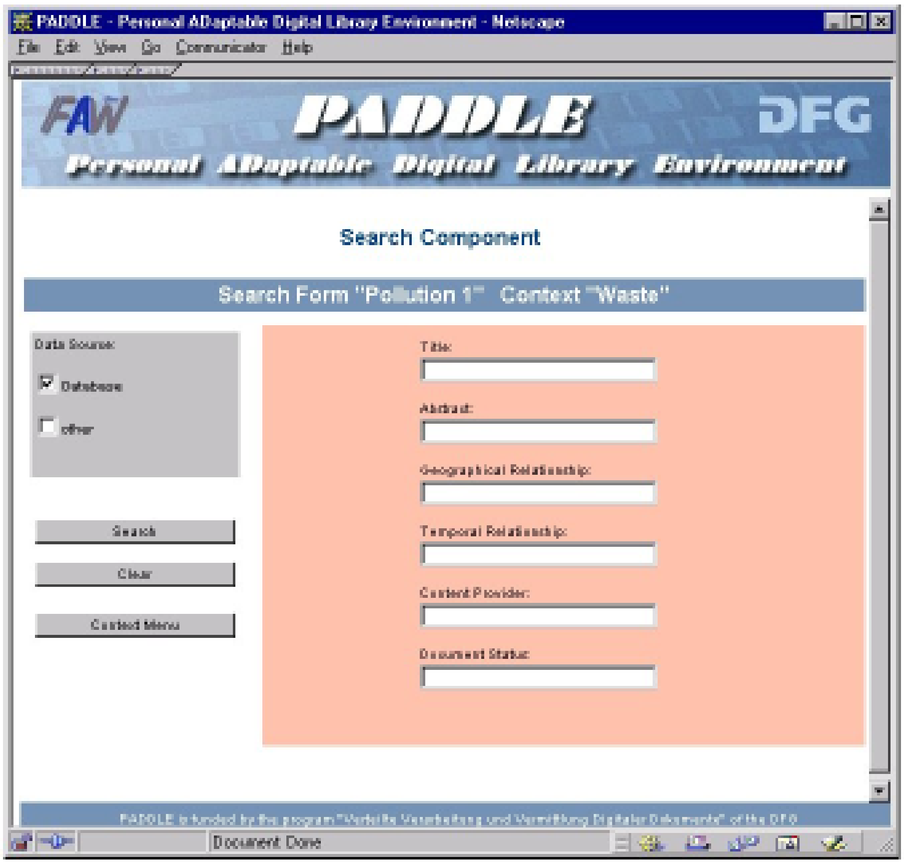

Figure 4
Search Component

With the current prototype it is only possible to customize metadata. In the near future we want to extend this concept to customize the information objects themselves (e.g., define a personal layout for text documents or define personal colors to be used in graphics). To follow this line a remote data resource containing XML documents will be connected to the PADDLE prototype. The advantage of XML is its strict separation of structure and layout. Therefore, the personal layout of an XML document can be defined in a corresponding XSL description. Another issue to be addressed in our future work is that currently any changes of the remote data resources will not be noticed by the CMDM. This can lead to a situation in which the CMDM tries to retrieve a data object which does not exist in the remote data resource any more. In addition, management concepts for several remote databases, the use of different protocols (e.g., Z39.50) and distributed search mechanisms in the remote databases will be part of our future research.

5 Acknowledgments

The research at FAW was conducted as part of the PADDLE project (Personalizable Adaptable Digital Library Environments) funded by the program "Verteilte Verarbeitung und Vermittlung Digitaler Dokumente" of the Deutsche Forschungsgemeinschaft (DFG).

Literature

DBG (1999): Dokumentation der Datenbank Grunddaten Version 3.1., Berlin

Hicks, D.L., Tochtermann, K., Rose, T., Eich, S. (1999): Using Metadata to Support Customization, in: Proceedings of the Third IEEE Metadata Conference, April 6-7 1999, Bethesda/MD, http://computer.org/conferen/proceed/meta/1999/papers/19/dhicks.html

Lurk, A., Alber, C. (1998): Entwicklung eines allgemeinen Systems zur teamorientierten Erstellung hypermedialer Umweltberichte, in: Tagungsband des 1. GI-Workshops "Hypermedia im Umweltschutz" in Ulm, Marburg

Tochtermann, K., Riekert, W.-F., Kadric, L., Kramer, R., Schmidt, R., Geiger, W., Peter, N., Reissfelder, M., Doberkat, E.-E., Sobottka, H.-G., Keitel, A., Zitzmann, W., Schuetz, T., Burkhardt, J. (1998): HUDA: A toolbox for environmental report production, in: Tagungsband des 12. Internationalen Symposiums "Informatik im Umweltschutz" in Bremen, Marburg

Tochtermann, K., Riekert, W.-F., Wiest, G., Seggelke, J., Mohaupt-Jahr, B. (1997): Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs, in: LNCS 1324, Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries, Pisa, September 1997, S. 73-86

UDK (1998): http://www.uis-extern.um.bwl.de/www-udk/RPT_DATA/about_d.html

Van House, N.A. (1995): User Needs Assessment and Evaluation for the UC Berkeley Electronic Environmental Library Project: a Preliminary Report, in: Proceedings of ACM Digital Libraries'95, Austin/TX, July 23-26 1995, pp. 71-76