

Integration of the Authorities' Environmental Information in the State of Baden-Wuerttemberg

Thorsten Schlachter¹, Renate Ebel², Werner Geiger¹, Annette Sawade³, Martina Tauber² and Rainer Weidemann¹

¹ Forschungszentrum Karlsruhe, Institut für Angewandte Informatik,
Postfach 3640, D-76021 Karlsruhe, Germany
Thorsten.Schlachter | Werner.Geiger |
Rainer.Weidemann@iai.fzk.de

² Landesanstalt für Umweltschutz Baden-Württemberg,
Postfach 21 07 52, D-76157 Karlsruhe, Germany
Renate.Ebel | Martina.Tauber@lfuka.lfu.bwl.de

³ Ministerium für Umwelt und Verkehr Baden-Württemberg,
Postfach 103439, D-70029 Stuttgart, Germany
Annette.Sawade@uvm.bwl.de

Abstract. With the citizens being entitled to be provided with environmental information, the quantity of the latter increased as did the efforts needed to find the desired information on the many distributed web sites. The Environmental Information Network of Baden-Wuerttemberg (UINBW) presented here shall serve as a central access platform that facilitates search by offering a thematically structured approach and various search options to the user.¹

1 Motivation

In Germany, supply of environmental information belongs to the obligations of public administration. According to the Environmental Information Act (UIG) [7], citizens already are entitled in principle to obtain the environmental information available at an authority dealing with environmental tasks. For this reason, many authorities provide the corresponding information offers in the worldwide web. Often, these developments have been made by the individual authorities in their own responsibility and are not embedded in a larger context. The revised version of the UIG makes an active supply of information obligatory for public authorities. Thus further enhance the role of the internet as a means for active provision of environmental information.

Many of these authorities' web sites offer rudimentary access to the data only. Frequently, a full-text search cannot be found and no meta data, e.g. keywords, are added to the usable contents. Links to related offers are often lacking to a large or complete extent. The contents have often been processed in line with the authority's structure, but not according to criteria that seem logical to the user.

¹ This paper is a revision of the paper given in [8].

For the citizen, this means that the information searched for is often found only when the authorities and their structure are known in detail. Even internet search machines are of no great help, as the large number of hits hides the information searched for.

With the project German environmental information network gein© [3] in 2000 an attempt was made to establish an environmental information portal on the federal level which offers search functions over information provided by federal and state authorities. Based on the model of gein© such an environmental portal also is considered a reasonable approach on the regional level for the state of Baden-Wuerttemberg.

When the environmental information offers of the state of Baden-Wuerttemberg and its subordinate authorities were listed in 2003, roundabout 100 relevant offers were found on the internet and about the same number on the intranet of the state administration. Most of this information was accessible under an own domain or subdomain.

For the selection among these offers, the "Landesanstalt für Umweltschutz" (LfU, State Authority for Environmental Protection) is presently providing an internet portal named "Portal Umwelt" (Environment Portal). It comprises the offers of the Ministerium für Umwelt und Verkehr (UVM, Ministry of the Environment and Transport) and in particular those of the LfU itself, together with a full-text search and a search of the entries of the "Umweltdatenkatalog Baden-Württemberg" (UDK, Environmental Data Catalogue Baden-Wuerttemberg). The portal is supported by hand directly on the respective HTML pages. As also the meta data of the UDK are maintained manually, discrepancies in the scope of the reached offers result. Congruence of these offers and meta data with the search function and the catalogue of environmental issues is envisaged.

2 Requirements and Basic Concepts

Further development of this portal of the environmental information system of Baden-Wuerttemberg (UIS) to an environmental information network of Baden-Wuerttemberg (UINBW) is aimed at improving the networking of the distributed environmentally relevant web offers of the state of Baden-Wuerttemberg. The user shall be given comfortable access from a central point.

Meta data on all information offers are compiled centrally by the editorial staff. This database represents the starting point for the operation of a portal. The data are stored persistently using a content management system (CMS) and updated in this system via a WWW interface. The CMS provides interfaces to other partial systems, e.g. full-text search and later automatic keyword search. Moreover, its templates allow for the presentation of the data, the layout, and the generation of navigation.

A major constraint of the operation of such a portal is that no or only a minimum expenditure is needed for the maintenance of the web sites referenced therein. Thus, the portal will meet with the acceptance of the web site operators.

Although the expenditure required for the integration of information offers in the UINBW shall be minimized, individual interfaces have to be generated for certain

information systems. This especially applies to offers that are generated dynamically, as they e.g. query statistical data or measurement values from databases. Also, the expenditure needed for the development of these interfaces shall be minimized.

The users of the UINBW are offered several accesses to the individual information offers, a thematic approach, full-text search, and other specialized access options.

According to the requirements outlined in the Act on Equal Opportunities of Handicapped Persons, the entire presentation is tailored to barrier-free access, i.e. the layout and structure of the contents offered are implemented largely semantically in HTML, while the layout is described by means of style sheets.

2.1 Data Collection

To obtain an overview of the current offers of environmental information, the inventory was taken.

For this purpose, the information offers already contained in the existing Environment Portal were selected first. In addition, the sites of authorities were checked for further relevant contents. Moreover, it was specifically searched for offers from the environmental field using full-text search machines.

Due to the large variety of small and smallest offers from the municipal sector, it was decided to consider them exemplarily first. This means that the offers of Karlsruhe and Stuttgart were taken into account, while decision on the integration of other municipal offers was postponed to the time of routine operation.

Analysis of all offers resulted in a list of about 100 information offers on the internet and about 80 offers made on the intranet. Depending on their relevance to the UINBW, these individual offers were assigned a priority between 1 and 3. Offers of first priority shall be integrated in the UINBW immediately, while the offers of second priority will be integrated later and offers of third priority will be accounted for by a link only.

To check the expenditure needed for integration in the UINBW, the web sites were analyzed for their technical parameters and contents. For this purpose, meta data of the individual offers were collected and the technical fundamentals and prerequisites were settled by questionnaires and interviews of the operators. The information gathered was incorporated into a database and now serves as contents base for the UINBW meta data storage.

2.2 Qualitative Evaluation

First qualitative results were obtained by questioning the persons responsible for the selected web offers on the internet. Here, the results shall not be presented quantitatively, as most of the offers studied come from a few authorities only. This gives rise to accumulations that distort the results.

On most web sites, no CMS are being used at present. On the remaining ones, various CMS products are applied, some are own developments.

Most systems are provided with an own full-text search. Here, various products are applied. Use of the existing index files of search machines of individual web sites for the search machine of the UINBW mostly is difficult or even impossible for technical and safety reasons. However, practically nothing speaks against the use of a search machine with a crawler. Nearly all persons interviewed have agreed to such indexing without any reservations. Technical problems will arise in a few cases only.

A search for keywords assigned by the editorial staff is impossible, as hardly any systems are enriched with respective meta data.

Several operators had difficulties in grouping the contents into a scheme corresponding to the gein© portal. A strictly hierarchical structure appears to be too rigid, as many issues are treated "marginally" only.

At the moment, practically no system is barrier-free. For coming updates of some systems, implementation of a barrier-free structure is currently considered.

Some systems will be modified to a large extent within the next year. Partly, CMS shall be introduced.

3 System Architecture and Components

The UINBW system architecture consists of various individual components. Use of a maximum of standard components is envisaged. These components include:

- Central data storage
- Maintenance interface for administrators and editorial staff
- Presentation component
- Data interface for external components
- Search machine for full-text search
- Search machine for keyword search

As demonstrated by an investigation and the prototype implementation of the core function, a CMS is particularly suited for the implementation of some of these tasks. With this, a first approach to using a database and manually implementing the software for these components was abandoned.

Now, the CMS is the central component of the architecture. With its back-end database, it provides for the persistent storage of data and offers interfaces for administrators, editorial staff, and the users of the UINBW. In addition, the CMS supplies the necessary data to other components and in particular full-text search (VTS) via a specially implemented interface. Later on, this interface shall also be used for automatic keyword search (SNS).

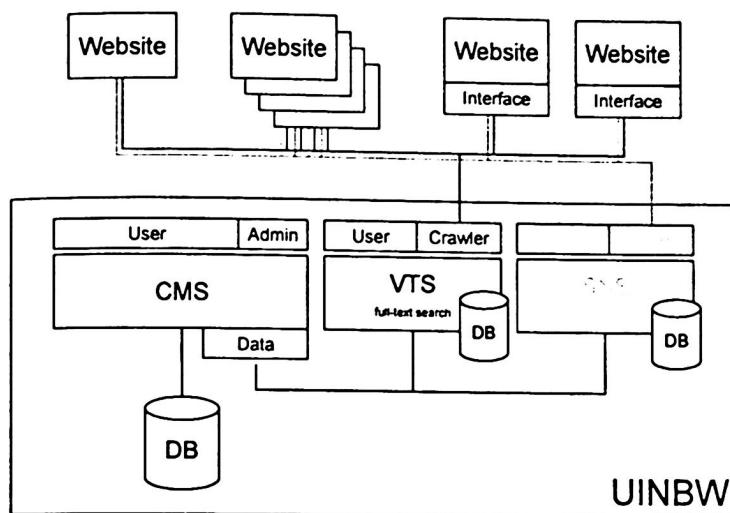


Fig. 1. System architecture of the UINBW. Components and interfaces

3.1 Content Management System (CMS)

The meta data are managed by a content management system (CMS). Thus, many functional parts of such a system can be used for the UINBW. The most important function is the storage of the necessary data in the CMS or its back-end database. The maintenance interface for administrators and the editorial staff may easily be mapped via the workflow support and the CMS authorization system.

The data interface for external components may be implemented by making use of the programmability and extendability of the CMS. Presentation of the contents as well as the automatic generation of navigation and menus for the portal may be accomplished using the template-based presentation mechanism of the CMS. This technology also facilitates the implementation of a barrier-free web presentation [2].

3.2 Thematic Approach and Navigation

By a thematic approach, the variety of offers is limited by consistent grouping to certain selected environmental issues, such as "soil", "water", "nature protection", etc. Experience gained from the use of the gein portal shows that a lean and flat structure complies most with the wishes of the user.

Administration of the environmental issues also takes place by means of the CMS. The environmental issues are defined as an own object class and may be linked with other contents of the CMS via relations. On the basis of these data, the

corresponding templates generate corresponding menus and, thus, navigation in the UINBW.

Each of the offers integrated in the UINBW is assigned to few environmental issues of high priority by relations. Moreover, it may also be assigned to other issues of lower priority. This assignment then is reflected by the order of the presentation for a certain environmental issue. For later extensions, these issues may be refined.

3.3 Full-text Search

A full-text search unlocks all web sites connected to the UINBW. Via a corresponding data interface, it uses the meta data stored in the content management system and the data are also made available to further components and possible extensions.

Indexing of the individual web sites takes place by means of a "crawler" that indexes complete web sites in a fully automatic manner based on the references contained. For individual web sites, adaptations to this type of full-text indexing have to be made and the respective interfaces have to be generated.

Full-text search allows to search either all or only part of the indexed web sites. The user has the possibility of limiting full-text search to such web sites that are assigned to one or several of the above-mentioned environmental issues. This thematic limitation of search means a major progress as compared to conventional search of all web sites, which may result in a number of irrelevant hits.

3.4 Semantic Network Services (SNS)

In the future, the contents of the web offers integrated in the UINBW shall no longer be opened up by full-text search alone, but also via a keyword search. In this connection, applicability of the semantic network services [1] developed on behalf of the Umweltbundesamt (Federal Environmental Authority) and used in the *gein*© portal shall be checked. These semantic network services offer a fully automatic keywording of WWW sites under semantic integration of an environmental thesaurus, geographical names, and chronology. Ambiguities are resolved by a context analysis. Keywords are weighed with respect to their significance to a special document.

3.5 Other Access Options for Users

Some environmental information items are updated regularly, partly at very short intervals. Among them are current air and radiation measurement values, flood forecasts, and water levels. In a special area, the user is granted access to these frequently requested information items.

Via another access, the offers integrated last in the UINBW and containing major novelties are presented to the users.

In addition, a list of suppliers of environmental information is provided, such that the data can be accessed via names of authorities or institutions. To generate this information, the data collected in the database are used.

3.6 Incorporation of Special Offers

Many, but unfortunately not all offers can be integrated in the UINBW without any further adaptation needed. This especially applies to systems that entirely or partly consist of dynamically generated sites that can be reached via form-based queries only.

An example of such an offer is the web site of the Statistical State Office of Baden-Wuerttemberg [4], which largely consists of dynamically generated tables that are transferred to the user following a selection via a form.

For such web sites, the corresponding interfaces or adaptations have to be generated. In the case of the Statistical State Office, an additional, automatically generated site was established apart from the already existing web offer. This site is used as a starting point for the crawler of full-text search and offers links to all indexed subpages.

3.7 Quality Assurance

To reach a maximum quality of the information offered, quality assurance tools are available in the UINBW. In particular, availability of the web sites integrated in the UINBW is checked at regular intervals. Furthermore, the administrators and editorial staff are informed automatically about larger structural or contents-related modifications on web sites and, if necessary, may interfere with failed indexings of the web sites or adapt changed URLs.

4 Implementation

Prototype development in 2003 to demonstrate basic functioning was followed by the development of a first productive system in 2004.

While the prototype consisted of a simple database and several PERL scripts, the present implementation is based on a CMS with a back-end database. The software WebGenesis [5] developed by the Fraunhofer Institut für Informations- und Datenverarbeitung (IITB, Fraunhofer Institute for Information and Data Processing) and a MySQL database are applied.

Both in the prototype and the productive version, the Open-Source search machine [ht://Dig](http://Dig) is applied for implementing full-text search. Due to its variety of configuration options, it guarantees sufficient flexibility for indexing web sites and the search functions required. Configuration files for the full-text search machine are generated regularly by a PERL script via the data interface of the CMS.

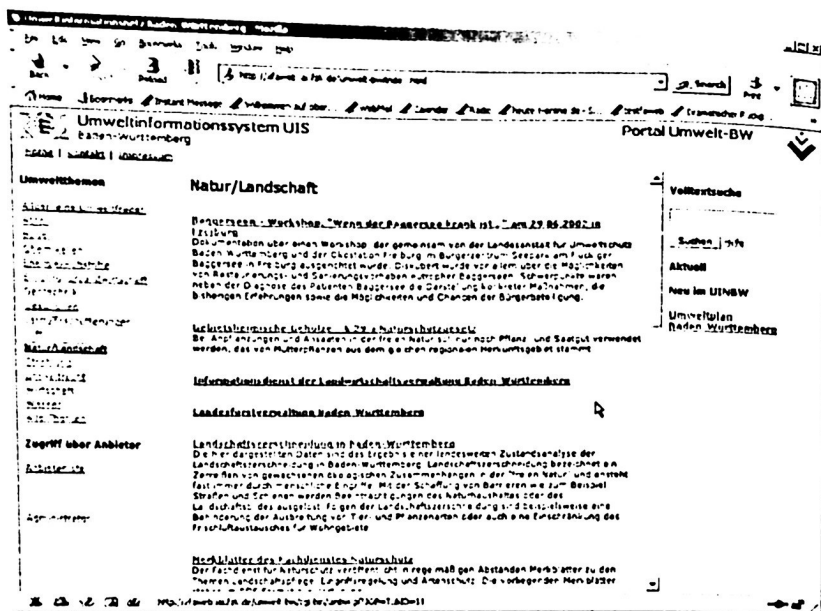


Fig. 2. User interface of the UINBW

5 Conclusion and Outlook

With the UINBW, information is offered by the environmental administration of the state of Baden-Wuerttemberg to the user in a transparent and clear manner.

Following commissioning of the UINBW [6], the expenditure required mainly consists of the administration of meta data and integration of new information offers. For this, the UINBW makes available an interface to the user, via which proposals can be made for the integration of further contents and information on modifications of existing sites can be transmitted.

The future extended version of the UINBW shall provide the user with a keyword search facility to specifically find searched for information. Moreover, the tools of automatic quality assurance shall be refined.

References

1. Thomas Bandholz: Erstellung eines semantischen Netzwerkservice für das Umweltinformationsnetz Deutschland - German Environmental Information Network, Forschungsbericht 201 11 612, Kurzfassung, Schlumberger Sema GmbH, Köln (2003)

2. Fernando Chaves: Personalisierung in WebGenesis: Individuelle Barrierefreiheit von Internet-Diensten gesteuert durch Beeinträchtigungsprofile, *visiT* Nr.2, S.4-5, ISSN 1616-8240 / ISSN 1617-1578 (Online) (2003)
3. gein, German Environmental Information Network [gein](http://www.gein.de), www.gein.de
4. stala, Statistisches Landesamt Baden-Württemberg, www.statistik-bw.de
5. webgenesis, WebGenesis Content Management System, www.webgenesis.de
6. umweltbw, Portal Umwelt Baden-Württemberg, www.umwelt.baden-wuerttemberg.de
7. Umweltinformationsgesetz, BGBl I 1994, 1490, bundesrecht.juris.de/bundesrecht/uig/
8. Thorsten Schlachter et al.: Environmental Information Network of Baden-Wuerttemberg – Integration of the Authorities' Environmental Information, *EnviroInfo* 2004, 18th International Conference Informatics for Environmental Protection, Geneva (2004)

