

Sharing Environmental Data with *gein*[®]

Thomas Vögele , Fred Kruse¹, Oliver Karschnick²

Abstract

In response to *EU Directive 2003/4/EC on Public Access to Environmental Information*, public administrations need to provide access to all environmental information available. The German Environmental Information Network (*gein*[®]), and the Catalog of Environmental Data Sources (Umweltdatenkatalog - UDK) are tools that support environmental agencies in Germany to comply with the directive. Both of them are currently being merged into a new system, *gein*[®] 2.0. This system will feature improved an improved user interface, better capabilities for indexing and searching of online information sources, and an extended access to meta-information-systems and databases.

In acknowledgement of the close relation of environmental and geospatial information, *gein*[®] 2.0 implements a metadata component (UDK 5.0) that is interoperable with the emerging regional, national, and European spatial data infrastructures due to an implementation of the relevant ISO standards and OGC specifications. In addition, new components for spatial information retrieval and the visualization of spatial information will be added.

gein[®] 2.0 implements a flexible modular system architecture, adaptable database interfaces, and a cascading query propagation. As a result, *gein*[®] 2.0 can be used to build a network of multiple information nodes on different levels of the administrative hierarchy. This may be useful for the integration of information providers on the community level, as required by the EU directive. It is expected that like the UDK, *gein*[®] 2.0 will play a leading role to the process of (meta)data and interface harmonization.

1. Introduction

Electronic media and the Internet are valuable tools in the effort to distribute environmental information and data in an efficient and cost-effective way. In Germany and Austria, efforts to use modern information technology to manage and disseminate environmental information in the public sector date back as far the early 1990's and the first release of the Umweltdatenkatalog (UDK), a digital Catalog for Environmental Data Sources (Swoboda et al., 1999, Swoboda et al., 2000, Karschnick et al., 2003). Almost ten years later, the *German Environmental Information Network gein*[®] was inaugurated at the world exhibition Expo2000 in Hannover (Bilo and Streuff, 2000). *gein*[®] is an online portal and search tool that serves as the official broker of environmental information for the federal government and the federal states in Germany.

Since spring 2003, both *gein*[®] and UDK are funded through a cooperative agreement between the German federal government and the German federal states (VwV-UDK/GEIN, 2003). Since then, operation, maintenance, and development of *gein*[®] and UDK is managed by the *UDK/GEIN Co-ordination Office (KUG)* located at the Ministry of the Environment in Lower Saxony.

In the same year, *EU Directive 2003/4/EC on Public Access to Environmental Information* went into effect (EU, 2003). This directive considerably extends the responsibilities of public administration with respect to the dissemination of environmental information and data.

One goal of the joint operation of UDK and *gein*[®] is to create an organizational structure that will help public authorities in Germany to comply with the EU directive and the respective state- and federal laws. Another objective is to better integrate both information systems in the emerging national and interna-

¹ Niedersächsisches Umweltministerium, Koordinierungsstelle UDK/GEIN, Archivstr. 2, D-30169 Hannover, kug@numis.niedersachsen.de, <http://www.udk-gein.de> ² Niedersächsisches Umweltministerium, Referat 14

tional infrastructures for environmental and geospatial data. To meet these objectives, the joint operation of UDK and *gein*[®] will be followed by a technical integration and a step-wise improvement of the two systems. The first phase of this integration is currently under way. The first release of the new *gein*[®] in 2.0 system is expected for the end of 2005. This paper gives an overview of *gein*[®] and UDK, the goals and concepts of the new *gein*[®] 2.0, and the current status of this development.

2. The German Environmental Information Network - *gein*[®]

gein[®] was the result of a research and development project supported by the German Federal Office for Environment Protection (Umweltbundesamt – UBA) between 1998 to 2000. The goal of this project was to build a tool that supports an easy and integrated access to environmental information distributed among numerous websites and databases operated by public authorities and associated organizations in Germany.

As a "Portal for Environmental Issues", the current version 1.1 of *gein*[®] offers both expert and lay users simple and user-friendly access to environmental information via one single Internet address - <http://www.gein.de> (Figure 1). In contrast to most internet search engines, *gein*[®] offers access to regular Web content (the "visible Web"), as well as to data bases (the "invisible Web"). Currently, *gein*[®] taps into eight environmental databases, including the UBA's *Geographic Information System for the Environment* (*Geografisches Informationssystem Umwelt, GISU*), the *Database of Environmental Publications* (*Umweltliteraturdatenbank, ULIDAT*), the *Catalog of Environmental Objects* (*Umweltobjektkatalog, UOK*) run by the federal state of Bavaria, and the *Catalog of Environmental Data Sources* (*Umweltdatenkatalog, UDK*) (see below). To make them accessible for the *gein*[®] search tools (both simple and advanced search), the databases are connected through the g2k interface, a standardized connector based on XML/SOAP (Kruse et al., 2001, Kruse et al., 2003).

As of July 2004, *gein*[®] had access to more than 400.000 web pages and approximately 500.000 database entries from 90 information providers. The high quality of the information referenced in *gein*[®] is ensured through a restrictive participation policy. Only suppliers of environmental information from governmental agencies, other public institutions at the federal and state levels, and organizations closely related to such institutions are allowed to become *gein*[®] information providers.

The *gein*[®] portal offers a number of paths to access environmental information: The latest news (including current events and publications) and up-to-date monitoring data (on air quality, water quality and other environmental parameters) published by various government agencies are displayed on news pages that are updated in hourly intervals. A hierarchy of environmental topics (following the UBA classification of environmental topics (UBA, 2002) leads to web pages that are the entry point to in-depth information on specific environmental issues. And an environmental calendar provides a time line of important environmental events.

All information sources in *gein*[®] can be accessed through a keyword-based full-text search. In addition, *gein*[®] offers advanced search- and indexing capabilities, including thesaurus-based indexing and query expansion, based on the *Semantic Network Services (SNS)* (Angrick et al., 2002) run by the Federal Environmental Agency UBA. Through the so-called "retrieval assistant", the controlled SNS vocabulary of environmental topics, geographic names, and time periods can be used to specify complex queries. For this purpose, the SNS integrates an environmental thesaurus of close to 40.000 topic terms, a gazetteer with 50.000 geographic objects, and a synopsis of historical and contemporary environmental events. All three types of semantic structures are interlinked and encoded as a topic map.

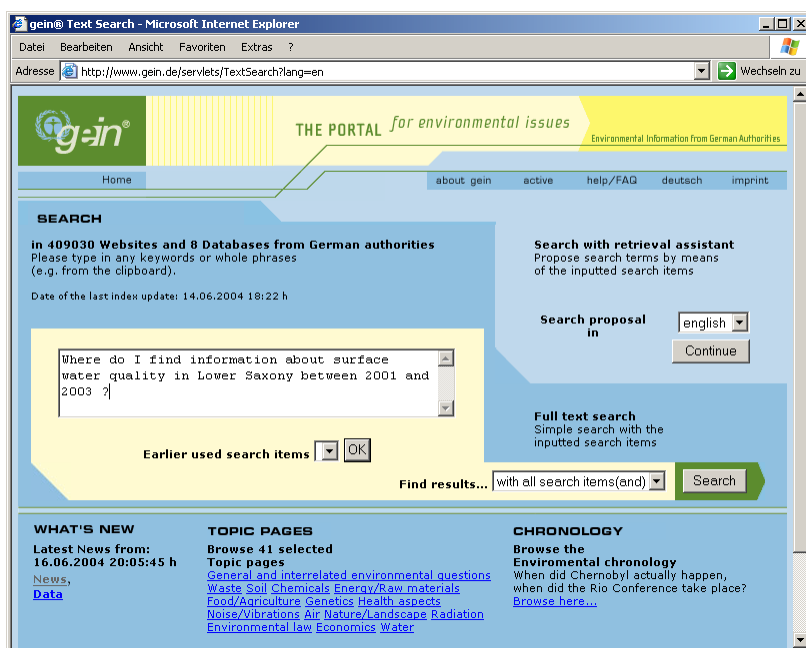


Figure 1: User interface of gein® 1.1

In addition to query expansion, the SNS is also used for automatic metadata generation. All documents (i.e., web pages, PDF documents, Word documents, etc.) that are directly accessible through *gein*® are subject to an analysis and automatic metadata annotation using SNS. The resulting metadata descriptions are simple, but suffice to describe the thematic, spatial and temporal reference of each document.

3. The Umweltdatenkatalog - UDK

The UDK is an information system based on metadata descriptions of information objects. Introduced in 1991, the system has a long and successful track record in field of referencing and management of environmental data sources, including geospatial data and services. Today, the UDK data model can be considered to be a quasi standard for environmental metadata in Germany and Austria. Recent enhancements of the UDK make the current version UDK 5.0 interoperable with spatial data infrastructures (SDIs) based on ISO and OGC standards and specifications.

The UDK metadata were designed to answer questions like: *Who* holds *where* *what* kind of data in *which* format? The metadata records are stored as so-called UDK objects, the attributes of which are e.g. object identification number, object name, general descriptions, key words as well as temporal and spatial coverage. The UDK distinguishes 6 object classes (Table 1) to accommodate different types of information objects. In addition, the catalog provides for the handling of addresses of responsible organizations and persons. As a special feature of the UDK data model, objects and addresses can be linked hierarchically (parent-children relation) and in parallel (object-object relation), based on the object identification number.

A comfortable, windows-based user interface (the Win-UDK) assists information providers and expert users with data entry, metadata management, and information retrieval. General public access to the UDK is provided through a web-based client-server application (the WWW-UDK). Each federal state and the federal government maintain an instance of a WWW-UDK database. Through a central information broker

(the Virtual- or V-UDK) these instances are interconnected. Consequently, a query submitted to the V-UDK (<http://www.umweltdatenkatalog.de>) is distributed among all participating UDK instances. The V-UDK is also directly connected to *gein*[®] through the g2k interface. Each search query submitted to *gein*[®] is therefore (among others) applied to all WWW-UDK instances.







	1.) data collection / data base
	2.) service / application / information system
	3.) document / report / publication
	4.) geodata / digital map
	5.) organisational unit / assignment
	6.) proposal / project / programme

Table 1: UDK Object Classes

The Umweltdatenkatalog was initially not designed as a catalog for geospatial data or (geo)services. However, it became obvious that the UDK can be an important component in the ongoing development of regional and national spatial data infrastructures in Germany. For one, the UDK is well established in almost every German federal state. This includes the necessary organizational infrastructure and the funding necessary to maintain and manage the catalog. Secondly, the UDK object model is proved to be flexible enough to handle geospatial data and geo-services. In fact, the original UDK data model included objects for *geodata/digital map* and *service/application/information system* (Table 1). As a result, the UDK is already used by the mapping agencies of three federal states (Hamburg, Hesse, and Lower Saxony) to reference geospatial data. To be able to fully integrate the UDK into the evolving SDI in Germany (i.e., GeoMIS.Bund[®] and GDI-DE[®]) and potentially international initiatives (e.g., INSPIRE), the UDK data model was modified to comply with ISO standards ISO 19115, ISO 19119 and ISO 19139. The system will also be equipped with an interface compatible to the OGC CS-W 2.0 catalog service interface specification.

4. *gein*[®] 2.0 - Integration of UDK and *gein*[®]

Since January 2003, both UDK and *gein*[®] are managed and maintained by the *UDK/GEIN Co-ordination Office (KUG)* located at the Ministry of the Environment in Lower Saxony. This is the result of a formal co-operation agreement between the German federal government and the governments of 15 German states with the intent to use synergies and make system operation more effective. The operational integration of UDK and *gein*[®] is accompanied by a re-design and technical integration in which both components will be merged into a new system, dubbed *gein*[®] 2.0. The end of 2005 expects completion of *gein*[®] 2.0.

One of the objectives of re-building *gein*[®] is to create a tool that will help government agencies (and other organizations) on all levels of the administrative hierarchy to better comply with the requirements of the EU directive 2003/4/EC on Public Access to Environmental Information.

4.1 Meeting the Requirements of EU Directive 2003/4/EC

In February 2003, the *EU Directive 2003/4/EC on Public Access to Environmental Information* went into effect. The directive extends citizen rights of access to environmental information. For government agencies in the European member states, this implies a significant increase in their responsibilities with respect to the display and distribution of environmental information. Information request, for example, have to be answered within a fixed time period. In addition, the directive calls for the “active” distribution of environmental information. This includes making publicly available not only lists of data sources and the respective responsible parties (i.e., metadata), but also providing direct access to the data and information themselves.

But not only new conditions for the access to environmental information were stated, the directive re-defines and extends the meaning of the term of “environmental information” itself. In addition to the “traditional” environmental topics, information about aspects of human health, genetically altered organisms, and economic issues are included in the new definition. The directive also extends the notion of who is a provider of environmental information. The new definition of a “public authority” includes government agencies on the federal and the state level, as well as local authorities (on the community level), and private and semi-private companies that provide services to public authorities.

To meet these requirements, the directive explicitly points to digital media and the Internet. In Germany, *gein*[®] and UDK already do support federal and state agencies in the dissemination of environmental information. However, both systems do still have a number of shortcomings that have to be overcome to meet the new challenges. The objective of the merger of the two information systems and the implementation of *gein*[®] 2.0 is to overcome these shortcomings.

4.1.1 Dissemination of Environmental Information

gein[®] 2.0 will be designed to allow for an effective dissemination of environmental information. This relates to the layout and the functionality of the application, but also to the ways in which information and data can be accessed. In *gein*[®] 2.0, the portal functions will be stressed, with a stronger emphasis on the display of up-to-date information and the possibility to limit the scope of the system to specific geographic regions (personalization). The thematic scope and organization of environmental topics will be in compliance with the EU directive, which means re-organizing the topic hierarchy and adding new topics (e.g., environmental aspects of human health). The technical architecture of *gein*[®] allows a selective inclusion of parts of a large web site. This will make it easy to integrate relevant information from organizations that do not have a thematic focus on environmental issues.

4.1.2 Active Access to Environmental Information

The EU directive calls for the „active“ distribution of environmental information. In the context of *gein*[®], this requirement means that the system will not only give access to highly aggregated information like web-pages and online documents, but also to raw data stored in databases. Through an XML/SOAP-based interface (g2k interface), *gein*[®] already enables access to databases that implement their own online search capabilities. In *gein*[®] 2.0, direct access will also include geospatial data and environmental information distributed through OGC compatible Web Mapping Services. The necessary interfaces and viewers will be integrated in the new system. In the long run, the implementation of new interfaces to tap into various types of environmental databases is being considered.

4.1.3 Lists of Information Providers

The EU-Directive calls for the establishment of “*registers or lists of the environmental information held by public authorities or information points, with clear indications of where such information can be found*”. This type of information can be found in UDK metadata descriptions of information sources and information providers. The functionality of the Web-component of the UDK will be fully integrated in *gein*[®] 2.0, hence providing the system with the ability to reference information sources and information providers registered in the more than 16 UDK databases maintained by the federal government and the federal states. This is particularly important for information that is not publicly available online.

4.1.4 Reports on the State of the Environment

The EU directive asks the EU Member States to “*ensure that national, and, where appropriate, regional or local reports on the state of the environment are published at regular intervals*”. In the long term, the responsible environmental agencies will be able to use *gein*[®] to produce such reports more effectively. Key topics in this respect are the *gein*[®] in search- and data-access capabilities, but also new tools and data flow networks that will be implemented in the future. This includes interfaces to national and international data infrastructures, like EEA ReportNet.

4.1.5 Harmonization of Metadata and Interfaces

In the last years the UDK has significantly contributed to a harmonization of metadata and interfaces in Germany and Austria. With the UDK being the metadata component of *gein*[®] 2.0, this process will be continued in the future and include not only environmental, but also geospatial data. In the long run, the harmonizing effect may also extend to standards and specifications used to build interfaces for the exchange of data and metadata. Version 5.0 of the WWW-UDK (which will be integrated in *gein*[®] 2.0) implements a catalog interface based on an OGC CS-W 2.0 compliant application profile. This profile is currently being developed by a German special interest group that brings together representatives from government agencies and industry (including the UDK/GEIN Co-ordination Office) on a national level. This application profile has the potential to provide an implementation-level specification for catalog services that can be used to homogenize developments on the regional, national, and potentially international level.

With respect to the exchange of environmental metadata, *gein*[®] already implements the g2k interface, an XML/SOAP based interface that has been successfully used to integrate a number of environmental databases. In the future, *gein*[®] 2.x will also provide interfaces to access raw data in environmental databases. Through a definition of the necessary profiles and interfaces, *gein*[®] 2.x will contribute to the building of a harmonized information infrastructure for environmental data.

4.2 Technical Integration of UDK and *gein*[®]

While the organizational merger of UDK and *gein*[®] has been completed, the technical integration is still under way. Full integration of both systems will be achieved in a phased approach. Completion of the first development phase, which will yield *gein*[®] 2.0, is expected by the end of 2005. The basic concept of the integration is to combine both systems into a single system with a modular and service-based software architecture. In the new system, existing data interfaces will be improved, and additional interfaces will be added. Likewise, user interfaces will be re-designed and new tools (e.g., to visualize digital maps) will be added.

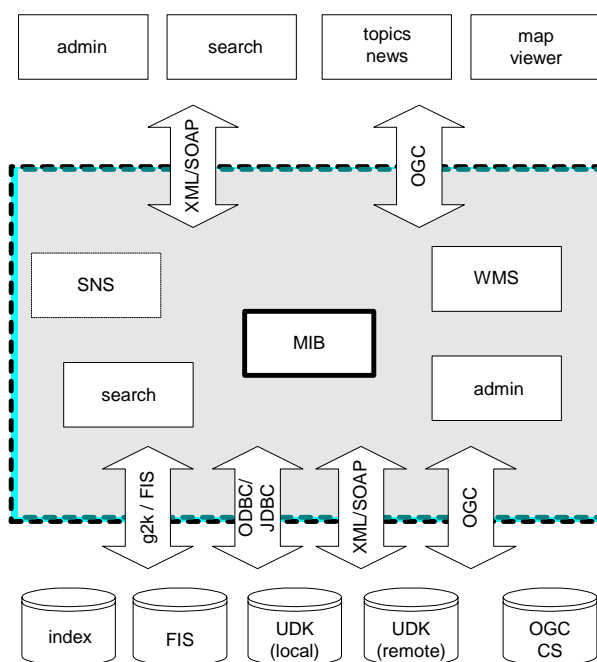


Figure 2: *gein*[®] 2.0 system architecture (schematic)

The core part of the system will be a set of modular components, with the UDK as the central metadata component of the new *gein*[®] 2.0 system (Figure 2). This *Meta Information Broker* (*MIB*) mediates between data sources (i.e., the Web, metadatabases, data catalogs, environmental databases, and index databases) on one hand, and the *gein*[®] 2.0 client- and visualization services on the other hand. To connect to different data sources, the MIB is equipped with a number of specialized interfaces. These include database interfaces on XML/SOAP and ODBC/JDBC basis, an OGC compatible catalog service interface, and an extended version of the *gein*[®] g2k profiles. The modular system architecture will support the installation of customized *gein*[®] applications. Users interested only in the metadata component (i.e., the former UDK) will therefore be able to run *gein*[®] 2.0 in a configuration that will provide the full functionality of UDK 5.0.

Among the new components will be an OGC compatible Web Mapping Service (WMS). Together with the respective clients, this service will be used in support of spatial information retrieval (i.e., query definition along the lines of geographic location) and the visualization of digital maps.

Other important modules are improved services for the creation and keyword-based search in fulltext indices, system administration services, and web clients for an improved *gein*[®] user interface. Some of the *gein*[®] 2.0 functionality will also be added through external services. An example are the Semantic Network Services (SNS) hosted by the Umweltbundesamt (see also section 2).

4.2.1 A Hierarchical Network of *gein*[®] 2.0 Nodes

A major goal of the *gein*[®] 2.0 environmental information network is to provide access to environmental information and data from all levels of the administrative hierarchy. To achieve this goal, the base of information providers will have to grow, in particular on the lower levels of the hierarchy, i.e., on the com-

munity level. In *gein*[®], the quality of the accessible information providers and data plays an important role. As a result, many processes are difficult to automatize, and a significant number of administrative tasks have to be done manually. While discussions about how a considerable increase in the number of information providers could be handled organizationally are still under way, the *gein*[®] 2.0 technical concept tries to provide a flexible and scaleable solution that can be adapted to different scenarios.

With *gein*[®] 2.0, it will be technically possible to build a network of *gein*[®] information nodes located on different levels of the administrative hierarchy. Each node will have a specific scope (in terms of the selection of registered information providers), and it's own administrative embedding. An information node may bundle information providers and other information nodes on the same level of the administrative hierarchy, as well as on the levels below. The information providers registered with an information node will typically belong to the geographic region defined by the administrative unit, the information node can be associated with.

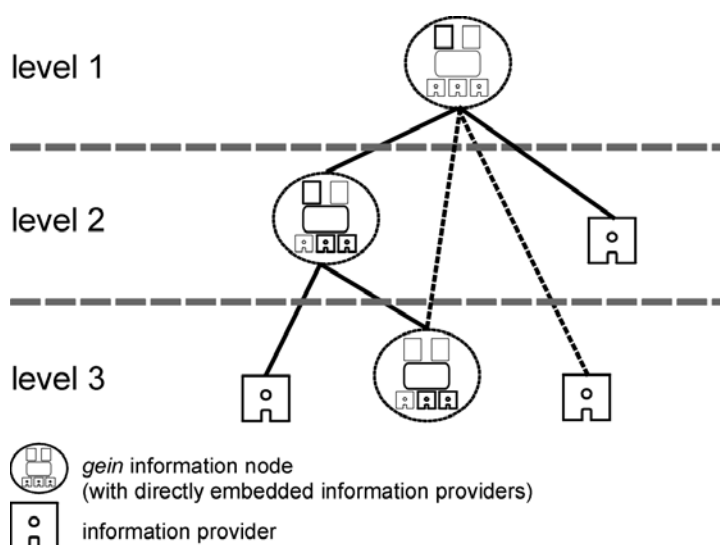


Figure 3. A hierarchical network of *gein*[®] 2.0 information nodes

An information node on a specific level of the hierarchy may be a registered information provider for nodes on higher levels. The *gein*[®] 2.0 catalog interface will provide the functionality needed to cascade information requests downward, and to percolate data and information upwards through the hierarchy. A user logged on to the federal *gein*[®] portal, for example, will be able to directly search through web-pages run by information providers on the community level. In future releases of *gein*[®], additional interfaces will allow for cascading access to other types of information in *gein*[®], like topic and news pages.

5. Summary and Perspectives

As a result of *EU Directive 2003/4/EC on Public Access to Environmental Information*, public administrations on all levels of the hierarchy are faced with the need to provide better access to environmental information and data. With the *German Environmental Information Network (gein*[®]) and the *Catalog of Data Sources (UDK)*, federal and state authorities in Germany already maintain an online information broker and a meta-informationsystem that can be used for this purpose.

Since 2003, both systems are jointly managed by the UDK/GEIN Co-ordination Office (KUG) located at the Environmental Ministry of Lower Saxony in Hannover, Germany. By the end of the year 2005, UDK and *gein*[®] will not only be integrated organizationally, but also technically. The result of the technical integration will be *gein*[®] 2.0, a modular system that incorporates the UDK (as a central metadata component), as well as functionalities and the general layout concept of *gein*[®]. The purpose of this new system will be to provide environmental agencies in Germany with an improved tool to comply with EU directive 2003/4/EC, and the respective federal and state environmental legislation.

Through new interfaces, *gein*[®] 2.0 will provide active access to information and data. To meet the increasing interest in geospatial data, tools to manage and access geospatial (meta)data will be included. Integration with emerging regional, national (e.g., GDI-DE) and European (e.g., INSPIRE) spatial data infrastructures will be enabled through the implementation of relevant standards and specifications.

Due to its flexible modular system architecture and database interfaces, *gein*[®] 2.0 can be used to provide access points to data and information on different levels of the administrative hierarchy. This may prove to be particularly useful for the integration of information providers on the community level, as required by EU Directive 2003/4/EC. It can be expected that *gein*[®] 2.0 will, like the UDK did in the past, contribute to the process of harmonization of metadata and interfaces.

6. Bibliography

- Angrick, M., R. Bös, et al. (2002). *Semantic Network Services (SNS)*. Proceedings of the 16th conference "Environmental Informatics 2002" (EnviroInfo'2002), Vienna.
- Bilo, M. and H. Streuff (2000). *Das Umweltinformationsnetz Deutschland - GEIN2000 - Fachliche Anforderungen an ein Forschungs- und Entwicklungsvorhaben*. 3rd workshop Hypermedia im Umweltschutz, Ulm.
- EU (2003): *Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on Public Access to Environmental Information*. In: Official Journal of the European Union, L 41/26, 14.2.2003, http://europa.eu.int/eur-lex/pri/en/oj/dat/2003/l_041/l_04120030214en00260032.pdf
- Gein (2004): *German Environmental Information Network (gein)*. <http://www.gein.de>
- Kruse, F., Eichler, M., et al. (2001a): *Die XML-Schnittstelle des UDK 4.2 - Der Schlüssel zur Integration von Umweltsystemen*, Tagungsband des 4. Workshop des GI-Arbeitskreises Hypermedia im Umweltschutz und 3. Workshops der GI-Initiative Environmental Markup Language, Ulm, Germany
- Kruse, F., Eichler, M. (2001b): *Der Windows UDK 4.3 auf Basis der SOAP Technologie*, in: M. Hilty, P.-E. Gilgen (Eds). Proceedings of the 15th International Symposium Informatics for Environmental Protection „Sustainability in the Information Society“, Zurich, Switzerland
- UBA (2002): *Umweltklassifikation*, <http://www.umweltbundesamt.de/uba-datenbanken/klassen.htm>
- VwV UDK/GEIN (2003): *Verwaltungsvereinbarung zwischen Bund und Ländern über die gemeinsame Entwicklung und Pflege des Metainformationssystems Umwelt-Datenkatalog UDK und des Umweltinformationsnetzes Deutschland GEIN*. <http://www.udk-gein.de/allgem/vwv.shtml>.
- Swoboda, W., Kruse, W., et al. (1999): *The UDK Approach: the 4th Generation of an Environmental Data Catalogue Introduced in Austria and Germany*, Proceedings of the 3rd IEEE Meta-Data Conference, Bethesda, Maryland, USA
- Swoboda, W., Kruse, W., et al. (2000). *Harmonisierter Zugang zu Umweltinformationen für Öffentlichkeit, Politik und Planung: Der Umweltdatenkatalog UDK im Einsatz*, In: K. Tochtermann, W.-F. Riekert (Eds). Proceedings of the 14th International Symposium „Computer Science for environmental Protection“, Bonn, Germany

Karschnick O., Kruse F., et al. (2003): The UDK and ISO 19115 Standard, in: Proceedings of the 17th International Conference "Informatics for Environmental Protection". Cottbus, Germany

Kruse F., Karschnick O., et al. (2003): gein® - Planning the next generation, in: Proceedings of the 17th International Conference "Informatics for Environmental Protection". Cottbus, Germany