Geographic Data Warehouse to Consolidate Dispersed Environmental Information

Martin Huber
Condesys Consulting LLC
Switzerland
University of Salzburg
Austria
martin.huber@condesys.com

Hanspeter Eberle
Environmental Protection Office, Princip. of Liechtenstein
Liechtenstein
Hanspeter.Eberle@aus.llv.li

Samuel Schäpper
Personnel and Organization Office, Principality of Liechtenstein
Liechtenstein
Samuel.Schaepper@apo.llv.li

Abstract

Though the Principality of Liechtenstein is a very small country in the heart of Europe it has to assure the same legal standards regarding environmental protection as its much larger neighbours. Only 15 employees cater for around 350 legal procedures, a task covered by more than 300 people at the federal level in neighbouring Switzerland.

A detailed business analysis revealed 280 different data collections for environmental monitoring and for the management of applications, authorizations and orders. These data collections were managed in different database management systems and spreadsheets, so an integral view of the state of the environment and of all the authorizations was impossible. However, given the diversity of tasks, a common database structure and a unifying application context was hardly imaginable, yet the limited resources imposed exactly this.

In the course of a Master Thesis and with support from a GIS business consultant, the GIS expert of the Environmental Protection Office analysed the requirements and modelled the functionality of the Liechtenstein Environmental Information System. With abstraction and a generic approach, a unified data structure was designed to host the environmental protection data from diverse sources. The key element of this data structure is the fact that all data is associated with common reference systems. Furthermore, a minimal set of user functionality and interaction elements was designed with the help of user interviews, use cases and application screen layouts.

To prove the suitability of both the data structure and the functionality, a prototype application was developed. Different software solutions were evaluated, commercial packages as well as open source components. The prototype convinced decision makers and end users alike. With minor adaptations, the original design is now being implemented.

The project profited from a systematic modelling approach, which enabled an intense exchange of ideas between end users and solution architects. The resulting generic geographic data warehouse concept can be used not only in environmental protection, but for all governmental offices struggling with dispersed data sources.

Introduction

Environmental Information. The continuous extension of the environmental legislation brings about a widening spectrum of duties and tasks for the environmental administration and, as a consequence, new data collections to be managed. Unlike a scientific description of the "environment", where an ecosystem with flows of energy and matter builds the framework for data collection (cf. Huber 1990, 1994, Leser 1991), the administrative data collection is focused on individual cases of applications for resource use or "pollution permits". All data collections that are not treated in an individual case is subsumed under the notion of environmental monitoring and reporting. Though the creation of environmental laws in cases like water or forest protection dates back more than a hundred years, it is only since 1999 that the European Environmental Agency (EEA) established the DPSIR framework (Smeets & Weterings 1999), a correlation system or feed-back control system, with which it is expected to be able to model the full causal chain leading to environmental damage as well as the influence of environmental policies to break that chain. Today, environmental monitoring and reporting in OECD and EU countries and by the United Nations Organization (UNEP 2007) builds on the DPSIR framework. DPSIR stands for **D**riving Forces-**P**ressures-**S**tate-**I**mpacts-**R**esponses. It is not the intention of this paper to describe the DPSIR framework. Crucial for the understanding is only that DPSIR dictates to a large extent, what kind of parameters are to be collected to describe the state of the environment, the impact of human activities on the state of the environment and the effect of environmental policies to reduce or neutralize human impact. An environmental information system will have to measure its value against this framework.

Environmental Information in Liechtenstein. The Principality of Liechtenstein has an area of 160 km^2 and slightly more than 35,000 inhabitants. The western part of the country is in the Rhine valley plain at around 450 m.a.s.l, the eastern part is very mountainous ascending steeply to more than 2,500 m.a.s.l. The high economic activity of Liechtenstein is concentrated on a small stripe of 20 km length in the plain and on the foot slopes of the mountains. Due to the very diverse industry, almost all environmentally regulated sectors have at least one case in the country. A detailed internal study revealed that the Environmental Protection Office has to cater for approximately 350 legal procedures, for which 280 different data collections were established. Though there are usually only a few cases per year for most of these 350 procedures, for each case one of the 15 staff members of the office has to know the exact procedure and has to collect the relevant information. A redesign of the environmental information system under such circumstances could only mean to cut the Gordian knot, because every modification of one or the other data collection would not really reduce complexity or help solve the resource problem.

Redesign of the Environmental Information System. In spring 2005 the Environmental Protection Office started a project to design and implement an environmental information system (EIS). Based on the description of all the 350 legal procedures and a metadatabase describing the 280 data collections, a system and solution design process was started. First of all, similarities in the procedures were analyzed leading to 34 typical tasks or workflow steps. In thorough interviews with all users, these tasks were further specified and prioritized in terms of the frequency of occurrence and the value of a support by an information system application. Based on the user interviews, a prioritized list of requirements was established. In the subsequent exemplary design process, all the typical artefacts of a solution and system design were produced: a business design and a business view, use cases, a functional view, application screens, a systems architecture (cf. Figure 1), a prototype implementation to test the feasibility of the core concepts and a detailed system specification. The implementation of the core module of the EIS is currently ongoing and will be completed by the end of 2008. The design process is described and methodologically commented in a master thesis by H. Eberle (Eberle 2006). This paper will concentrate on two crucial points for the consolidation of dispersed environmental information:

- The unified data structure for environmental data for the execution of environmental laws.
- The generic functionality around the unified data structure to generate whatever task and process specific interface that is needed.

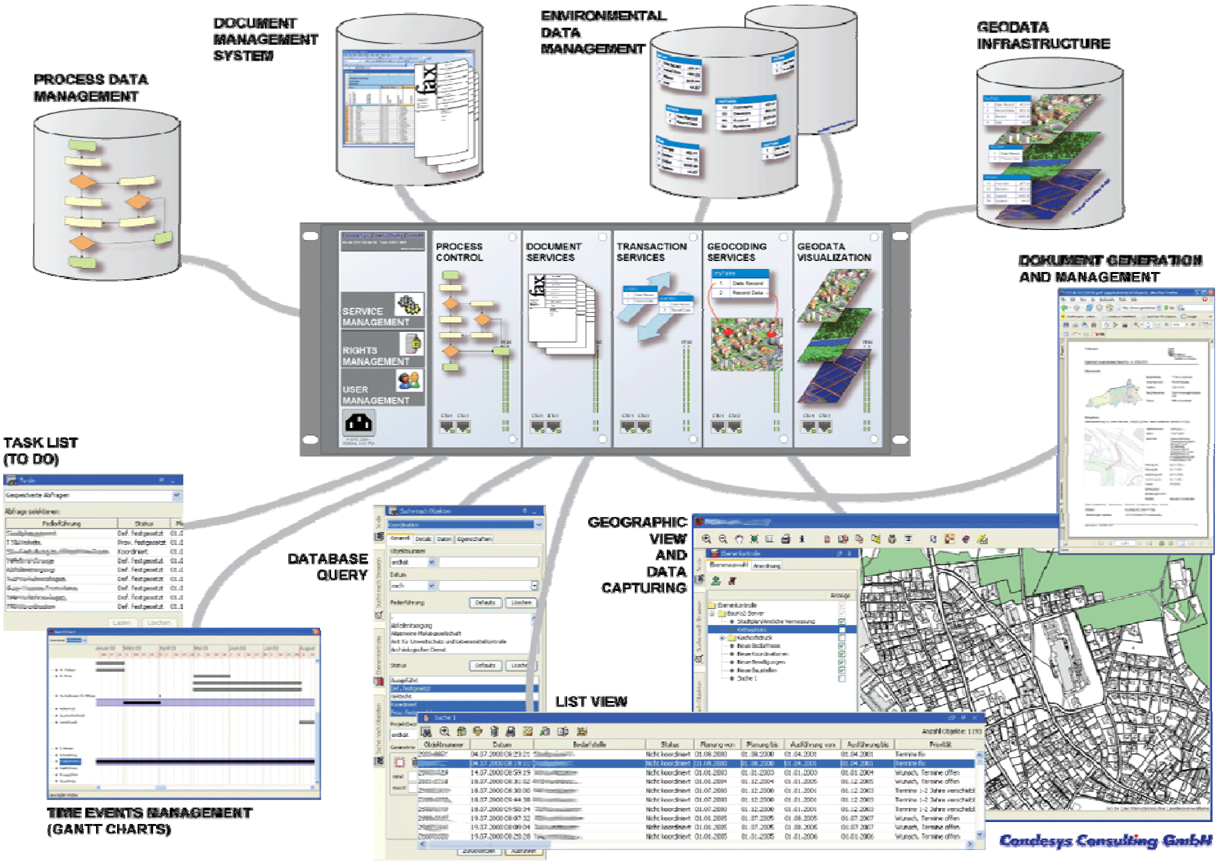

Figure 1: Master systems architecture of the Liechtenstein Environmental Information System.

A Unified Data Structure for Environmental Data

When analyzing a major selection of data sets at the Environmental Protection Office, it was not evident, that there was a common denominator. Each data set was designed exactly to the requirements of the specific law or regulation for which it was collected. As a first step to consolidation, the few things in common were established in a very general data structure. For the prototype implementation, a group of the most important data collections and databases was selected for the integration into an environmental data warehouse. With each table, the insight into the real information needs grew and the draft data structure could be refined. After the integration of about 20 to 30 tables the picture became much clearer. In fact, there were only 7 domains of information that were collected (cf. Figure 2):

- Objects: all installations with relevance to the environment i.e. plants, machinery and measurement sites for environmental monitoring.
- Cases: all applications, permits and restraints related to the use of environmental resources or to plants and machinery with impact on the environment.
- Registry: business cases and plants under the environmental legislation always have a relationship to natural or legal persons and can be localized with addresses or within parcels of the land registry. Persons, addresses and land registry are kept in their own specific

systems which are generally used by the administration of the Principality. It is therefore reasonable not to copy all this information which quickly becomes outdated, but to link directly to the leading systems, i.e. the registry of persons and the land registry.

- Measurements: independent on whether one looks at a plant with emissions of pollutants or at the state of the environment, in either case a measurement or observation installation has to be made and measurement values according to the methodology of the measurement installations are collected. From the organization of most of the measurements analyzed in the prototype study it was becoming clear that their structure was adapted rather to the traditional measurement forms than to a long term environmental monitoring. Furthermore, a lot of metadata was only in the heads of the (oftentimes external) persons who carried out the measurements. Based on these findings, a data structure was defined that was able to host all data from diverse sources without any compromise on data description (metadata) or on data analysis capabilities.
- Methodology: the static metadata on how measurements are taken can be stored in the methodology section. It describes the sector of the environment concerned, the method applied, the units of measurement etc.
- Spatial Representation: an environmental information system is not primarily a GIS, but the data and measurement series stored in an EIS always have a relation to geographic space. This relation can be explicit with a geographic attribute (point, linestring, polygon) to a plant or a measurement site or implicit via an address or a land parcel. The spatial representation hosts primarily a location, which in turn can be geographically represented as point, linestring, polygon, as a layer in three dimensional space (soil, geology, atmosphere) or as a segment in a network (rivers, roads).
- Dates: cases are always related to events: an application has a deadline, a permit expires or a restraint has to be regularly controlled etc. To optimize the work organization of the very limited human resources, all events scattered in the many files can be grouped in one repository for events and appointments.

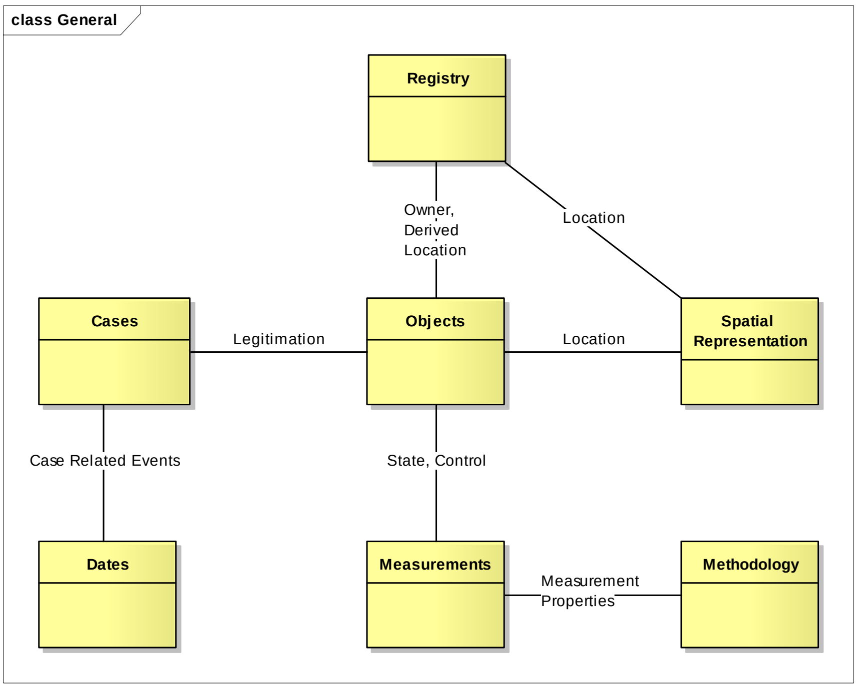

Figure 2: General view of the information domains needed for the Liechtenstein Environmental Information System.

One hypothesis at the beginning of the analysis was that data structuring patterns would be found which then would have to be adapted for the data collection domain at hand. It was never expected that it would be possible to unify all environmental data into one data structure, which is outlined in Figure 3. The unified data structure has several positive effects:

- Staff members can switch from one environmental domain to another because the information about cases, objects and measurements is structured the same way as in their home sector and properly documented with metadata.
- For external contractors in data collection an easy to communicate data documentation can be provided, which enables them to deliver data that can be integrated into the EIS without frictions, independent of the environmental sector concerned.
- Applications for often recurring tasks like the analysis of time series, the spatial interpolation of measurements, the creation of emission and immission maps or the quantitative description of storages and flows in a process system have to be implemented only once for all environmental sectors.
- With a proper application design, the data structure is not restricted to a data warehouse approach only, but can be used for operational data as well. It is even extensible for new data collection requirements with minimal to no changes on the database schema.

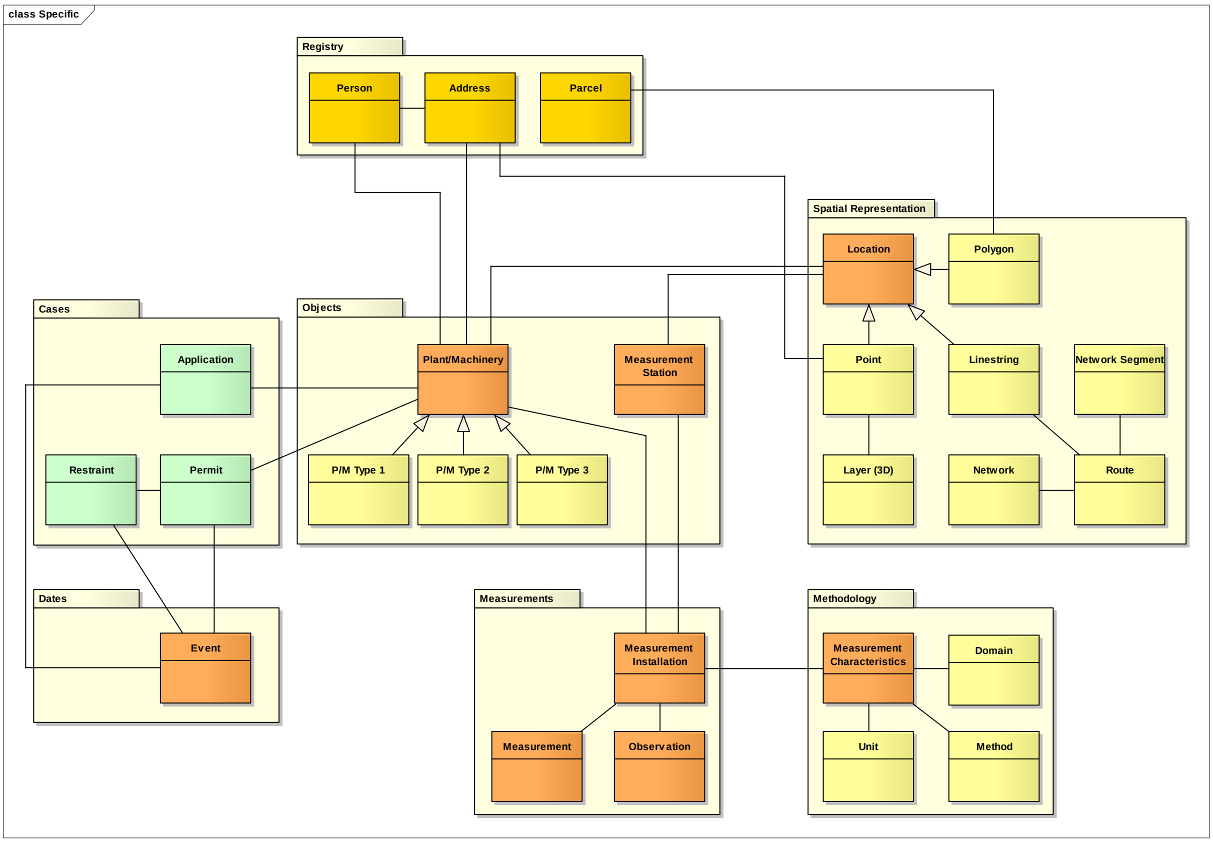

Figure 3: Overview of the unified data structure for the Liechtenstein Environmental Information System.

The main precondition for reaping theses benefits is an end user application that is capable of analyzing the metadata in order to present data specifically for a business case, a reporting task or an environmental sector.

A Generic Application Design

Inspired by the breakthrough in data modelling, it became obvious that also the application system could be much more generic than expected. Instead of building an application for each procedure, a generic case management system could be designed, whereby the user context or the procedure specific application configuration selects the relevant cases, object types, procedures, functionalities etc. An even greater impact is expected at the reporting end of the data treatment chain: while previously cross-sector data integration was nearly impossible, it is not relevant anymore from which environmental sector data to be reported comes from because all data can be treated and visualized with the same tools.

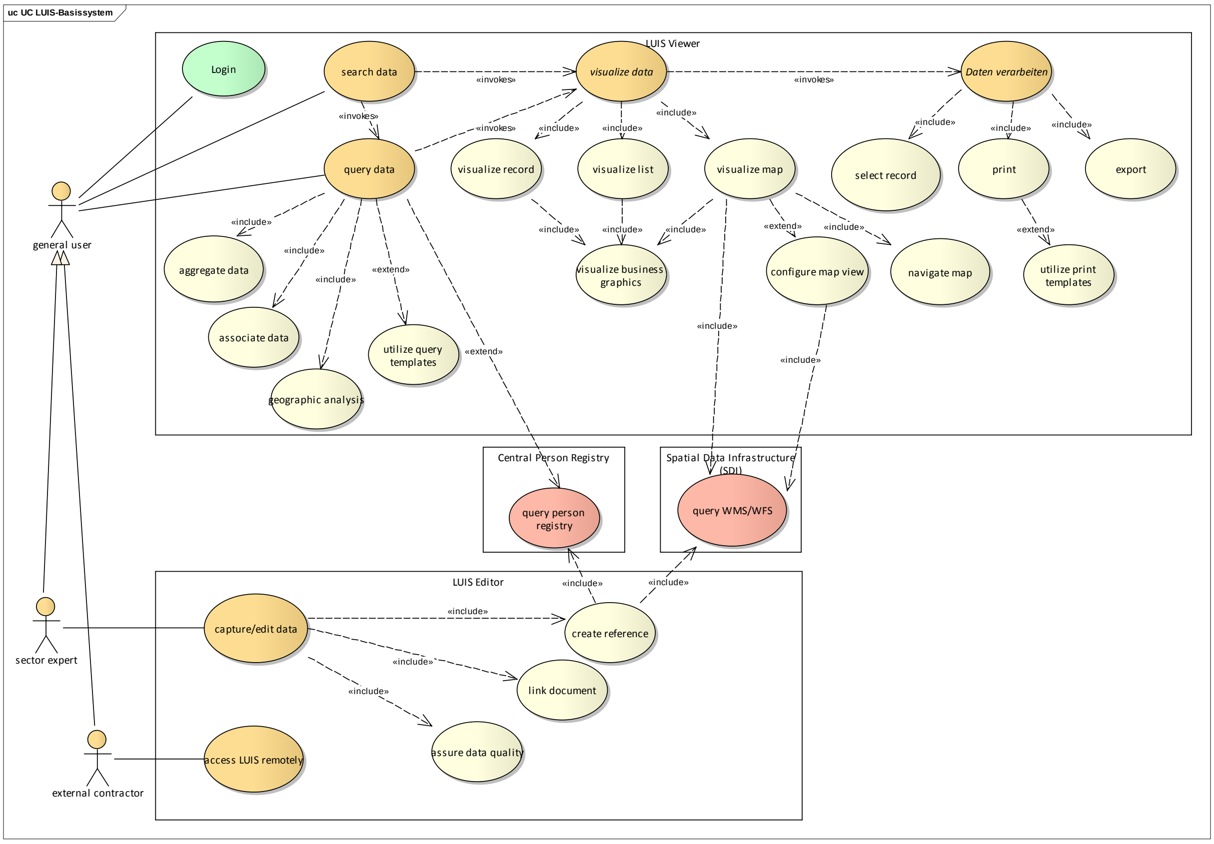

Figure 4: Use cases for the base application of the Liechtenstein Environmental Information System.

Figure 4 shows the use case diagram of the generic base application of the Liechtenstein Environmental Information System. A lot more use cases have been defined for the detailed system specification, particularly for the configuration, customization and adaptation of the generic system, but their description would go beyond scope of this paper.

The general workflow through the base application starts with the selection of the required data sets. From there one can either refine a query and perform analyses or move directly to the visualisation. Visualisation supports explorative data analysis by browsing through records or displaying the information on maps and in business graphs. Several data sets and analysis results can be visualized at the same time. From visualized data sets, individual records can be selected for further treatment or editing or a whole data set can be exported or printed. In addition to this, sector experts and external contractors can capture new or edit existing data. What can be done with a specific data (sub-)set is always determined by its metadata.

On top of the base system an application configuration tool is required to combine selections of data sets, customized query masks and analysis and visualization templates for specific tasks and business cases. Figure 5 shows a configured application screen of the prototype of the Liechtenstein Environmental Information System. The vertical bar on the left displays query templates for specific data collections. In the centre, one query is displayed with a tree view synchronized with a map display. By providing preconfigured views, the end users will very seldom be confronted with the generic nature of the system, while continuously profiting from its advantages, mainly the consistent presentation of all environmental data and the flexible adaptation of the application for new procedures and data collections.

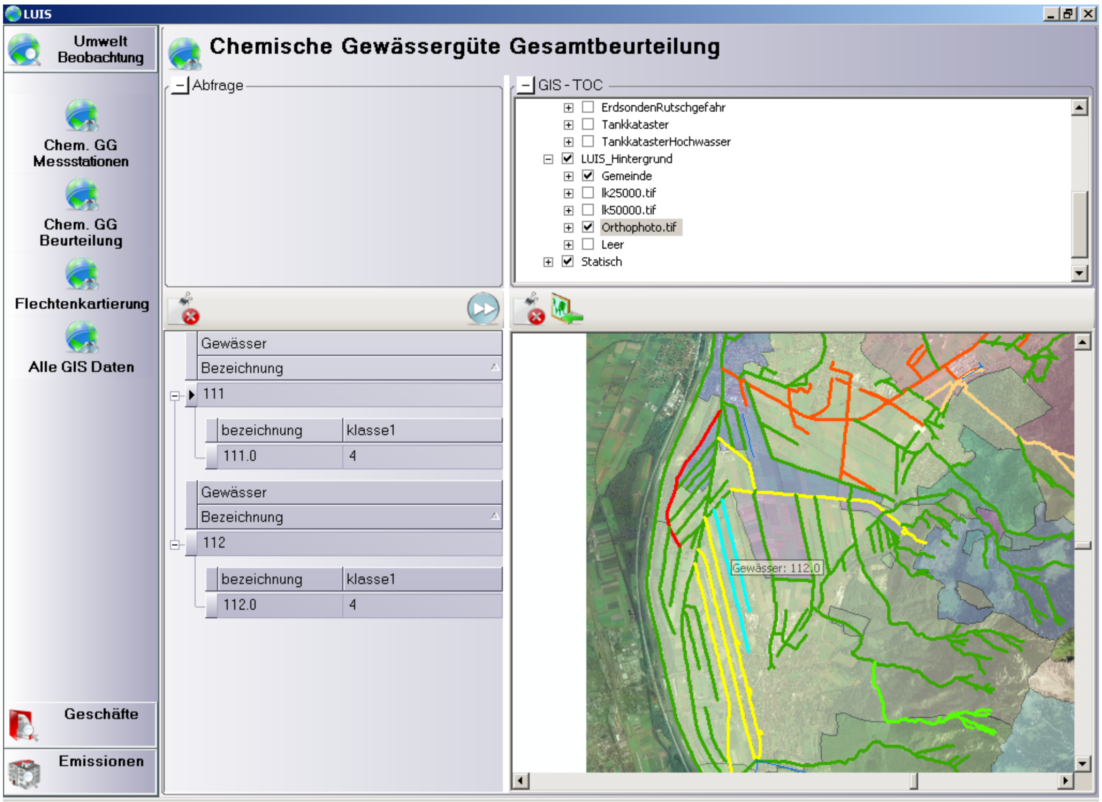

Figure 5: Application snapshot of the prototype application of the Liechtenstein Environmental Information System.

Conclusion

For a small country like the Principality of Liechtenstein, the use of unified data structures and generic applications in the fulfilment of the environmental legislation is not an academic exercise, but an essential design requirement to avoid the fatal data chaos and to reduce the constant work overload of the limited human resources. In an exemplary design process, the Environmental Protection Office of Liechtenstein created a solution and system design for a consolidated environmental information system. The feasibility of the design was proven in a prototype implementation, which paved the way for the realization of a base system.

The unified data model hosts environmental data from all environmental sectors in the same structure. It also caters for the management of individual business cases related to environmentally relevant installations and activities. The generic application supports the selection and filtering of individual data collections as well as the display and analysis of alphanumeric, spatial and temporal data. Metadata-driven application configurations help reduce the complexity of the generic application by narrowing down the data selection capability and the functionality to the needs of specific procedures. The resulting generic geographic data warehouse concept can be used not only in environmental protection, but for all governmental offices struggling with dispersed spatio-temporal data sources.

References

Eberle, H., 2006, Realisierung eines Umweltinformationssystems mit IT-Standard Methoden, Master Thesis University of Salzburg, http://www.unigis.ac.at/club/bibliothek/Abschlussarbeit/1102.PDF.

Huber, M., 1990, Elektronische Datenverarbeitung in der Physiogeografie - Allgemeine Konzepte und Werkzeuge für Studium und Forschung sowie ein Entwurf für die computergestützte Erarbeitung der geoökologischen Karte 1 : 25 000, Master Thesis, Basel.

Huber, M., 1994, The Digital Geo-ecological Map - Concepts, GIS-Methods and Case Studies, Physiogeographica 20, Basel, 144 p.

Leser, H., 1991, Landschaftsökologie, Stuttgart: Ulmer, 647 p.

Smeets, E., R. Weterings, 1999, Environmental Indicators: Typology and Overviews, Technical Report No 25, European Environmental Agency – EEA, Copenhagen, http://reports.eea.europa.eu/TEC25/en/tech_25_text.pdf.

UNEP, 2007, Global Environment Outlook GEO4, Nairobi & Valletta,540 p., http://www.unep.org/geo/geo4/report/GEO-4_Report_Full_en2.pdf