

Envirocat: a Swiss Catalogue for Sharing Environmental Information

Karin Fink¹, Véronique Ortner¹, Hy Dao², Jean-Philippe Richard³,
Frédéric Vogel³

Summary

This paper presents the envirocat⁴ project, the Swiss Catalogue of Environmental Data. Envirocat deals with different aspects of sharing: federal and cantonal institutions share information about environmental data with other institutions and the public in a system accessible to all. Sharing is also significant when it comes to providing the contents of the catalogue: many different partners share the workload of updating.

Recently several different initiatives concerning metadata and centralized databases have been launched at different levels (national and cantonal) in Switzerland. Sharing therefore implies connecting these initiatives and coordinating the collection of metadata.

Envirocat has been developed entirely with open source technology, making it possible for others to share the experience gained through envirocat and to take part in the future development of the application.

This paper focuses on different aspects that are equally important for a successful envirocat catalogue: the institutional mandate, environmental data concepts, open source technologies and the data provider community.

1. Introduction

Although at the present Switzerland has only signed (but not ratified) the Aarhus Convention (1998), which aims to promote public access to information, participation in decision making and access to justice in environmental matters, a growing

¹ SAEFL (Swiss Agency for Environment Forest and Landscape), Environment Monitoring Section)

² University of Geneva, Geography Department & UNEP/DEWA/GRID-Geneva

³ UNEP/DEWA/GRID-Geneva

⁴ The envirocat application can be visited at <http://www.umwelt-schweiz.ch/envirocat> or at <http://www.envirocat.ch>

number of politicians and local authorities soon understood that “environmental democracy” is not possible without massive participation of the citizens, meaning easy public access to information about the environment. A consequence of this was the 1996 decision of the Swiss Federal Council to give a mandate to the Swiss Agency for Environment, Forests and Landscape (SAEFL), in cooperation with the Swiss Federal Statistical Office, to introduce an environmental data catalogue, to improve public access to environmental sources.

It was therefore decided to introduce throughout Switzerland the Catalogue of Data Sources (CDS), a tool proposed by the European Environment Agency (EEA). Between 1997 and 1999, the cantons and the federal offices put the initial environmental metadata in the Swiss catalogue by means of WinCDS, which is stand-alone software provided by the EEA. In 2001, the providers of metadata called for a tool allowing them to manage their metadata on-line. There were several attempts made to implement existing software, but the specific demands could not be met so that the SAEFL decided to develop its own application. The new system is called envirocat and has been available since October 2003. The old CDS was closed down in July 2004, and the new envirocat system is now used by all partners and by the public.

2. Objectives

The main purpose of the envirocat catalogue is to encourage the use of available environmental data by making it easier for the administration and the general public to access these data. The description of data in a catalogue is essential for quick retrieval and for correct interpretation of the data by users. Envirocat aims to answer the following questions:

- What environmental data are collected in Switzerland?
- On what legal basis are the data collected?
- Where, when and by whom are the data collected?
- Who distributes the data and how are the data accessible?
- What products derive from these data?

Sharing environmental information using such a catalogue is crucial for Switzerland since the 26 cantons mainly act individually in collecting environmental data and it is often difficult to have a national overview of the data collected or of the flows of data between the federal level and the cantons.

Envirocat gives an overview of the environmental data available, contributing to the establishment of an environmental monitoring and reporting system in Switzerland. In this respect, networking activities with data providers are of the utmost importance.

Furthermore, envirocat is a way of presenting environmental monitoring activities to the public and of making the process of generating environmental data more transparent and easier to understand.

3. The process of generating environmental data and the envirocat data model

The types of environmental data sources described in envirocat represent various steps in the process of data generation: the legal basis, projects, data collection (observational networks), data and products (such as documents and Internet sites). Each data source description can be linked to one or more descriptions, and to any number of contact addresses (persons, organizations). These links allow the process of data generation to be followed. For instance it is possible to describe data in relation to the observational network where the data were collected, or it is possible to connect products to data. The elements can also be linked to the legal basis on which they were produced. Figure 1 shows an illustration of the environmental observation process and related element types.

The element types are defined as follows: *persons* are contacts for described data and can be associated with an organization. *Organizations* are institutions that produce or distribute environmental data; they can be associated with other organizations through “parent-child relations”. In the *law* element type, Swiss environmental law and international agreements can be described. The *project* element type describes the organizational framework of environmental observation activities. In *survey* research campaigns, networks of measuring stations etc. are specified. The emphasis is on the description of measurements and methods rather than results (compared to the *data* element type). Instruments such as specific software or measurement tools can be characterized in the *tool* element type. The most important element type in our view is *data*. It contains descriptions of raw data, and of data that have been processed, analysed or interpreted. GIS layers, inventories, lists etc. also belong to this category. *Documents* (books, articles, printed maps, graphs etc.) are the products that result and *hyperlinks* describe important websites.

Envirocat can ensure direct access to data via http addresses, provided that the data are available on the Internet. It is possible for every metadata description to include links to further information on data, methods, visual forms of the data (web map browsers or graphics on the Internet) and so forth. In addition, the *hyperlink* element type allows reference to selected websites. So envirocat also acts as a specific guide to environmental information on the web.

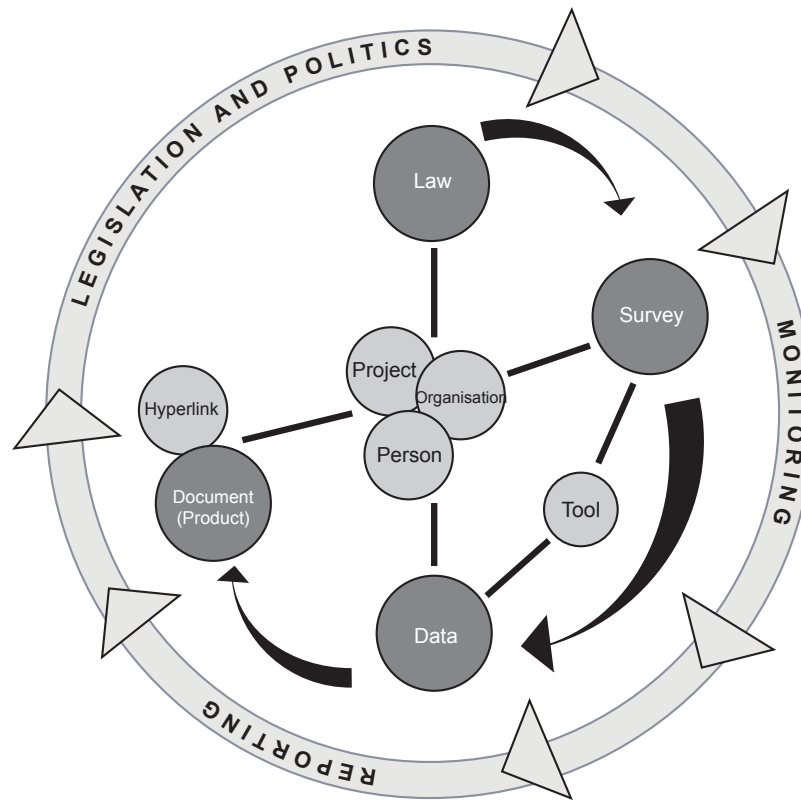


Figure 1: Main element types in envirocat representing the process of generating environmental data (dark grey) and other related elements (light grey) (illustration adapted from EEA 2003, Reportnet)

A foundation stone in the development of envirocat is a data model to ensure the transfer of all information from the former CDS into the new database. Another specification of the data model is that it considers the Swiss standard for geographical metadata (GM03). This standard is a profile of ISO 19115, the standard for geographical metadata respected throughout the world. The envirocat data model contains all mandatory attributes of the GM03 core⁵, as well as some other optional attributes. The decision to implement only a small part of the GM03 core responds to the need for a model particularly specialized in describing environmental data and the process of generating environmental data.

⁵ For further information (in German or French with English summary) on the GM03 metadata model see http://www.kogis.ch/frameset/gm03_d.htm.

4. Content definition

From a thematic point of view, a data source is considered as environmentally relevant if it fits into the DPSIR (Drivers – Pressure – State – Impact – Response) model developed by the EEA. However, envirocat first focuses on topics and keywords considered in Swiss environmental law. Currently, envirocat contains data descriptions concerning the following keywords: waste, contaminated sites, biotechnology, soil, electro smog, energy, water protection and fisheries, noise, air, atmosphere, substances and products, incidents, tank installations, EIA, forest and timber, landscape, nature, wildlife and hunting.

At the moment, one of the main tasks of the project is to harmonize the contents of the database and to elaborate detailed references to topics for the project partners.

Different opinions emerged when it was tried to separate environmental data from other data such as statistical data on population, traffic, health and so forth. Often such data are used as a reference when environmental statements are made. Nevertheless, it was decided to focus on a narrower definition of environmental data – in particular to minimize the workload for our partners. Furthermore, the Swiss Catalogue for Geographical Data (geocat.ch) ensures access to a relevant part of basic data needed for environmental statements. Also the Swiss Federal Statistical Office provides its collection of data on the Internet. Descriptions in envirocat contain links and addresses leading to these data sources.

When it comes to standardization of the contents, the interpretation of the model's element types and attributes is crucial. Since the metadata are collected online, it was necessary to add on-line help texts to the attributes since label names are not always self explanatory. We found that explanations are not always considered in depth, and that regular checks by the project team are necessary to ensure data quality. We expect this to change once our partners have become familiar with the data model of envirocat.

Very often, important information is also given in the free text fields intended for additional explanatory remarks. Free text fields therefore seem to be necessary and they show that standardization has its limits.

Regarding the contents, it is also a question how much detail should be used to describe a dataset. Should different time series be described independently or does it make sense to sum these up in one set of metadata? Is it always necessary to describe the dataset and the data collection, even though there is not much extra information to give? What choice of documents should be described in envirocat? Such questions cannot be answered easily and we still need a great deal of discussion to find solutions practicable for all partners.

In addition to these general questions, Switzerland's federal structure creates other difficulties: data are often collected at the national level (e.g. federal inventory of mires and raised bog landscapes) and these should therefore be described by the federal authorities. However, in the former CDS, cantons often used to describe the

same dataset just for their canton. Since we want to prevent double entries (or up to 26 different entries) about the same dataset, a rule must be established about who collects the data in envirocat. The metadata description should be done directly by the owner of the data, but this regulation is not always practicable – e.g. for data collected in cooperation between cantons and federal offices. Such questions are also political questions – it was therefore decided to describe some of the national datasets in a “neutral” group – neither belonging to the cantons nor to a specific federal office.

We have found that it is very difficult to achieve homogeneity of the database contents. This has to be constantly revised by the project team – a very time-consuming task. In the end, the data owners have to decide themselves what to describe and how to describe their data. Still we consider this as one of the great benefits of such a system: the reflections and decisions of the partners gathering the metadata add extra value to the information.

5. System architecture and technological choices

The architecture of the Swiss environmental catalogue differs from that in other federal countries such as Germany and the USA. Unlike these two countries where the focus was on communication protocols between distributed databases, and since most of the cantons were already used to a common tool (WinCDS), Switzerland decided to build a central database with great flexibility in the management of remote users and groups.

Despite the centralized architecture of envirocat, special attention was given to the definition of a data model allowing further exchanges with other systems.

Envirocat runs on a Linux operating system (Redhat or Suse), data are stored in a PostgreSQL database and the application is built on Zope⁶ architecture. Zope was chosen because it offers a complete toolkit for the development of complex applications containing dynamic web pages, the management of access rights, upload / download modules, etc. Zope’s features and modules and its object-oriented language capacity allowed a complete multilingual environment to be built, proposing different interface options depending on the users’ rights (see figure 2 for an example of the user interface). Through this technological choice, a flexible and open source based application is proposed.

⁶ For more information on ZOPE see: <http://www.zope.org/>



Figure 2 : envirocat user interface

The basis of envirocat is a platform called IRMI⁷ (Information and Resource Management Interface). This platform is common to three applications: envirocat, the Swiss Biosafety Clearing-House⁸ and the Swiss Clearing-House Mechanism for Biodiversity⁹.

On request to the SAEFL, other administrations or institutions can obtain the envirocat application and modify it according to their own needs.

From a technical point of view, communication with other systems is of great importance. An interface with the Swiss Catalogue for Geographical Data (geocat.ch) will be developed during 2004. This interface will allow users entering the geocat.ch portal to search for geographical data within the envirocat database. The search will be carried out only within the *data* (GIS data) and *document* (printed maps) element types. The list of results will show the basic information about the dataset (title, description, format etc.). In order to get all information available, the original dataset

⁷ see Kleindienst H. : IRMI - An open metadata catalogue for information and resource management, EnviroInfo Conference 2004.

⁸ see <http://www.ch-bch.ch/>

⁹ see <http://www.ch-chm.ch/>

will be opened in envirocat. An interface in the other direction (from envirocat to geocat.ch) will be developed in a second step.

6. Metadata management and quality control

The main reason for developing a new application for the Swiss catalogue of environmental data was the need to have a web-based interface for editing metadata. Building up the content of the database is a time-consuming task. Sharing this task between all data providers is particularly important in Switzerland, as the collection of environmental metadata is done on a voluntary basis and partners, in general, do not have specific human resources dedicated to metadata.

Since 1997, with the former CDS, a semi-manual procedure has been applied to get data from cantonal partners and to merge it in a central database each time an update is required. The problem with this procedure was not only that it is tedious but also that the workload was assigned to a unique contact person in the partner office.

In the new envirocat system, by using user/group and rights management tools, partners can manage their own internal network of persons, which can comprise an unlimited number of authors, allowing better sharing of the work. Experience shows that some cantons still prefer to work in a more central way with one contact person gathering the metadata (smaller cantons), while others have created many different groups and authors who enter metadata into envirocat. The system allows both ways of organising.

This decentralised responsibility in the management of metadata implies a quality control strategy. Different features of envirocat should ensure data quality:

- The assignment of specific roles and the use of a workflow: authors can change and create metadata for specific groups and submit these to the editor of the group. The editor validates and publishes the metadata. Envirocat editors receive all information regarding guidelines and recommendations regarding data collection.
- Mandatory fields and the formats of some fields (i.e. numbers only) are automatically checked when the status of the metadata is changed.
- The envirocat project team supports the partners regularly with updates concerning recommendations published on the Internet. Newly published metadata are also checked by the project team.
- Changes in the contents of envirocat and the application itself are discussed in a yearly workshop organized by the project team.

Many details concerning the contents of envirocat still have to be defined in order to guarantee that there is an easily-understandable, high-quality metadata base¹⁰.

7. Current contents and user statistics

Envirocat gathers descriptions of environmental data collected by the Swiss cantons and by the Swiss federal offices.

The contents of the former CDS have been transferred into the envirocat system. However, the data quality was not entirely satisfactory, so the metadata were not directly published. The partners have now started updating the contents of the envirocat database. Currently, the envirocat database contains about 6,000 entries of which 750 have been revised and published so far. The overall amount of data will decrease compared to the 6,000 entries in CH-CDS. CDS contained a great deal of metadata that had to be deleted (double entries, incomplete entries). Approximately half of the entries in CDS were addresses and organizations. Envirocat however focuses on the description of data, addresses are only kept if they relate to a dataset.

The envirocat network connects thirteen Swiss cantons, five federal offices and federal research institutions and two cities. This corresponds to about 50% of all potential environmental data providers. The continuous enlargement of this network is a medium term goal of the project, while primary activities will be concerned with the qualitative and quantitative improvement of the contents of the envirocat database.

During the first half of 2004, envirocat had an average of 1,200 visits per month (the website was officially opened only towards the end of June), and around 15% of visitors logged on to the system. Analysis of the apache log file showed that around 1,500 search queries per month were submitted (80% using the standard search option – efforts for the development of enhanced search features seem to be quite great compared to their usage). In envirocat more than 50% of the metadata are viewed in detail. This is a great improvement compared to the analysis of CH-CDS, which indicates that only 10% of the queries lead to an opening of the detail view of an element. It seems that the functionalities of envirocat are better understood and/or that users more often find datasets matching their searches.

In order to tap the full potential of envirocat it will be necessary to acquaint possible user groups with the tool. Promoting envirocat within research institutions, private companies and the federal and cantonal administration is a mid term goal of the project.

The more data that is available, the more envirocat will be appreciated. Once this positive feedback loop is launched, sharing environmental information will become reality.

¹⁰ See chapter 4 of this paper

8. Conclusion

Envirocat is an example of a tool in which different aspects of sharing are of great importance. Sharing information can sometimes be quite challenging. In the case of envirocat, different problems had to (and still have to) be solved in order to put sharing into practice. From a technical point of view a new application had to be developed on condition that the former data stock could be transferred. The contents of the information system had to be more clearly defined – a process which will be ongoing. The network of partners had to be re-established and extended. The workload of updating the envirocat database had to be shared – these activities need careful coordination and communication.

Initial user reactions and the development of user statistics indicate very good acceptance of the envirocat tool. During a workshop held in Berne in December 2003, user requirements were refined, while initial experiences were presented to other envirocat partners. Still, people are not yet completely familiar with the importance of metadata, so the partners need to be motivated.