Ellipse – Long-term and Permanent Protection and Accessibility of Geodata

Krystyna W. Ohnesorge
Swiss Federal Archives
Archivstrasse 24
3003 Bern, Switzerland
Tel. +41 58 464 58 27
krystyna.ohnesorge @
bar.admin.ch

Chiara Marciani
Swiss Federal Archives
Archivstrasse 24
3003 Bern, Switzerland
Tel. +41 58 462 48 70
chiara.marciani@
bar.admin.ch

Alain Mast
Swiss Federal Archives
Archivstrasse 24
3003 Bern, Switzerland
Tel. +41 58 463 5074
alain.mast@
bar.admin.ch

ABSTRACT
Archiving of geodata historically focused on methods of keeping digital geodata alive "almost forever". Project Ellipse is a joint effort by the Swiss Federal Archives (SFA) and the Federal Office of Topography (swisstopo) running from 2011 to 2016. Its aim is to find a common solution for the archiving of geodata in order to implement the applicable legislation. Ellipse follows the entire archiving process chain: from the inventory and appraisal of geodata, to its submission to the digital archives and finally to the users, who expect geodata in a form that is authentic and accessible in a future technological environment.

Archiving of geodata is a complex task that demands intensive cooperation among all stakeholders. Despite our efforts, not all questions have been solved successfully. In this paper, we will report our findings and solutions as well as the obstacles we encountered during the course of Project Ellipse.

Keywords
Preservation of geodata, long-term availability, geoinformation system, Geo-SIP, Geo-Dossier.

1. INTRODUCTION
Geodata is (digital) information that identifies the geographical location and characteristics of natural or constructed features and boundaries on the earth's surface, typically represented by points, lines, polygons, and other complex features (vector data) or pixels (raster data). These descriptive items are not understandable if they are not linked to geospatial reference data, for instance to topographical maps. The combination of spatial orientation with other thematic sets of geodata or with geospatial reference data creates geoinformation. In today's modern society, geoinformation is the foundation for planning, measuring and decision-making at private and federal level. It is an integral part of state action and has to be preserved. Thus, Project Ellipse is developing a geodata archiving solution.

According to its remit, Ellipse shall achieve the following objectives:

- To develop an integrated solution for all geodata produced in the federal administration
- To achieve a worthwhile enhancement of the long-term availability of geodata and of archiving
- To allow geoinformation to be restored and interpreted from archived geodata at a later date

2. INFORMATION ABOUT THE PROJECT
The Archiving Act (ArchA)[1] and the Geoinformation Act (GeoIA)[2] require geodata produced in the federal administration to be preserved. For this reason, the Federal Office of Topography Swisstopo and the Swiss Federal Archives (SFA) were asked to develop a geodata archiving solution, in order to implement the applicable legislation. The scope of the project is limited to the geodata listed in the appendix to the GeoIO[3] (official geodata catalogue).

Between January 2011 and March 2013, Project Ellipse developed a concept for archiving official geodata. The concept describes the fundamentals for archiving geodata along the process steps production – geodata management – planning of conservation and archiving – acquisition (to the archive) – preservation – use. The main emphasis here was the collaboration between the producers of geodata and the Swiss Federal Archives. Based on the legislation on geoinformation, these institutions are required to mutually ensure the availability of geodata for the long term.

In the spring of 2013, the concept was approved by the SFA, swisstopo, as well as by the coordinating agency for federal geographical information (GCG), thereby initiating the implementation phase, which will be concluded by the end of 2016. Collaboration with cantons[4] and specialist organizations within both the archiving and the geoinformation communities will continue throughout this phase. The objectives were combined into four different work packages, each of which contains a range of tasks:

- **Work Package 1 – Conservation and archiving planning CAP**[5]: responsible for planning and conduction of appraisal of all geodata and for developing a tool to support this appraisal.

[1] https://www.admin.ch/opc/de/classified-compilation/19994756/index.html

[2] https://www.admin.ch/opc/en/classified-compilation/20050726/index.html

[3] The appendix can be found at: https://www.admin.ch/opc/de/classified-compilation/20071088/index.html#app1

[4] Switzerland is divided into 26 administrative regions called *cantons*.

[5] http://www.geo.admin.ch/internet/geoportal/de/home/topics/archive_planning.html

- **Work Package 2 – Formats and GeoSIP**: responsible for documentation and specification of archivable file formats and for developing a specification for a Geo Submission Information Package.
- **Work Package 3 – Access and Use**: responsible for creating and assuring access to geodatasets from archiving search platforms, and specifically for developing a link between the geometadata online catalogue (geocat.ch) and the archive metadata online catalogue (swiss-archives.ch).
- **Work Package 4 – Operational Organizations**: responsible for geo knowledge accumulation in the SFA and for developing and defining a second level support solution for end users.

Swisstopo assumed responsibility for work package WP1, the SFA leads the other three work packages. Below, the organisational structure of Project Ellipse is depicted:

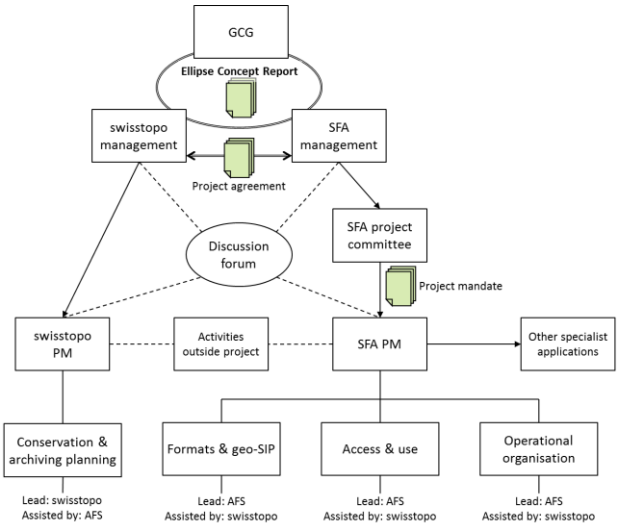

Figure 1 Project Ellipse organization

3. RESULTS

3.1 Conservation and Archiving Planning

A key issue in the management of geodata is to define what data must be available on which platform, to what purpose and for how long. In Switzerland, there is a legal framework for answering these questions, which distinguishes between *conservation* for a limited time at the authority responsible (long-term availability) and *archiving* for an unlimited time by the SFA.

The CAP includes all geodata sets and other spatial data sets of the Federation with the corresponding appraisal of the long-term availability and its value for archiving. The responsible offices of the Federal Administration conducted the first part of the appraisal. The SFA conducted the second part. The combined appraisal results identified the majority of geodata sets as archivable (313 out of 342 datasets). As most datasets are based on a decree, the need for their archiving was implicitly given. Furthermore, many datasets were built with a considerable effort and they promise broad potential for future use, adding another reason for their archiving. Of the remaining 29 datasets, 9 datasets were not considered for archiving, because they were products of data that was archived in a different, more suitable channel already. The remaining 20 datasets could not yet been appraised because either the datasets themselves or their juridical foundation was not yet completed.

3.2. Formats and GeoSIP

The work package *Formats and GeoSIP* concerns itself with file formats suited for the archival of geodata and with the definition of a submission information package for geodata.

With TIFF+Extended World File (TIFF+EWF.XML), a format for the archival of georeferenced image and graphic raster data has been defined and its specification has been published. The format consists of a baseline TIFF image and a simple XML sidecar file, which stores a minimum catalogue of attributes that permit the description with regard to space, time and content. Both files are linked by sharing a common filename that differs only in the extension.

The XML file contains ten attributes, of which six are the attributes that also make up a world file[6]. The remaining four attributes are:

- ReferenceSystem: Indicates the geographic reference system used in the form of a text reference in accordance with EPSG, for Switzerland "CH1903 / LV03" or "CH1903+ / LV95".
- BeginTemporalExtent and EndTemporalExtent: Temporal extent of the content of the geodata or best possible approximation of the period in ISO 8601 format.
- ImageDescription: An optional free text to describe the image.

The decision to specify TIFF+EWF.XML instead of using GeoTIFF[7] was driven by the fact that there is currently very little use of GeoTIFFs in the Swiss Federal Administration. This means that the barrier of introducing a simple but new format like TIFF+EWF.XML is actually lower than introducing an established, but more complex format such as GeoTIFF. Additionally, there is no standard set of metadata tags to be used in GeoTIFF which, in our opinion, further compromises long-term understandability of this format.

For archiving georeferenced vector data, no straightforward solution was found. In the Swiss Federal Administration, the products and thus the file formats of ESRI[8] (Shape and the Geodatabase family) are dominant. These proprietary formats however are not ideal candidates for long-term storage. Currently, there are only two candidate formats for archiving georeferenced vector data: On the international level there is GML, and on the national level, there is the Swiss format INTERLIS2[9]. It is important to note that the translation between any vector format is a challenging task, it is often ambiguous and therefore difficult to automate.

For lack of a better solution, INTERLIS2 will be named as format for archiving georeferenced vector data, since it is more widely used in the Swiss administration than GML. When submitting data in INTERLIS2, it will also be possible to additionally submit the most current ESRI format (of the same data), in the hope that this will facilitate the transition to other formats in the future.

---

[6] https://en.wikipedia.org/wiki/World_file

[7] http://trac.osgeo.org/geotiff/

[8] http://esri.com/

[9] http://www.interlis.ch/interlis1/description_d.php

The Swiss federal archives receive digital data for the archive encapsulated in SIPs (submission information packages). Initially in the project, it was planned to extend or adapt the SIP specification to assist archival and retrieval of georeferenced data. This plan will be postponed though (and thus pushed beyond the finishing date of Project Ellipse), as currently, several parallel undertakings to adapt or improve the SIP specification are on the way that must be streamlined to minimize impact on surrounding systems.

Instead, focus was shifted to the definition of a set of rules on how to organise geodata inside an SIP container in a structured way that takes into account the multi-dimensionality (time, space, layering) of geodata. The goal is to define a simple structure for storing geodata that is primarily understandable by humans, and secondarily aims toward automatic or semi-automatic machine-readability. The solution that will be proposed, while suitable for simple use cases, will be flexible enough to accommodate more complex use cases and geodata of different producers and systems. Additionally, it will be indifferent of formats and allow storing of primary data, necessary metadata and accompanying documentation. For that, we coined the term Geo-Dossier[10]. At the time of writing, a Geo-Dossier contains three first-level folders for storing documentation, models and primary data. It defines the mandatory splitting of data into subfolders if there are multiple views of the same data (e.g. multiple reference systems or multiple quality levels). It also allows for optional splitting of data into subfolders for arbitrary criteria (e.g. in thematic layers or spatial regions).

3.3 Access and Use

In this work package, requirements for the user interfaces were identified in order to enable the information retrieval system of the SFA to cope with geodata. Furthermore, geometadata required for archival purposes was selected and the linking between the access system run by swisstopo for long-term availability and the access system run by the SFA for archival was defined.

In the SFA, requirements for the user interfaces, for information retrieval and for search were defined in a series of workshops. A minimal set of requirements suitable for implementation within the currently operational access systems has been identified. Additionally, an extended set of requirements to be implemented in a possible future access system was defined. As the primary purpose of the access system of Swisstopo already is the handling of geodata, no additional functionality is needed there. In a next step, a set of geometadata to assist categorisation, search and information retrieval in the archival system of the SFA was selected:

- UUID: The identifier as it is used in geocat.ch, the metadata system of swisstopo (mandatory)
- Official geodata set ID: Another identifier that is mandatory for certain kind of Geodata
- Abstract: A short textual description of the data (optional)
- Preview: A thumbnail of a selected area of the geodata (mandatory)
- Georeference Data UUID: The identifier of a reference geodata set, if the actual geodata set is based on such (optional)
- Additional Georeference Data: If above UUID does not exist, a textual description of any reference geodata (optional)
- Geocategory: A categorisation based on the standard eCH-0166[11] (mandatory)
- Keywords: Keywords for describing the geodataset (optional)

By having the UUID as it appears in the long-term availability system of Swisstopo as a metadata, it is possible for both access systems to link to a dataset as it appears in the other system. Thus, users of one system can see that related datasets exist in the other system (usually the case if older data of a dataset has already been archived, while newer data is still only found in the long-term availability system).

3.4 Operational Organizations

The two activities of this work package consisted of appraising existing processes and identifying necessary changes in relation to acquisition and archival of geodata by the SFA, and of identifying and building up geo expertise inside the SFA.

The need for adjustment of the SFA-internal processes for data receiving and end user support proved to be negligible. Consequently, only a small internal workshop to convey basic geo knowledge for the SFA staff was conducted. Furthermore, a test run of the process of delivering geodata from the producer to the archival systems of the SFA was conducted between May and September 2016.

4. CONCLUSIONS

One of the key factors that have led to successful results was the broad guided dialogue we have had with all producers of geodata and the very efficient cooperation between all stakeholders. We maintained a high level of communication among SFA, swisstopo, the affected geodata producers and the Swiss geo community as a whole, which proved to be invaluable for the success of the project.

Undoubtedly the largest influencing factor in the project were the results of the CAP. Before the conclusion of the CAP, it was unknown how much, if any, geodata would be chosen to be archived and how long this data would remain in long-term availability before being sent to the SFA. Also unknown was the amount of data that the SFA had to expect and at what time intervals the data was to arrive, so it was difficult to judge the amount of automation that had to be built for ingesting.

The specification of a Geo-SIP was postponed, so that a more generic approach for a new and more flexible SIP specification can be developed, in which the accommodation of geodata will only be one part. This postponement however freed valuable resources for definition of the Geo-Dossier, a task that was not initially planned ~~for~~ but proved to be important.

The results of the work package Access and Use will influence the future of information retrieval in the SFA and will be a valuable input to the definition of the generic new SIP format.

The work package Operational Organisations has shown us that our processes are already flexible enough to accommodate themselves to various kinds of information, including geodata.

[10] Definition of the Geo-Dossier is still ongoing and results will be available at the end of 2016.

[11] Vgl. eCH-0166 Geokategorien, Version 1.1 (Minor Change) of the 23.09.2013, http://www.ech.ch

The main work that will go on past the conclusion of Project Ellipse at the end of 2016 is the

- definition of an archivable vector format and the
- definition of a new and more flexible SIP format which is better suited for various kinds of digital data, such as geodata or hypertext data.

Even with these activities still outstanding, we feel that Project Ellipse has successfully addressed the important aspects of archiving of geodata, and we are confident that with the current level of cooperation between all involved parties, the ongoing work can be adequately addressed.

5. REFERENCES

[1] Projektteam Ellipse, 2013, *Concept for the archiving of official geodata under federal legislation,* p.14. https://www.bar.admin.ch/dam/bar/en/dokumente/konzepte_und_weisungen/konzeptbericht_ellipse.pdf.download.pdf/concept_report_projectellipse.pdf

[2] Schweizerisches Bundesarchiv, 2016, *Bewertungsentscheid Geo(basis)daten des Bundes*. https://www.bar.admin.ch/dam/bar/de/dokumente/bewertungsentscheide/Geobasisdaten%20Bewertungsentscheid%202016.pdf.download.pdf/Bewertungsentscheid%20Geo(basis)daten%20des%20Bundes%20(Projekt%20Ellipse,%20AAP),%202016-02-19.pdf

[3] Schweizerisches Bundesarchiv, 2015, *Merkblatt Spezifikation archivtaugliches Geoformat für Bild- und Grafikrasterdaten TIFF + EWF.XML*. https://www.bar.admin.ch/dam/bar/de/dokumente/kundeninformation/Merkblatt%20TIFF%20EWF%20XML.pdf.download.pdf/Merkblatt%20TIFF%20EWF%20XML.pdf