

## Werk

**Titel:** Archiving Electronic Information: Some Aspects

**Autor:** Bütikofer, Niklaus

**Ort:** Graz

**Jahr:** 1995

**PURL:** [https://resolver.sub.uni-goettingen.de/purl?514854804\\_0005|log71](https://resolver.sub.uni-goettingen.de/purl?514854804_0005|log71)

## Kontakt/Contact

[Digizeitschriften e.V.](#)  
SUB Göttingen  
Platz der Göttinger Sieben 1  
37073 Göttingen

✉ [info@digizeitschriften.de](mailto:info@digizeitschriften.de)

# **Archiving Electronic Information: Some Aspects**

NIKLAUS BÜTIKOFER  
*Swiss Federal Archives, Bern*

## **Introduction<sup>1</sup>**

Everybody who has been using computers for some years is no stranger to messages from the suppliers like: "Now, we are able to offer you our new enhanced system. Unfortunately, our support of the old system cannot be guaranteed later than 1996. Due to engineering changes the new system is not downward-compatible". Such messages are the beginning of the inevitable end of your system. You have to buy a new system and to migrate your applications and your data. By doing so you are likely to lose not only a lot of money, but also a lot of information. This example illustrates what difficulties archivists and librarians are facing today in dealing with electronic records, electronic books or electronic maps.

The revolution in computing and communication brought us new fascinating information products and more comfortable ways to disseminate these products. The revolution is steadily going on, it is transforming our work, our profession, and our institutions. We are all aware of that and cannot even imagine how our archives and libraries will look like in twenty years from now.

In this article I would like to show you how an archivist is approaching the issues emerging from the revolution of data processing. As an archivist of the Swiss Federal Archives I have to deal with records produced by the bodies of the Swiss Confederacy in conducting their official business. Records are in many ways different from books and maps, no matter if they are stored on paper or on digital media. Records give, or should give, evidence of communicated business transactions. In order to understand a record you have to know not only its content and structure, but also the context in which the transaction has taken place.

<sup>1</sup> I thank Hugo Schwaller for his helpful review of this paper.

I do not want to stress the difference between archives and libraries. On the contrary, I believe that our issues and our methods in the electronic age are converging, as I will point out when preservation of online GIS (geographical information systems) are treated.

### Objectives of archives

The mandate of archives is essentially to preserve information with a permanent value over time. There are three fundamental requirements for the preservation of information:

- 1) The preserved information must continue to be **accessible and retrievable**. It must be possible to find the information searched for and to output it. These requirements address principally the technology used for preservation and the finding aids available for research.
- 2) The preserved information must continue to be **understandable**, so that it can be correctly conceived even in hundreds of years. This requirement addresses principally the relationship between elements of information content.
- 3) The preserved information must continue to be **authentic** in a way that a user can be sure that the information he is getting through the system is the same information which the author originally has created.

While these requirements are quite trivial for long term preservation of paper records and paper books, they become critical in an electronic environment.

Below, I would like to point out on a general level, problems, consequences, and possible solutions. The first part is focused on preservation issues regarding technological change, the second concentrates on the issues of historicity and authenticity as they arise mainly in online databases like GIS.

### Preservation regarding technological change

In order to make electronic maps accessible and retrievable, e.g. on a CD-ROM, one needs a complete system that is capable of reading the CD-ROM and finding and displaying the view of the map the user is looking for. Consequently, we have to preserve not only data on a storage medium, but a complete system if we want to keep electronic maps accessible. By rigorously reducing the complexity of the matter in hand, we can say that the system we need consists of data, software and hardware. Each of these elements has its own life cycle. Figure 1 shows a possible succession of hardware and software, as they become obsolete within 3 to 10 years. Reality, however, proves much more

complicated; there you have a lot more components that are undergoing technological change in their own rhythm. Data for permanent preservation must follow the innovation cycle of hardware and software, and it is necessary to convert the data time and again to new technological environments.

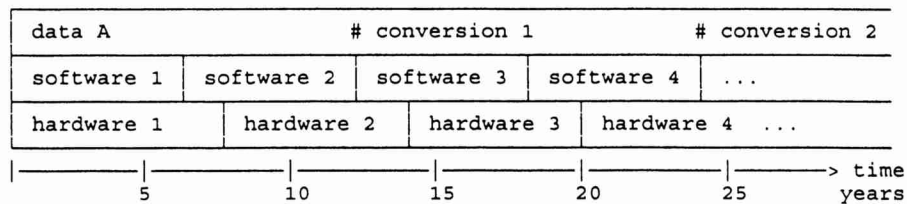


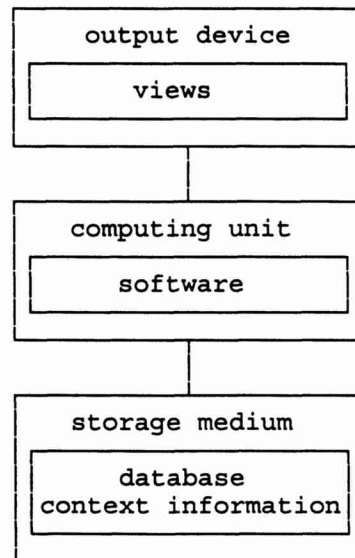
Figure 1: Succession of hardware and software generations

Although the time scale I use in Figure 1 is pure speculation, we can draw an important conclusion from it. We have to understand that we will not escape the necessity of spending a lot of our resources on converting data and migrating applications. I cannot see any possibility of archiving computers and software and keeping them running our electronic maps. Components of a computer do not have a long life, they must be replaced from time to time, and if each single component of a computer must be manufactured outside mass production, this becomes so expensive that we cannot afford it any longer.

If we cannot preserve computers and software we can, at least preserve what is essential, and that is information. It may already be commonplace that electronic information is not bound to a physical medium. A given piece of information can be transferred or copied from one medium to another with no significant consequences whatsoever at the logical level.

Information in electronic maps consists of all possible views of a given database and - as an archivist I have to point this out - of knowledge of the provenance, i.e. the context of the production of the database in order to be able to judge the reliability of the views. (Figure 2)

Figure 2:



There are three principal different ways to archive such a system:

- 1) Create and capture all possible views of an electronic map, or at least the most important views, and print them on paper or microfilm. This is the traditional method of archiving. The views can be preserved by conventional means, but the users will not be very happy with it because they all would like to work with computers and, consequently, they have to digitize the view whenever they want to work with it.
- 2) Preserve only the raw data in a form that is software independent as far as possible. Future users will have to import the data into their own system whenever they want to work with them. By archiving only raw data we are probably losing the original views because future systems will not have exactly the same software functions as the original system, and they will, therefore, not be able to create the same views as the original system.
- 3) Convert data and software functions into the successor technology whenever a system has become technologically obsolete. This method provides us with the possibility of shaping succeeding systems so that they can create views equivalent to those of the original system.

Each method has its advantages and its shortcomings. Method 1) may be a practicable way to archive systems with a complicated database on the one hand and few possible views on the other. Method 2) may be good enough for systems with a very low frequency of access, while method 3) yields the most comfortable result of archiving. But this last method is also the most expensive one.

### **Keeping historicity and authenticity**

Electronic maps distributed on removable storage media will probably only be of transitory relevance. They will soon be succeeded by huge online GIS which are given access by using public networks. GIS consist of numerous datasets from different sources, and these datasets are continuously updated.

Such systems present both the archivist and the librarian with an enormous challenge. How can the historicity of information be kept and how can evidence of information authorship be kept? Both are essential for scientific work and for any future use.

There are, as far as I can see, two different ways of keeping historicity in GIS:

- 1) We can periodically make a 'snapshot' of the whole database and preserve the state of the database at that moment. After a certain lapse of time there will be a series of databases in our archives that will allow the reconstruction of historical change and evolution.
- 2) We can implement a facility in our GIS which automatically writes a record of every modification of the database in a history file. This file would allow us to reconstruct exactly the state of the database at any time we want. This is an essential requirement for systems which serve as a basis for official decisions. For audit purposes responsible authorities have to be able to show what information has been available when a certain decision was taken.

Authenticity as well as historicity is an important point in big GIS which contain information from different sources. Users both now and in the future want to know by whom and when the information they are reading was produced, and moreover, they want to be sure that the information is the same the author originally produced, and the same a colleague made reference to in his footnote. This is not trivia in a system that is continuously updated.

There are two means which both can ensure authenticity:

- 1) Secure the system from unauthorized alteration.
- 2) Create metadata of each modification so that we know author, time, and content of each modification.

**Role of archives and libraries**

To conclude I would like to ask about our role in the field of GIS. GIS usually are not produced and updated by or in archives and libraries. They are maintained by other profit or non-profit institutions which normally provide access to users. What roles should archives and libraries play in these circumstances? I can see four different roles:

- 1) Do nothing and concentrate on paper fonds. Leave the problem altogether to those who are producing and running GIS.
- 2) Transfer GIS to archives or libraries when they are abandoned by their producer and preserve them. In this case we are not able to ensure historicity and authenticity of the GIS because we have to take what we get, and that is the final state of the database.
- 3) Acquire regularly 'snapshots' or history files from producers and preserve them together with the initial state of the database. This method should be based on an agreement with the owner of the GIS.
- 4) Charge producers of GIS by law or by payment to ensure historicity and authenticity of their system and to provide access to any user. In this case, archives and libraries would reduce their functions to information locator services providing only reference and access information of available information systems.

There can be no question what I recommend to do. Only by playing role three or four can we preserve our profession over time.