

Les nouveaux enjeux de la conservation du patrimoine documentaire /  
New challenges for the preservation of documentary heritage  
<https://sierre2017.wordpress.com/>

## Preserving cultural heritage: Better together!

**Barbara Signori**

Department e-Helvetica, Swiss National Library, Bern, Switzerland



Copyright © 2017 by Barbara Signori. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:  
<http://creativecommons.org/licenses/by/4.0>

---

### Abstract:

*The Swiss National Library has a mandate to collect, catalogue, store and disseminate the cultural heritage created in Switzerland and abroad by and about the Swiss, both in print and digital. This sounds like a clear enough mission, but dig deeper and this mandate raises all sorts of tough questions.*

*What exactly is cultural heritage? Obviously, it goes far beyond e-books and e-journals of well-established Swiss publishers. It is Swiss websites, newsletters of Swiss societies, and so on. However, what about all the digital data that is created by Swiss people every waking moment? The selfies, blogs, tweets, social media, personal digital archives. Surely not everything can be considered cultural heritage. But who decides what is and what isn't?*

*And then how do we cope with the enormous quantity of information being produced? How can we decide what to keep for future generations when we cannot even cope with the output of the current generation? Not to mention the costs. With budgets being cut all the time, what does that mean for our cultural heritage?*

*Amidst all these tough questions, one thing is clear: no single institution can possibly cope with collecting all that information nor be tasked with the decision on what to preserve and what not.*

*This paper will use the example of Web Archive Switzerland to show how trust and interoperability have led to constructive collaboration. Web Archive Switzerland was born in 2008 following 5 years of discussion with the cantonal libraries. Since then websites with a bearing on Switzerland have been selected, documented, preserved and disseminated collaboratively among 30 Swiss institutions.*

*The key lesson learned over the past 14 years is that to answer the tough questions and challenges we had to look beyond our own walls and borders. We learned to let go of the idea that we can do it alone, that we can control the world of content through clever curation. We learned how to create partnerships and strong networks of institutions, how to engage new sorts of curators, how to trust each other and share synergies and costs, all with the common goal of saving as much digital heritage as possible.*

*In summary, this paper is a call to arms to join forces, to forge partnerships, to bundle competences, and to build collaborative networks! It will show that curating collaboration between institutions is as important as curating cultural heritage and it will suggest ways forward to create more collaborative collections of digital cultural heritage within Switzerland and beyond.*

**Keywords:** preservation, cultural heritage, collaboration, Switzerland, web archiving

---

## **Introduction**

*What is cultural heritage?*

What exactly is cultural heritage? Wikipedia [1] says “Cultural heritage is the legacy of physical science artifacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations.” UNESCO [2] outlines the following categories of cultural heritage: tangible cultural heritage, movable cultural heritage (paintings, sculptures, coins, manuscripts), immovable cultural heritage (monuments, archaeological sites, and so on), and underwater cultural heritage (shipwrecks, underwater ruins and cities). From a national library perspective, cultural heritage is all that a country and its inhabitants produce as memories in text, sound and vision no matter in what medium. This is easy to imagine and define in the print world, but the scope of cultural heritage in the digital world is wide open to any memory published online in any way. Thus digital cultural heritage goes way beyond e-books and e-journals of well-established publishers and includes websites, newsletters and so on. You can go a step further and include all digital data that is created by the people of a country. Consider the selfies, blogs, tweets, social media and personal digital data like photographs that are created every minute of the day. Or think about the emails that are sent. Of course these were formerly known as letters! Where does cultural heritage end? Where are the borders and boundaries?

Some very interesting projects have emerged recently on personal digital archiving. Nestor [3] in Germany for example has established a task force dealing with the issue and in Italy, Arca Memorie [4] aims to curate the memories of citizens there. Indeed, every institution that has a mandate to collect and preserve the cultural heritage must define the scope and boundaries of their work, dictated either by its culture and history or by its financial resources and capabilities.

*What are the challenges with digital cultural heritage?*

Defining these boundaries is one of the many challenges that memory institutions must face when aiming to preserve digital cultural heritage. The enormous amount of data forces us to take decisions on scope, scale and costs. However, the real challenge begins only once you have set your boundaries: the preservation itself! Information on paper is relatively easy to preserve as long as you have stacks under the right climate conditions. Digital data is much more demanding. It takes much more effort and ongoing maintenance to keep digital data available and readable over time. In a recent article by Richard Kemeny [5] Martin Kunze, who is very skeptical of digital preservation, was interviewed. He firmly believes that our present memories will be lost forever if we choose to rely only on digital preservation. Kunze has developed a preservation strategy based on ceramic plates. He notes that “clay is the material that will last longest....the only thing that can threaten this kind of data carrier is a hammer!”. His goal is to create a ceramic time capsule of human knowledge (Archive of Memory of

Mankind, MOM) [6] and bury it in a salt mine in Austria. For a small donation anyone can engrave their own memories on a table, whether it is an item from an art collection, a photo' of your wedding or your CV [5, 7].

It is interesting to note that Kunze has widened the scope of his project and is looking for more content, more knowledge and more co-operation. This is an important point. No matter what kind of preservation you are aiming, the urge to collaborate comes automatically. National libraries should accept the challenge laid down by Kunze and show the world that libraries are time capsules themselves, a point made by Richard Ovenden in the same article [5]. If we take this role seriously, our descendants will not only find information on ceramic plates or ceramic microfilms but also in our long-term archives.

### **Web Archive Switzerland**

The Swiss National Library takes its role in long-term preservation very seriously and invests a great deal in digital preservation. The strategy [8] states: "Our collection is analogue. And it is digital. It is quick to find and easy to use". The Swiss National Library commits to preserve all information related to Switzerland no matter if it is printed or digital. The Library has been doing this for printed material for over a hundred years and for digital material since 2001.

The Archive was established using a step-by-step approach in a series of projects covering e-books, e-journals and e-theses. We worked closely together with the producers of the digital data, for example publishers and universities, and established processes as well as the technical infrastructure for ingesting and storing digital data. Once these experiences were embedded into our working culture and became routine, we felt ready to explore the (for us) unknown world of web archiving. We decided to join IIPC, the International Internet Preservation Consortium [9] in order to learn from the experiences of its members and to contribute what we could.

IIPC today counts more than 50 members that all are involved in archiving parts of the World Wide Web. It is a great source of expertise and knowledge. Alongside international cooperation, we also looked for national co-operation within the borders of Switzerland. From the start, we had realized that a single institution could never archive the Swiss Web on its own. Switzerland is divided into provinces called Cantons and our choice was to join forces with the Cantonal Libraries since they have a similar mandate to ours, namely to collect and preserve the cultural heritage of their Cantons.

We invited the directors of the 26 Cantonal Libraries to Bern in 2013 and shared with them our ideas for creating Web Archive Switzerland. It was an historic event. As we explained the concept that cultural heritage hides in the Web and that it is important to collect websites, our ideas were cautiously welcomed. We received a mandate to develop our ideas further and to start a pilot project for which 11 Cantonal Libraries joined us. The goal of the pilot was to establish the basis for collecting, cataloguing, archiving and disseminating Swiss websites [10]. Together we defined selection criteria and cataloguing rules. We aimed for a shared workflow in which our partners would be responsible for the selection and description of the websites, while the Swiss National Library would be responsible for checking rights and permissions, harvesting and archiving the websites, and giving access to them.

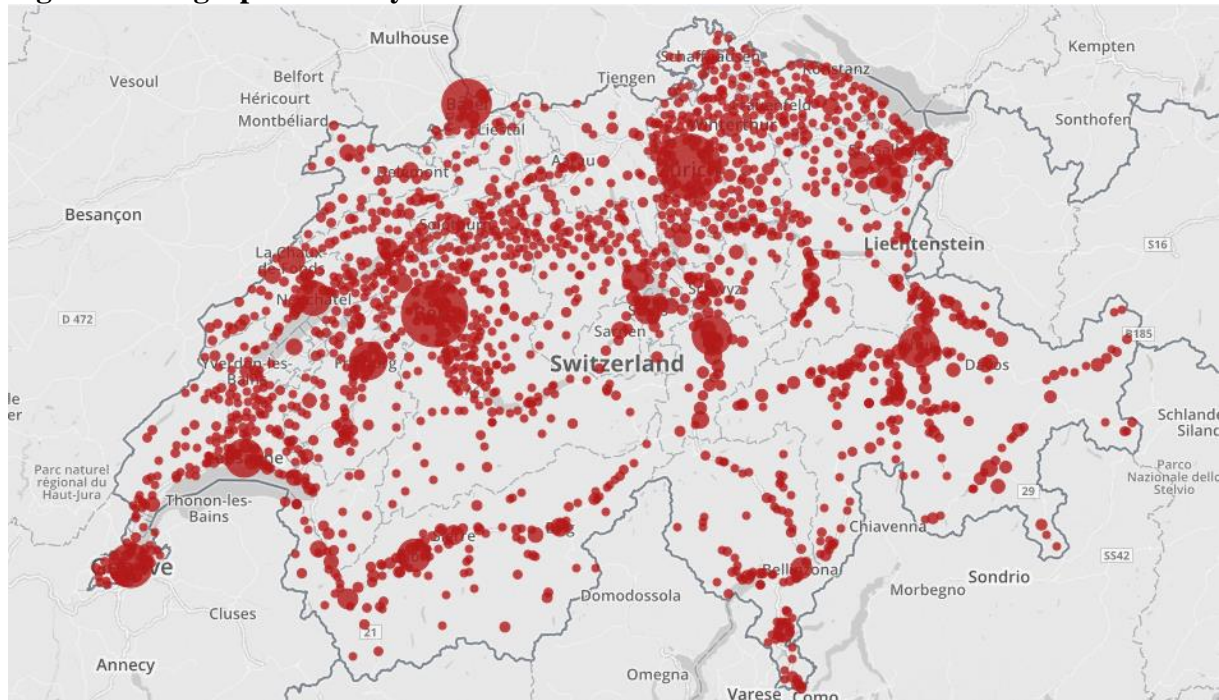
The Swiss National Library built the harvesting infrastructure in-house and set the archiving rules. We integrated the Wayback Machine into our access system. In 2008, the systems were

operational and Web Archive Switzerland was born. Over the years, more partners have joined us and today in 2017 Web Archive Switzerland has 30 members [11]. To date we have archived approximately

7'800 websites,  
27'300 snapshots,  
350 mio. URLs.

A recent geographical and thematic analysis showed that the websites are widespread throughout Switzerland [Figure 1] while the major topics are: public administration; geography and travel (Switzerland); sports; political science; music; economics; commerce, communications and transportation; education, school systems and teaching; recreational and performing arts; as well as social problems and services. Some topics are missing completely however; there is still much to be collected and archived.

**Figure 1. Geographical analysis of Web Archive Switzerland**



Looking back over the nearly 10 years since its inception, it is clear that Web Archive Switzerland was only possible thanks to collaboration. Collaboration outside the country to learn from others and collaboration within the country to make it all happen.

## **Challenges**

### *Scope*

The Swiss National Library has had the mandate to collect and preserve Helvetica [12] ever since it was founded. The term Helvetica is intended to cover all publications with a bearing on Switzerland, for example: the entire fiction and non-fiction literature published in Switzerland, all publications published abroad dealing with Switzerland and its inhabitants; all works and translations of works by Swiss authors in all languages and so forth. The amount of data that must be collected has increased tremendously since the emergence of the Internet.

Websites are a big part of this – after all it is how most communication and information exchange takes place today. It is how publishers, societies, companies, authors present themselves.

We see an increasing amount of previously printed material that is appearing now only as online content. Take for example the Memoriav Society [13] that has the goal to look after the audiovisual cultural heritage in Switzerland. It no longer prints its annual report anymore, but simply posts it to their website. It is obviously vital to develop strategies to archive this “born digital” information.

We have defined selection criteria guidelines [14] together with our partners, which each institution is free to add to with more detailed criteria if necessary. We chose to focus on websites that are freely accessible on the Internet. We chose not to collect data on intranets or data that is in some way access controlled or protected. The requirement that the information collected has a bearing on Switzerland and the Swiss is central to our aims, but we do not draw the line at our own borders. Thus, websites about Switzerland with server addresses abroad can be as relevant to us as a website hosted in Switzerland.

Selection is not easy. In fact, it was a completely new task for us. In the past, the Swiss National Library had formal criteria only for printed books and serials. For example, how many pages does the publication have? How many copies have been published? The content of publications was never questioned in the way that other libraries or archives often do. Hence, we needed to develop a way to evaluate the relevance of the websites we wanted to archive.

Together with our partners, who are specialized in their regions and their topics, we defined criteria, initially on the content but eventually we considered aspects like navigation, construction and layout of the websites. It is possible to judge a website as a whole by attributing votes to the different categories. This method helping us in decision-making was adapted from the Deutsche Internetbibliothek [15], a project started in the 1990’s to offer an information portal with over 6’000 links. Currently, we have 4 major approaches when selecting websites as follows:

- 1) Websites about the Canton: These are selected websites about things that are significant for a Canton. That could be the Matterhorn for the Canton of Valais for instance or the embroidery of St. Gall. In this area we knew from the start that we could not achieve anything like comprehensiveness. The goal is to bring as much Swiss heritage into the Web Archive as possible.
- 2) Websites representing the Canton as an authority: These are all the websites about a Canton, and its communities and major cities. It comes very close to the traditional library collection of printed official publications from national and local governments. In this area, we try to achieve a comprehensive collection with the goal that all administrations are well documented over the years.
- 3) Websites representing the Government: These are all the websites of the Swiss government and the websites of the governments departments and offices. Most of those websites are very large so we chose to set the harvest frequency as every 4 years, except in cases of major changes happening within a department.
- 4) Websites to events or other special collections: These are selected websites about major events happening in Switzerland or that are related to Switzerland, for instance the Euro2008 football championships or Swiss elections. They can also be about a specific topic, for instance literature or economics. This where our new specialist archive partners help us.

They are experts in their fields and therefore the best for selecting websites in their own topic areas.

Curators have a major role and it is a responsibility that weighs heavily on some of them. How can we here today decide what websites our descendants will find in the archive? How can we know what will be interesting for them? It is important to realize that we will never know. Instead, we need to rely on what we know today and what is relevant today. It is also very important that we have a mix of archived content because this best reflects the diverse society we live in. We should not try to save only the scholarly information that has scientific value. Social media and gossip are part of our daily life and if we want to show to future generations how we used the web this needs to be part of our collection. We also believe that we should be as comprehensive as possible when collecting a website - as long as legal constraints do not prevent us from doing so.

The other widely adopted approach is to do entire domain crawls from a particular point of view. For instance, we could have chosen to crawl the entire .ch domain. This approach requires no selection decisions others than the choice of which domain to crawl. However, you are effectively handing over a single mass of information to future generations in the hope that they will know better how to deal with it and what to preserve and what not to. We believe that this is too simplistic a view and that we have a responsibility to manage our current and future resources in a responsible manner.

Making selection decisions together with partners, helps share the responsibilities. We discuss the content of our archive in regular workshops. We spot holes in the archive by doing geographical and thematic analysis and take steps to fill them by actively recruiting new partners from new fields with new interests and new ideas. It is only because we have so many experts & specialists from different regions and in different topics that we are able to create such a rich and varied collection. A single institution could never achieve this on its own.

### *Scale*

Scope is not the only dimension that determines the size and manageability of an archive. Having decided on the extent of our collection our next challenge is scale. A printed book is a discrete object with a fixed number of pages. A website is an infinitely expandable bucket of information. The decision to collect a website is the start of a never-ending story. It means that a decision to collect is not simply a question of prioritizing today's resources but it is also a commitment to future investment to keep the archive in sync with the ever-updating websites.

Scale is not only a challenge for selection, but it impacts the harvesting and archiving and even the dissemination. Single websites can be so huge, that we are unable to parse them through our normal ingest system. At one point, our web collection grew so fast that we were unable to index the objects for a full text search without blocking the rest of our digital collection.

For the Swiss National Library scope is an ongoing challenge. We still have to learn how to deal with this mass of data since one thing is certain: it will not decrease. The only way we currently have to keep scope somewhat under control is the amount and the regularity of the website snapshots. Many other web archives are able to harvest websites on a weekly or even daily basis. Unfortunately for most of our collection we are only able to archive once a year. There are few exceptions for which we harvest twice a year, but for some very large websites, for instance the Government ones, we harvest only every four years.

Our system is currently configured so that the entire website is re-collected. This results in a lot of duplicate information in our archive. Harvesting only the differences over time would greatly help with our scale challenge, but it offers new challenges when displaying the content to the user.

Access is given through a common interface [Figure 2] that is also used by our partners to show Web Archive Switzerland. The whole archive can be searched or filtered by partner if desired. From the user perspective of course the more information that is in the collection the more interesting it is for them; the bigger the scale, the better for users now and in the future.

**Figure 2. Website search in e-Helvetica Access**

The screenshot displays the 'e-Helvetica Access' search interface. At the top, there is a header with the Swiss National Library NL logo and navigation links. The main content area is titled 'Advanced full-text search - Websites'. It features a navigation bar with buttons for 'All (131214)', 'Books (38086)', 'Serials (50923)', 'Websites (26392)', and 'Dissertations (15813)'. Below this, there is a 'Search in full-text' section with input fields for 'Search Term' and 'Search Phrase', and a 'Search filters' section with options for 'Language' and 'Year'. At the bottom, there is a 'Search with Dewey Code' section with a list of categories.

We also work internationally on the scale challenge. Within IIPC we have created a Curator Working Group and together we build collections on defined topics. Each country focuses on its relevant websites for the defined topic. In this way we can collectively scale the information collected on these topics.

## *Cost*

For the Swiss National Library as for many other national libraries around the world, the main factor holding back growth in scale and scope of their digital archives is cost. In many ways, cost is a challenge that you wish you did not have to consider when dealing with digital cultural heritage. Surely preserving our heritage should be a given thing.

However, even if a library had escaped the budget cuts that so many are facing, we believe that there is an obligation to spend precious resources wisely. Digital preservation is in many ways a blank cheque for future generations; our descendants will pay for decisions that we make today. No one really knows how much it will cost in future to keep all the data alive and accessible.

The collaboration in Web Archive Switzerland means that we are able to share resources and expertise. Doing it together bundles our collective forces and saves us all money. However, it is important to realize that keeping a network together requires effort too.

There are several international projects dealing with the question of costs, for example the 4C project [16]. The Swiss National Library covers the costs of web archiving itself. Only one partner institution makes a monetary contribution for archiving their websites, although all partners contribute resources in the form of staff time. We may consider different frameworks for the future, for instance charging a membership fee. However, we are very sensitive to the fact that costs should never stop us from caring about digital curation. It would be a worst-case scenario were we to ask for a contribution from our partners and as a result have some of them stop contributing websites to the archive. What we do not collect today will be lost for tomorrow.

We believe that it is very important to convince all stakeholders of the importance of archiving the web. Although awareness is much better than it was 15 years ago, it is still looked at as a niche task in many libraries. This is yet another area where national and international collaboration can help. The bigger your network is, the more strength and power you will have in persuading the key decision-makers of how important digital preservation and web archiving is. For instance, the Swiss National Library organized the international iPRES 2016 conference [17] and invited the President of the Swiss National Council and a very well known keynote speaker who attracted great press interest.

However, there is still a long way to go for us in Switzerland. An article [18] in the Swiss daily newspaper *Der Bund* published in May 2017 highlighted the successes of Brewster Kale and the Internet Archive in web archiving. Unfortunately, the journalist did not reference our efforts in Web Archive Switzerland albeit on a much smaller scale. We have to work harder so that Switzerland and the Swiss become aware of the fact that we have our own Internet Archive in Switzerland.

## **Conclusions**

The key lesson we have learned over the past 14 years of Web Archive Switzerland is that to answer all the tough questions and challenges we faced, we had to look beyond our own walls and borders. We learned to let go of the idea that we can do it alone. We learned that we could not control the world of content through clever curation. We learned how to create partnerships and strong networks of institutions. We learned how to engage new sorts of curators. We



learned how to trust each other and share synergies and costs, all with the common goal of saving as much digital heritage as possible.

We learned that web archiving is a time consuming and expensive thing to do. It is imperative to have full management support before starting to build a web archive. It really is a never-ending story, a commitment to the future. It is inconceivable to close a web archive once you have started. Moreover, a collection will become useless if it does not grow and is not maintained. Web archives become interesting only over time. They become indispensable when websites disappear from the live web. For these reasons, the decision to launch a new web archive should never be taken lightly.

We learned not to be perfectionists. Even if a snapshot is not rendered perfectly in the Wayback machine it is still worth to keep track of it. One thing is certain: that instance can never be retrieved if you delete it. It will be lost. It is always better to archive something.

We learned not to try to be comprehensive – it's simply not possible. There are still so many technical challenges to be solved in crawling and not every website can be downloaded entirely. We learned to live with holes.

We learned not to try to solve all the technical and legal problems regarding access before collecting. Our guiding principle is what you don't collect today will be lost tomorrow. Therefore, it is more important to have the website in your collection and to solve technical and some legal access issues later.

### **Looking to the future**

There are still many more ways to collaborate on digital preservation in the future. For instance, perhaps collections could be linked with each other to increase the scope covered. It is highly likely that there is a lot of Swiss cultural heritage hiding in other archives. Content that is not obviously Swiss at first sight or not on a .ch domain, but nevertheless relevant. And vice versa of course. We have content in Web Archive Switzerland that is relevant to other collections.

We would realize valuable network effects if we open up our archives to each other and let users search across all of them. Each archive that joins the network increases the value of the whole network. It is also a much smarter use of resources since it means that not every piece of data needs to be stored in multiple places.

We should stay open-minded and broaden the definition of cultural heritage in traditional memory institutions. We should build networks that widen our scope be that in text, video, sound, personal data, websites, apps or any other medium.

When it comes to preserving cultural heritage, we really are better together!

### **References**

[1] [https://en.wikipedia.org/wiki/Cultural\\_heritage](https://en.wikipedia.org/wiki/Cultural_heritage)

[2] <http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage/>

[3] <https://wiki.dnb.de/display/NESTOR/AG+Personal+Digital+Archiving>

- [4] <http://www.arcamemorie.net/>
- [5] <https://www.theatlantic.com/technology/archive/2017/01/human-knowledge-salt-mine/512552/>
- [6] <https://www.memory-of-mankind.com/>
- [7] <http://www.derbund.ch/leben/gesellschaft/festplatten-aus-ton/story/22640552>
- [8] <http://www.nb.admin.ch/org/00779/index.html?lang=en>
- [9] <http://netpreserve.org/>
- [10] [http://www.nb.admin.ch/nb\\_professionnel/01693/01695/01705/index.html?lang=de#sprungmarke0\\_5](http://www.nb.admin.ch/nb_professionnel/01693/01695/01705/index.html?lang=de#sprungmarke0_5)
- [11] For a full list of members see here: Webarchiv Schweiz : Grundlagen, Version 1.15, 22. Mai 2015  
([http://www.nb.admin.ch/nb\\_professionnel/01693/01695/01705/index.html?lang=en](http://www.nb.admin.ch/nb_professionnel/01693/01695/01705/index.html?lang=en))
- [12] <http://www.nb.admin.ch/sammlungen/helvetica/index.html?lang=en>
- [13] <http://memoriav.ch/>
- [14] Webarchiv Schweiz : Merkblatt Sammeln, Version 1.8, 22. Mai 2015  
([http://www.nb.admin.ch/nb\\_professionnel/01693/01695/01705/index.html?lang=en](http://www.nb.admin.ch/nb_professionnel/01693/01695/01705/index.html?lang=en))
- [15] [https://de.wikipedia.org/wiki/Deutsche\\_Internetbibliothek](https://de.wikipedia.org/wiki/Deutsche_Internetbibliothek)
- [16] <http://www.4cproject.eu/>
- [17] <http://www.ipres2016.ch>
- [18] <http://www.derbund.ch/digital/internet/Warum-peinliche-Partyfotos-nie-fuer-immer-geloescht-sind/story/28957975>