ÖSTERREICH AM WORT: A PROJECT GIVING WEB ACCESS TO MORE THAN 5,000 RECORDINGS

Hermann Lewetz (Austrian Mediathek, Vienna, Austria)

This is the story of a project from the Austrian Mediathek (AM), a mid-sized institution with twenty-two employees.

1.   The project

Submitting concepts for projects is one of our strategies for acquiring extra funding to digitize our immense holdings. One recent project was Österreich am Wort (OaW). The main goal of OaW was to put selections of recordings on the website that represent each of the numerous collections stored in the archive of the AM.

In the world of scientific research, audiovisual recordings are often ignored as primary re-sources. One reason is the difficulty for scientists to get continuous access to them. One aim of OaW, therefore, was to publish potential sources in a way that would secure persistent access for researchers who want to use these sources in their work, as well as for those who want to evaluate the primary sources used by these researchers.

When we proposed OaW, we aimed to publish approximately 1,000 video and 4,000 audio recordings during a period of three years.

We used some of the project resources to finance additional equipment, especially for the vid-eo digitization. But above all, the project resources went towards developing the infrastructure for managing metadata and, especially, for clearing copyrights. The digitizing of audio recordings and part of the video digitization was done with existing resources at AM.

The project ended 1 September 2012. The result can be accessed online (http://www.oester-reich-am-wort.at/).

OaW was funded with money from the federal government of Austria's Department of Education, Arts, and Culture.
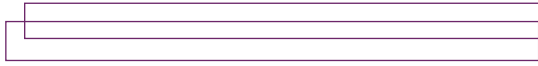
2.   The plan

The plan was that two persons would evaluate all collections and make a list of potential recordings that could represent the content. With the first selections in hand, another team would start to clarify the copyrights. Then the workflow of digitization would be initiated.

While the AM has been digitizing audio for more than ten years and, therefore, was ready for the task of digitizing large collections of audio recordings, this project marked the begging of the AM's foray into the world of video digitization. The way forward looked quite simple. There was an eagerness to purchase and use the recommended video system (SAMMAsolo), and the project offered the funding. With the new equipment in hand, we just had to start digitizing the selected video recordings.

Another team would enrich the metadata in the catalog. Meanwhile an outside company would design, program, and install a website in which not only the new, but all the AM audiovisual content would be presented. The new idea was to present each recording in an individual page, instead of having several items on one site as we did before. This way, it would be possible to generate an individual link for each recording. In the era of Europeana and other large search aggregators and in light of the Internet's reliance on linked content, this became an essential feature.

This change in our web structure resulted in a complete new architecture that also required a complete update of existing web exhibitions.

We generated web copies of the audiovisual resources from the digitized archive copies. Although most of such projects are funded only to give web access to content, we insist on digitizing in a way that is equally adequate for long-term preservation. This often results in the fact that the digitized files do not meet the requirements for web copies. For example, for preservation goals we aim to make an image of the complete content of a carrier. For a web copy it is not acceptable to watch three minutes of color bars before the desired content begins. Therefore after digitization we had to edit many files for use online.

A CMS was used to manage all data and metadata for the website. The majority of the data was generated automatically by importing metadata and links from our existing catalog. Some metadata had to be altered manually.

3.   The bumps in the road

As it happens in all good movies, this project did not come off without a hitch.

3.1. Evaluation

Many carriers just had poor descriptions about their content or, even worse, were labeled incorrectly. In many cases this made it necessary to listen to, or watch, the recordings in advance in order to make an evaluation possible. Multiple formats require multiple machines. But at the AM the broadcast machines are attached in a rack inside the technical room and were needed for digitization. Long research in the technical room would disturb the digitization process.

The thought of winding all the original material forward and backward just for evaluation did not seem to be promising. Therefore, in many cases, it was quicker first to digitize and then evaluate the content.

Thus the expected amount of recordings that would be digitized increased immensely.

3.2. Copyright

The procedure for clarifying copyright took too much time. First we had to detect potentially sensitive rights issues by using the metadata for each recording. Then we attempted to track down the right holders and contact them. Digitizing revealed additional rights holders who were not mentioned in the metadata. Because of the limited time frame, digitizing had to be started before all rights were clarified. Otherwise the digitizing would not be finished in time. Some of the digitized recordings could not be published online in the end because of a negative response to the inquiries or just because the rights holders were undetectable. In order to reach our predefined goal, we had to digitize many more recordings than we originally expected. One positive side-effect of all the numerous letters we sent to clarify rights was that many people who had never heard of the AM suddenly developed a personal interest in the archive.

3.3. Video digitizing system

At first we intended to buy a SAMMAsolo. This system is able to digitize video and to generate a MXF container with the JPEG2000 codec (lossless compression) using a specially manufactured video card to do the processing. Tests however showed that it was quite difficult to use those archive files outside the SAMMA system. Seen from the perspective of an archive this would lead to an unacceptable dependency. So we searched for alternative formats/codecs that would satisfy the criteria for long-term preservation. We discovered the capabilities of ffv1, an open-source codec, but the SAMMA system could not work with this codec. We did not find any other system that could substitute for the SAMMAsolo system. So we invested the money that we were going to use to purchase a SAMMAsolo and we invested it in developing a complete video digitization system called DVA-Profession (www.dva-profession.mediathek.at).

All in all, OaW became more complex than we originally expected.

4.   The results

Figure 1. The homepage of OaW (www.oesterreich-am-wort.at).

From the OaW homepage, when you start a search, for example "Weihnachtsfilm," you get a list of possible matches.
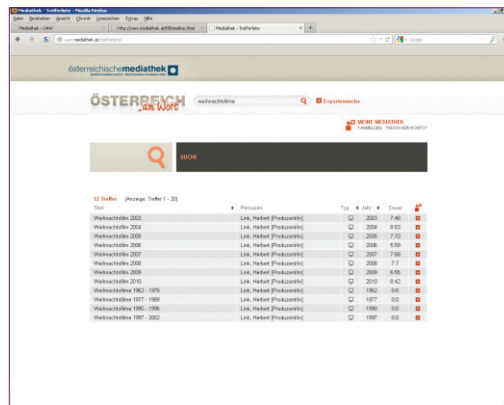

Figure 2. Results from a search in the OaW.

They all belong to a collection with videos made by the filmmaker Herbert Link. He recorded every Christmas evening with his family since 1962.

The short movies vary from super 8 to several consumer video formats and show the family members growing older year after year. Link continues to make these recordings.

This is just one example from about 100 collections, which have been made accessible through OaW. Choosing "Weihnachten 1962" we navigate to the page of Link's recording of Christmas Eve, 1962.
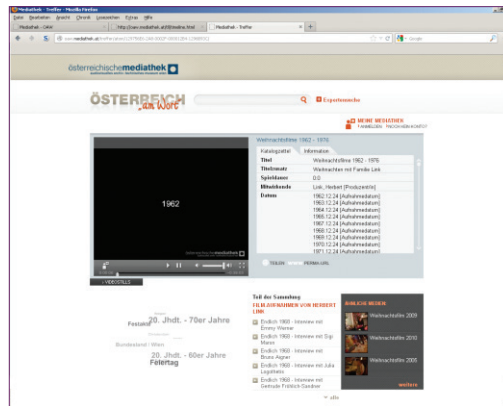
Figure 3. List of Link's recordings from Christmas Eve, 1962.

On the left, in Figure 3, you can see where the player is integrated and beside it is a frame showing the descriptive metadata imported from our catalog. Beneath you see some common features that give suggestions for other relevant items. On the left is a cloud generated from keywords that are defined in the catalog; on the right, some suggestions of similar holdings also generated from defined fields in the metadata.

During the digitization workflow, DVA-Profession creates thumbnails in a frequency of one for every 1,500 frames, which is the equivalent of one thumbnail per minute for PAL video. In this way it is possible to survey the length and the content of the recording. These thumbnails are used as markers in order to navigate through the web file, which in our case is a flash file.
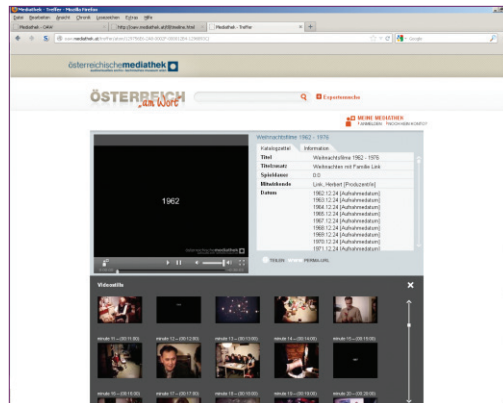

Figure 4. A view of thumbnails in OaW used to help users navigate through a given video.

For instance, when you click on thumbnail number 11, the curser jumps to the eleventh minute of the video file and starts to play (see Figure 4).

The URL is managed by a server that re-addresses any change to the link address. This is necessary to ensure the validity of the link address and the continuous access to the recording.

Clicking on the "Information" tab, which can be found in the gray box to the right of the video stream, shows extra information and descriptions that have been collected manually in addition to the descriptive metadata.

Users can register in order to create individualized playlists. These lists allow users to store, comment, and combine selected recordings in order to find them later or to share them with others. Users can define markers in order to segment the desired sources.

Instead of using the search engine to find a recording, you can navigate through lists that are generated by the timestamps defined in the metadata of the recordings.
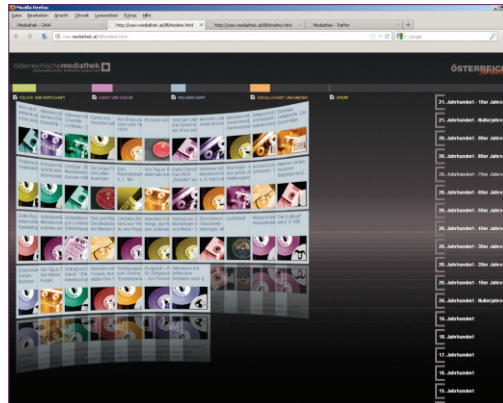

Figure 5. Chronological browsing in OaW.

In Figure 5, clicking on one of the time frames on the right side of the screen expands the chosen period to more detailed time frames and so on. This method of filtering a recording is more intuitive and demands less knowledge about the content.
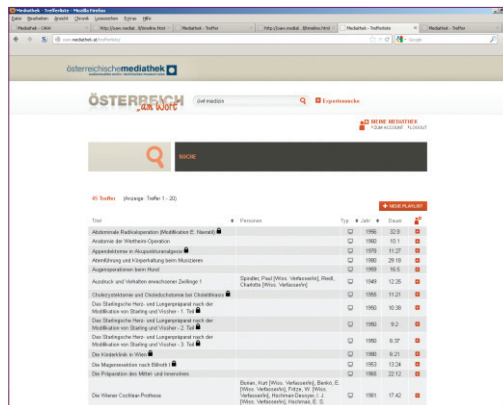

Figure 6. Handling access control for sensitive recordings in OaW.

In some match lists there are icons of locked recordings (see Figure 6). These are indications of sensitive recordings that should be published only to a dedicated audience. In such cases, the steps for how to acquire permission are included. After sending a request you get informed via e-mail if and when you can access the recording for a defined period of time.

5. Summary

OaW, which had the objective of increasing the continuous access to audiovisual source recordings, grew much larger than originally expected. Instead of using an existing video digitization system we developed a completely new and alternative system and published it under a Free Software License (GPLv3 - GNU General Public License). By taking over responsibility for the signal chain through the usage of a self developed system we learned a great deal about video signals, about converting video, and about the strange behavior of some delicate video recordings.

Workflow and time scheduling forced us to digitize in advance before we had clarified all copyrights. In the end we digitized over 2,000 videos and approximately 22,000 audio recordings. Some copyright clearances were acquired after the deadline. Now we are able to give online access to more than 7,400 audiovisual resources. Combined with the 6,000 recordings that had been published in earlier web exhibitions, through OaW, we made almost 14,000 audiovisual recordings accessible on our website. Meanwhile, we just finished another project called "Journale." This increased the number of accessible recordings on the AM website to more than 20,000.