

## Digital preservation and access systems, exemplified by the Österreichische Mediathek

Rainer HUBERT

De primero quisiera decirles cuante me hubiera gustado hablar en espanol, pero no puedo. My espanol es demasiado pobre y ademas me falta el vocabolario professional especial. Por esto dejanme hablar en ingles. –

My paper has the original title “Sistemas the almacenamiento digital para archivos audiovisuales y multimedial” As I am speaking in English I have to translate it and I use the opportunity to adapt it a bit. – I would put it thus:

“Digital preservation and access systems, exemplified by the system of the Österreichische Mediathek”. This is a clarification necessary because I am not able to speak about digital systems in general. That would exceed my expertise considerably.

So let me humbly show you what we have done in the field of digitisation. Doing this I will of course also deal with some general questions.

So, first of all I would like to stress, that the way to digitise and to devise a digital system will differ considerably if we are speaking about a small or a big AV archive. It will also make a big difference, if we aim at a radio archive or an research archive.

Therefore I should first tell you some basics about my institution. – national archive for sound and video

- division of: *Technisches Museum Wien*,  
“scientific institution constituted by public law“
- Holdings: 1 million recordings
- Staff: approx. 25 persons
- Digital system: since 1999, 2002 operational with all basic modules

When we opened this new system it was like the opening of a new building, a new wing of the Mediathek – a virtual building of course.

This virtual building has two facades, meaning the two surfaces by which access to the public is given.

So I am speaking of the access aspect first. –

The first access point is the Voxbox .- the local access. –

Voxboxes are just special multimedia PC's on which our catalogue database – “Dabis” – is running. With the help of this catalogue our users may look for our entire holdings or they can look especially for the already digitised sound material. We have not digitised very much so far, only approx. 3000 recordings. I hope this will augment considerably in the near future.

But now let's have a look at it!

What we see is our catalogue; an adapted libraries database which has been changed graphically a bit.

We can choose now if we want to search in our whole collection, or in the 3000 already digitised sounds. We will do the last. I have to choose an example carefully, because I didn't bring with me all 3000 mp3-files.

We see here a quite normal catalogue entry. The interesting part of the entry is the small icon and text here: mp3 file.

When I push this button a special player – the “Wave-Wizard” – opens the file showing us the mp3-file in its waveform. We can scale the dimension of the waveform and just begin to play wherever we want. If we want to playback the recording a bit more systematic we use a kind of content list: We call it the marker list. As you can see we have divided the whole file by time markers. With their help we can go directly to specific spots of the recording.

In our example the file consists of a compilation of historical sound recordings from Austria, recorded before the first world war.

## EXAMPLES

What I am showing you is a default marker list produced by our cataloguers. Users cannot change this. But they can produce their own marker lists, store them and use them again on their next visit.

This player was devised for us by the company Noa which also came up with the digitisation station we have.

These Voxboxes are local, that is they stand at our premises and at that of the technical Museum of Vienna. Here, as I have already said, the digitised recordings can be retrieved in full length.

This however is not the case with the access by internet. Here only short clips are accessible.

I have to stress again that by internet not the whole files, but only short clips can be used – and not 3000, but only 250 at the moment – and in a much lower resolution and in a streaming format. Main reason for this is the very complicated copyright situation, which is nearly prohibitive. It took one of my staff members nearly a year to clear the rights for these clips! In future we will concentrate on older, that is free material. –

I think it very strange how difficult it is to put cultural lore correctly into the internet. I do not know if this is as difficult in other countries, but I think basically yes.

Now to the preservation aspect of our system. –

I have shown you the two facades of the new building. But what is behind the facades?

### **THE DIGITAL ARCHIVE, THE MASS STORAGE SYSTEM**

When we started in 1999 we had the somewhat naive idea, that we should find one partner company which would arrange everything for us: all the modules of the system and then they would be responsible for its working.

We soon saw that we could not afford such a comprehensive system – and then it would not really meet our necessities. So we decided to build up the system module by module ourselves. That is, of course we had external help and had to find some companies which would work with us, but we had the say.

On the whole I think this was for us a wise decision, because one consequence was, that we had to build up a local expertise in the Mediathek itself. This was very time consuming, especially as we are a small institution, it is very important. I don't think, that to outsource everything is a realistic option: we ourselves must know what we want and we must be able to express ourselves and we must be able to control the work of our partner companies.

Now to the modules and to the very important question of automatic processes and the workflow.

The digital workstations. – We started very pragmatic with a normal PC and a conversion software called samplitude. At the beginning there was no algorithmic analysis of the signal while digitising. Our attitude was to start a process – optimising it when already running; that is there was planning of course, but we didn't plan every detail. We rather were learning by doing. On the whole I think, this pragmatic approach was fruitful.

Meanwhile we have another digitisation station and we do apply an analytic software.

On one station we digitise material in bad condition, not properly cut recordings – that is material which needs special treatment. The second station is NOA workstation. On his – with the help of a job database – three parallel processes can be organised. This is ideal for uniform material. The data of the algorithmic analysis are the basis of the second step of the digitisation: the produced file has to be controlled. I don't think we can do without a kind of quality control.

I need not stress that we do not change or subjectively improve the signal while digitising.

We digitise normally with the resolution of 96 kHz and 24 bit – with this following the technical recommendations of IASA you all know. These wav-files are our new electronic originals.

### **MASS STORAGE SYSTEM**

We have a band roboter by the company Grau.

For the storage we use AIT2 data cassettes; one can store 50 GB;

The Grau band roboter can store 17 TB and organises the files as if it were a hard disc. Therefore its name: Infinistore Virtual Disc.

The main thing about mass storage is security of course – the security for a very long time. We have found that this is a feature other users – that is non-archival users – of mass storage systems do not think that important.

First security network:

All AIT2-cassettes are mirrored, that is we have an identical second set of data cassettes stored on other premises.

Second security network:

For a real long term preservation it is necessary to be able to migrate the media files on to a new carrier and/or in a new format. There are two options to achieve this.

One is the normal way just to estimate when a migration might be wise or necessary. The other way would be to check the files as well as the carrier regularly measuring the error rates and to migrate only when we know by this measuring process that the time has come. WE aimed at this "preventive tape check", but unfortunately this doesn't work. That is, we can do the checking and measuring and we do get figures. But Sony, the company producing AIT2-cassettes, is not able so far to interpret this. So we have to follow the first option. I told you this in length because it stresses my opinion that you have to check and double check everything companies and experts are telling you. A system planned carefully on paper and a system running are two different things.

Third security network:

The analogue originals, which will be here hopefully for some years to come.

Now the catalogue data base.

We have chosen a libraries database by the Viennese company Dabis, because it applies given cataloguing standards and rules. It seemed very important for us to be compatible with the biggest pool of consistent and well-structured electronic data. So we use the middle European RAK-rules and the MAB2 format, which is a parallel development to MARC and I think bridges between them do exist.

We had to adapt the database considerably for our multimedia needs. The catalogue is connected with the mass storage system by hyperlinks.

In order that so complicated a system really works we need a realistic workflow and automatic procedures facilitating our work.

The whole thing starts with the cataloguing.

When the normal catalogue entry is made or an old entry has been checked the cataloguer starts a first automatic process. Additional entries for the future wav- and mp3 file are made:

Then automatically hyperlinks are written in these entries. They target the mass storage system. But as the corresponding wav-file and mp3-file do not exist the hyperlinks lead into nowhere at the moment.

Third automatic step: a text file is produced automatically and put into a folder called "In-Box".

This text file contains a set of directives, a set of orders to be executed with the wav-file, which will produced soon.

As we normally digitise in the same units as the analogue holdings, the signature of the analogue original becomes the file name of the text file (and of the wav-file).

Now the digitisers start. I should add that the cooperation between cataloguers and digitisers is facilitated by a job database organising the workflow in detail. But even more important than that is that they speak with one another, that they work together very closely.

I wont go into the details of digitising. I will do this in the workshop in the afternoon.

The digitisers produce a wav-file, giving it a signature-based file name. Then – after a quality check – this wav-file is put into the same folder.

In his folder we now have two files with the same name – a text-file and a wav-file. When this is the case, that is when these two files come together in this folder, another automatism starts, an automatism called “Wave-Butler”

This “Wave-Butler” was devised for us by the NOA-company. Wavebutler now executes the set of orders written down in the text file:

- Wavebutler normalises the wav-file
- It produces a mp3-file
- It takes basic information from the catalogue and puts it into the header of the audio file, f. e. title proper, signature
- It writes data of the digitising process into the catalogue data base
- It does several other things I cannot mention
- And it finally puts wav- and mp3-files in the mass storage system, controls if the hyperlinks in the catalogue do function now and then makes an entry in the catalogue, that all is running now.

From now on the files can be played back from the catalogue.

For us the work does not end here however, the cataloguers have to write marker lists and to choose clips for the internet.

A system like the described system of the Mediathek needs continual maintenance and development. It is a bit like a living system – it has to change in order to stay functional. Its modules will change as well as the workflow and the automatisms; data and metadata will have to be migrated once.