Historical and current perspectives of digital preservation of sound and audiovisual collections

ALBRECHT HAEFNER
International Association of Sound and Audiovisual Archives (IASA), Alemania

The idea of sustainability has developed in the 21st century as a model for political, economical and ecological procedures. There is the rather sophistic definition of the UN from 1987: "Humanity has the ability to make development sustainable - to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs." (United Nations, 1987)

In Germany we have coined the catchword "Enkeltauglichkeit" – suitability for our grandchildren, because we have to leave to our children and grandchildren an ecologically, socially and economically intact system. Sustainability is a principle, saying that you must not consume more than you can renew or regenerate; it is an action principle for the careful use of resources. Quite clear, sustainability is limited not only to environmental and climatic protection, it is essentially more. From that point of view it goes actually without saying that digital audiovisual archives are absolutely sustainable institutions.

The subject of my paper is "Historical and current perspectives of digital preservation". These are two chapters bound together: the current is the result of the historical.

Therefore, I'll first report about the story of my digital history and then, more or less as an outcome, give a brief survey of the current perspective of digital preservation.

However, prior to that, allow me to start with the fundamentals of our archival business: : Why archiving? Why do we archive? What is an archives good for? What is the "raison d'être" of archives, what the archivist's responsibility?

To collect knowledge, to preserve knowledge in order to remember it, and to disseminate knowledge is the basis of every culture, the fundament of every civilization. Structures of remembering, i.e. of memories, have always been essential traits of all civilizations. In this context, we can understand the establishment of libraries and archives and of collecting strategies.

I believe there is a global double "raison d'être" for every archives: Archives are the memory of the society on its way from yesterday to tomorrow, from the past to the future.

In particular, *documentary* heritage reflects the diversity of languages, peoples and cultures. It is the mirror of the world and its memory.

And And together with museums and libraries, archives hold and preserve the cultural heritage of the society and, thus, represent an essential part of its national identity.

Therefore, it is the chief duty and the most important task of the archives to avoid any lapses or losses of that memory and, further, to support all means which help to preserve our cultural heritage.

Here are the four basic tasks of every archive. Well, this is a kind of definition set up by myself but widely accepted.

- Acquisition.
- Documentation & Cataloguing.
- Access.

- The purpose of archive holdings is to re-use them - therefore, only an accessible archive is a good archive!
- Preservation.
- Preservation is the most important task of the archivist because preservation is the prerequisite for access!

Let me underline that all four basic areas have to be covered if an archive wants to be recognized as an archive. If access and/or preservation are missing then the rest is not justified to be called an archive – it is a collection only. On the other hand: we should not forget that every archive is based upon a collection and starts with a collection.

Finally, what is the AV archivist's responsibility? Basically, he has to deal with two categories: carriers and content. And to look in two directions: backward as well as forward:

- *Backward* to the time when the first audiovisual carriers came into being: Edison cylinders, Berliner disks, shellacs, acetate discs, steel wires, steel tapes, films, photographs, etc. He should, of course, look not only to the technical format of the carriers but also to the content and its historical and cultural significance, relevance and value.
- *Forward* to keep pace with the challenges of the technical development: New storage media, new formats, new technologies. We will see that the most influential development has been digital processing and all its attendant circumstances.
- *Forward* to watch the challenges of media-political development: New content, new services. This could mean to see the chances for new archive business, e.g. content or asset management, or multimedia.

Now my pioneering work, my findings. About 1990 the sound engineers in my company began using digital equipment in their studios for pre-production. Increasingly every time when they had to mix any analogue material from the sound archive, they had to digitize it first. This was the moment when I realized the confrontation with the outset of something new: the digital era. Consequently looking into the future, the idea of a digital sound archive took shape. 1992 we started with a task force on considerations about "safeguarding the radio sound archives", first of all dealing with all the everyday problems we were faced with.

Of the numerous obstacles and difficulties restricting the availability of and access to archival material in the analogue domain let me focus on the two most significant and essential ones: obsolescence and condition monitoring.

First obsolescence. What is obsolescence?

Obsolescent means: in the process of getting obsolete. Obsolete means: old, out of date, out of fashion, gone out, run out. Obsolescence effects threaten archives permanently: Obsolete formats, obsolete carriers and obsolete playback equipment are inherent archival phenomena.

We have to distinguish between:

1) chemical obsolescence of the carriers,
2) obsolescence of formats and
3) technical obsolescence of replay equipment.
1) All storage media deteriorate chemically more or less, the speed of deterioration depending upon

- their physical nature and composition (this is why we call that *carrier degradation/chemical obsolescence*)
- the proper storage conditions (temperature and relative humidity, to call but two effects)

- the proper handling

2) Audiovisual carriers, be it mechanical, magnetic or optical, are obsolete as soon as they are no longer manufactured. Hence, the availability of the reproduction technology is limited. We call that *format obsolescence*.
3) A product such as an audiovisual replay equipment is obsolete if it is no longer in service; no spare parts are available for maintenance and/or repair. We call that *technical obsolescence*.

In the long term, obsolescence effects are the archivist's biggest enemies.

Now condition monitoring. Only regular condition monitoring informs you about the deteriorating status of the carriers in a collection whether they are still in a health condition or just endangered, indicating that necessary safeguarding measures have to be taken. This is very likely the most time and cost consuming duty. Why? With analogue carriers in your holdings, condition checks can be performed only manually and in real time. It is the only method to find out in which condition your collection is. If you do not carry out regular condition checks, you are sitting atop a volcano!

By the way: Condition monitoring, even if continuously carried out, will not prevent carriers from a final decay, no matter whether in ten or in hundred years' time. This is of utmost importance: it means that *an eternal carrier does not exist!* As a consequence, we have to change our archival mentality entirely: it is not the carrier but the information embedded in or on it that has to be preserved. This is really an immense paradigm shift!

In the long term, the information embedded in the carrier has first priority. In the first line, it is the information we have to preserve. Storage carriers and formats play a secondary, an ancillary role only.

Now let me explain why digital technology is the only way to overcome all the problems we are faced with in the analogue domain, in particular these obviously insurmountable problems of obsolescence and condition monitoring. Digital means countable. Origin: digitus (latin) is the finger. When the old Romans counted, they said "numerare digitis", count with the fingers.

Contrary to the analogue signal, the digital signal can take on only discrete, defined values at discrete time intervals and, therefore, represents a physical process by an incremental, step-like course. Digitization is the replacement of the continuous course of an analogue signal's value by an equivalent incremental course in order to allow the steps to be allocated numerical (i.e. digital) values. It is the division into constant absolute units, the "principle of counting small steps". That sounds nice, but rather sophisticated. The best example I can quote and which you are very familiar to is currency.

The fundamental further stage is that today a digital signal means a *binary* signal. Binary = twofold, capable of two values, consisting of two units. Some simple examples: "on/off", "yes/no", "black/white", "high/low" etc. Only two different states need to be distinguished one from the other (remember: in the decimal system there are ten!).

This "yes/no" method is the most foolproof principle there is in signal processing!

This is the superior advantage of digital over analogue technology to reduce a signal consisting of infinite many physical values down to only two states. A simpler solution is not possible.

As a consequence of utmost importance, as long as two states can be differentiated, a digital signal is basically reconstructible, even if it has undergone alterations. This is possible only with binary signals, it is totally impossible with analogue signals. Why? Because binary coded signals have always a well known and well defined rectangular shape, and this is why they are reconstructible. This is the *essential and decisive* advantage of digital over analogue signal processing and the most important reason for digitization. In the analogue domain, signal alterations are either not detectable or, if detectable, reconstruction is impossible, difficult or expensive.

Result: Compared to the *analogue signal processing*, the accuracy of which is limited by imprecisions such as reading, measuring, processing, and transmission errors, mechanical wear, ageing, and external effects such as temperature, and the like, digital signal processing is more efficient, more precise and most secure. Most of all, there is the feature of lossless copying, contrary to analogue duplication which is always lossy. However, substantially higher "bandwidth" (i.e. bit rate) and larger storage capacity is needed.

Attention: Let's assume we have just digitized an AV collection. What results have we got so far? Our collection is transferred onto a new storage medium and into a new (digital) format. Has the transfer eliminated obsolescence? NO. The next format obsolescence is sure to come in the foreseeable future (digital formats get obsolete, too), forcing us to reduplicate the collection once more onto a another technical platform and/or another storage medium and/or another format, causing again efforts in terms of manpower, material and equipment. The one and only progress we have achieved: no more copy losses thanks to digitization. So, we have to realize that digitization alone does not really solve the problem of obsolescence, it simply postpones it only. Would that mean a steady transfer again and again and nothing else?

Well, digitization is done by means of a computer. Idea: the costly manpower, that is the manual work, has to be replaced by automated computer-controlled systems.

What is the manual work? Analogue audiovisual content threatened by obsolescence effects has to be transferred onto a new storage medium, a process that is always lossy and can be performed only manually in real time. The same is true for condition monitoring of analogue AV content which is possible only manually and in real time, also a process that is always lossy. However, as soon as the content is available as digital signals, as files, then they can easily be automated by means of a computer-controlled storage system – that is impossible with analogue signals. Digital AV signals are, of course, also threatened by obsolescence, but, unlike analogue signals, they can be duplicated, transferred and stored by means of a controlling computer lossless and automatically. Result:

Away from manual transfer towards automated migration

What features are requested for such an automated migration?

- It must be computer-controlled.
- Data must be stored digitally.
- Data must be kept redundant.

With these features, an automated system will work

- auto-controlling, i.e. the system itself checks the integrity of its data;
- auto-regenerating, i.e. the system itself "repairs" corrupt data (refreshing);
- auto-transferring, i.e. the system itself transfers data (migration).

These features mean nothing else than to overcome obsolescence effects and offer permanent condition monitoring! As soon as you have such an automated system in use then you may finally forget the question: "What is the life expectancy of this or that format?

Those were the results of my reflections, to have such a system was my dream. At that point a pure but lucky chance put me on the right track. At an exhibition dealing with "Satellite-based communication" in 1993 SONY presented their Digital Instrumentation Recorder DIR 1000, a digital mass store system with a storage capacity within the TeraByte range. Just that I was looking for!

Well, those systems have been used since the early 1980s for data storage mainly by EDP departments (scientific as well as military applications), and a variety of such devices has been offered on the market. There have been installations in branches which have to store vast quantities of data such as assurance companies and bank institutes. Also, other services which create a great many data within a very short time period prior to final data processing such as interferometry used in cosmic radio research or earth observation by satellite-based remote imaging have made use of digital mass storage systems.

Hence, the technology was already available, it was only the application in a sound archive which was new! All we had to do was to see how such a Digital Mass Store could be applied to our requirements.

In 1996 we started a pilot project called DMAS (Digital Media Archives Solution) for two years. Outcome: A digital mass store is the only successful way to overcome the problems in front of us. And in 2000 the call for tenders, evaluation, recommendation and award followed.

So far the history of my story. Now the current and future perspectives.

Having a computer-controlled AV storage, obsolescence effects and condition monitoring are no longer a nightmare for us. Confronting conventional analogue with digital technology: there are many more advantages of digitization:

- *Permanent availability and accessibility*: all files are at any time available, no physical carriers such as tapes and discs in use. Benefit: as soon as the entire or main parts of the holdings have been transferred into the DMS, no staff is needed anymore; reasonable saving of staff resources. Released staff can be used for digitization of analogue material.
- *Multiple access, parallel use*: parallel/multiple access possible at any time. Users have at any time access to all parts of the holdings, no staff needed. Benefit: As no more copying is needed, saving of staff resources, saving of money for new carriers, saving of machinery resources.
- *No longer shelves and racks*: no special storage facilities needed, as hard disc arrays are part of the DMS. Benefit: Saving of investment and operation costs.
- *Access from any point of the world*: the internet allows access even from or to overseas.

- *No longer wear&tear, damaged or lost carriers*: a DMS is a closed system, no external manipulations by users possible, e.g. improper treatment of hard discs or cartridges. Benefit: No additional monitoring needed: saving money for staff and machinery.
- *Copying faster than real time*: high data rates reduce essentially the time for copying.
- *No longer labeling and housing*: Physical carriers need special protective containers (cardboard boxes, cassettes, cartridges, covers, sleeves etc.), containers need labeling and marking that is no longer necessary. Benefit: Saving of material costs and personal resources.
- *Considerably less storage space*: As a result of the miniaturization of all system parts, essentially less storage space required.
- *Nearly no air condition, no efforts against dusty air and electro-magnetic stray fields*: special storage conditions only in a much lesser extent needed. Benefit: Substantially reduced expenditures for air condition, essential savings of investment and operation costs.
- *Operation around the clock*: Holdings are available and accessible 24 hours around the clock without staff. Despite the continuous availability/accessibility no additional staff needed. Benefit: saving of funds.
- *Fast transfer of files*: high data rates reduce essentially the time for file transfers.
- *Extremely large storage capacity*: today, storage capacities of PetaBytes possible.
- *Low error rate, error correction*: create redundancy by adding additional bits, e.g. parity bits.

In other words: Away from the ephemeral carrier towards the eternal file.

Here is my short Conclusion:

Let's remember the meaning of sustainability: We have to leave to our children and grandchildren an ecologically, socially and economically intact system. This is, of course, true for leaving our audiovisual heritage to the future generations. Digitization makes it possible – and this is why digital AV archives are sustainable.

References

United Nations (1987). Report of the World Commission on Environment and Development: Our Common Future: http://www.un-documents.net/wced-ocf.htm