YLE digital sound archive

P. Gronow and M. Petäjä
Yleisradio (Finnish Radio)

Like most other broadcasters, Finnish Radio is addressing the problem of what to do with its mounting volume of sound archive material, much of it still stored on quarter-inch analogue tape. A digital archive system is planned for the year 2000 and this article looks at the thinking which has gone into the planning of this new system.

Introduction

For several decades, the tape recorder was the standard tool of every producer of radio programmes. Most programmes were recorded on tape before transmission and, if for any reason it was decided to preserve a programme permanently, one simply took the reel of tape and put it on a shelf in the archives, and it was subsequently catalogued. Quarter-inch professional-quality tape was an excellent archival medium; it had a life expectancy of half a century. Over the years, Finnish Radio – *Yleisradio* – has acquired a collection of about 250 000 archive tapes, with a total playing time of over 100 000 hours.

Now the days of tape are numbered. *Yleisradio* is no longer buying any new analogue tape recorders, because for several years new programmes have been recorded on digital audio tape (DAT) machines. Soon we will be working in a completely computerized environment where all the elements of the system – from newsgathering to transmission – use digital technology, as illustrated in *Fig. 1*. Our newsroom has acquired a new production system, where journalist use their workstations to write news reports, edit interviews and record their own commentary. This takes place on a local network, enabling all producers to share the same information. The system will also record all newscasts as they are aired. No tape is required to preserve the text and audio parts of the newscasts: they are stored on the hard disk of the computer.

This is all very wonderful but, from the viewpoint of the archivist, the system has one critical shortcoming – it only has the capacity to store our newscasts for about one year. After that, we'll have to move this archive material elsewhere, in order to make room for newly-archived production material.

Of course there are several ways to solve this problem. We could copy the newscasts onto DAT machines, recordable CDs (CD-R) or even analogue tape, but in doing so, we would lose the connection between the written text and the sound. We could expand the storage capacity of the newsroom system and buy additional hard disks, but this would not help our other production departments, which also need new means of archiving their programmes – as their tape recorders are also being replaced by workstations. After all, *Yleisradio* has five national

Original language: English
Manuscript received:: 23/4/99.

and over thirty regional radio channels, and all of them produce material – in varying quantities – that needs to be archived.

We have calculated that the annual growth of our audio archives will be between five and ten thousand hours of sound. Ten thousand hours of linear audio (stereo) equals about 7 terabytes (10^{12}) of storage capacity. Such storage capacity is no longer an unattainable goal: in terms of cost per hour, it is actually far less expensive than the cost of analogue tape. We have found that there are at least half a dozen companies who can deliver this storage capacity, off the shelf.

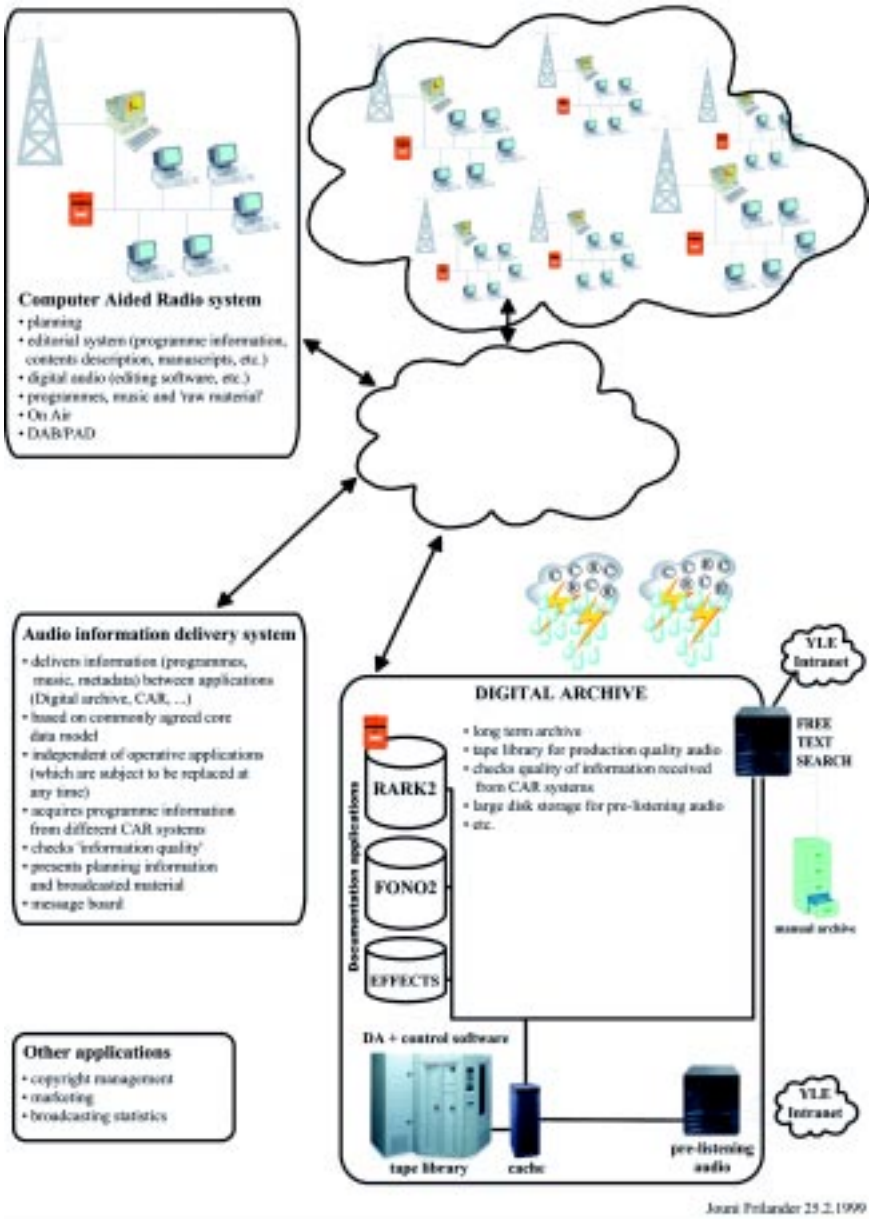

Figure 1
YLE digital sound archive environment.

However, storage capacity is not enough. What we need is an archive system which has the following capabilities:

- **Best possible sound quality**. Ideally, this means linear audio (48 kHz stereo). In practice, most of our current production systems use some type of audio compression (bit-rate reduction). We have not found any reason to convert compressed sound to a linear format for archival purposes so, in practice, this will considerably reduce the amount of storage capacity needed.

- **Compatibility with existing catalogue databases**. We already have extensive catalogue databases covering our analogue holdings. Users of the archive system must be able to search the catalogue simultaneously for analogue and digital materials, if they choose so. At present this would seem to indicate that it is advisable to keep the catalogue database (metadata) separate from the sound files; in the audio world, we do not see any practical benefits from a database structure that combines metadata with content.

- **Compatibility with existing and future production systems**. The various radio channels of YLE have different profiles, and a production system which is suitable for a popular music and current events channel will not be ideal for the arts channel. The archives must be able to serve both.

⇨ **A two-tiered system which combines browsing-quality sound for quick reference, with production-quality sound.** As new items are added to the sound archive, the system will automatically produce a "browsing" copy.

The architecture of the planned *Yleisradio* archive system is illustrated in *Figs. 2 - 5*.

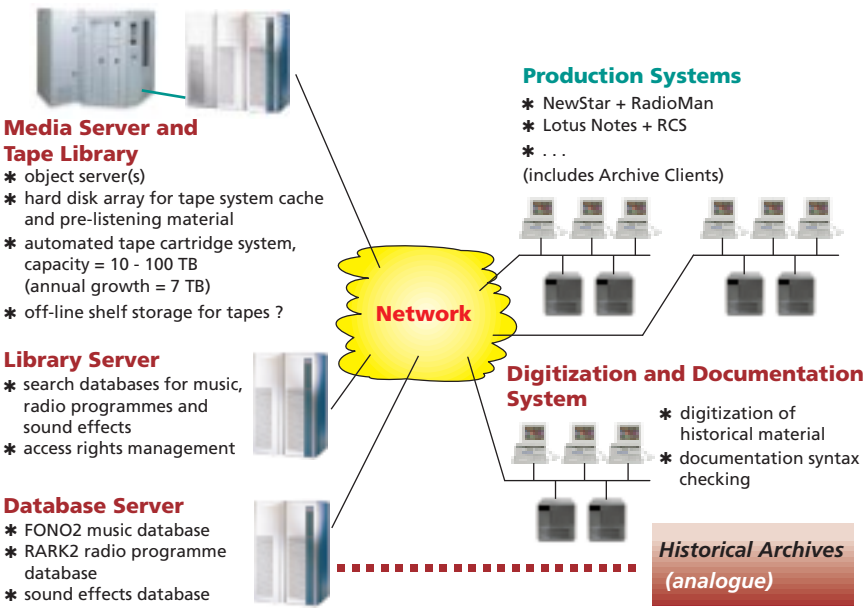

Figure 2
System architecture.

Working with the digital archive

We hope to have the new digital archive system installed in the year 2000. Integration with the various production systems will require additional time but, after a few years, we expect to have the system fully working, with enough archive material to make it useful for producers.

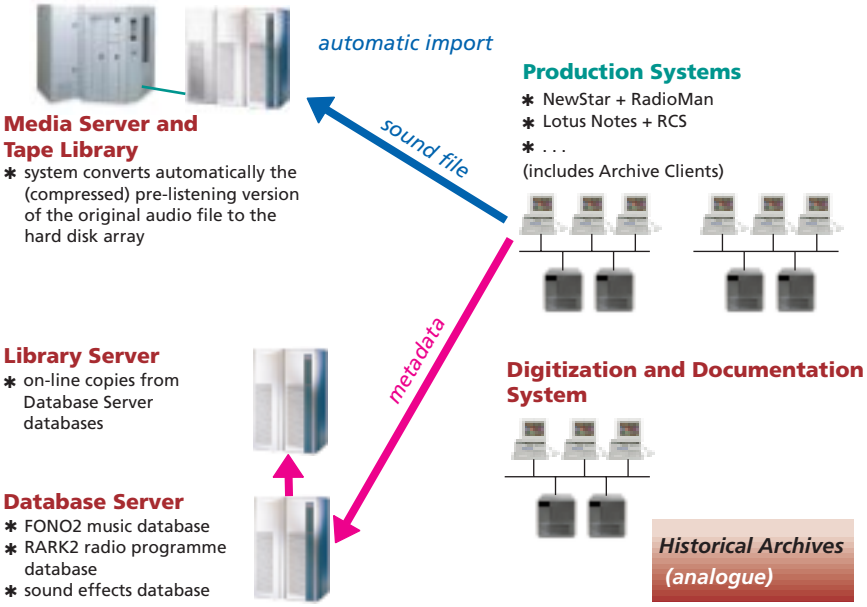

Figure 3
Archiving from production systems.

In the future, a journalist looking for archive material will simply open the workstation in his / her room and type in the commands for the required material. Should he / she be searching for the Finnish president's statements on the Kosovo crisis, he will probably type AHTISAARI and KOSOVO. The journalist can limit the search to a certain time frame (for instance, February 1999), or add other qualifiers. The database will respond with a list of "hits", indicating the dates and titles of the programmes where the requested combination exists. The database will also indicate whether the items are archived in digital sound, or are perhaps older items not yet digitized.

If the journalist finds items which look interesting, with the click of the mouse, he or she can open an audio player with the normal functions (play / stop / quick play / forward, etc.), and begin listening. The response time should be a matter of seconds. If the item is suitable, another click will send a request to the system for a production-quality copy, which should arrive at the journalist's workstation in a reasonable time. If only part of a longer programme

is required, the journalist can limit his/her request to only that section of the sound file, to avoid unnecessary loading of the system.

The digital archive will be open 24 hours a day. It addition to spoken word, it will also contain selected items from the record library, ranging from current "hits" to the "evergreens" always played on request programmes. There will even be a selection of music for the events feared by all radio producers – the serious crises which require the rehauling of the complete programming schedule at short notice.

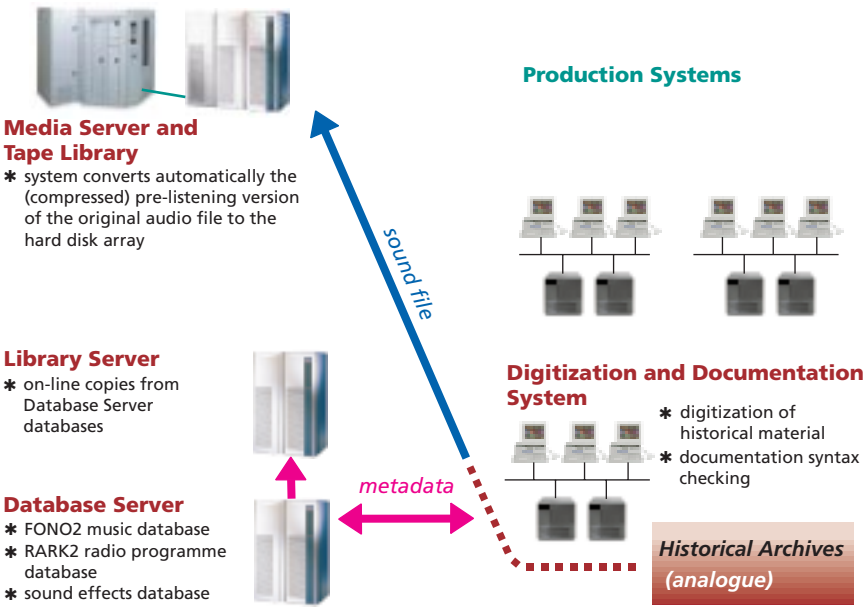

Figure 4
Digitization of historical materials, and metadata checking.

Catalogues and content

Unlike books, tapes do not have title pages. Usually there is just a reference number – typed on both the box and the reel. Up to the 1970s, every reel of tape was documented on a file index card. Additional cards were prepared to facilitate searching by means of cross-references. The cards were typed by trained cataloguers, working from information provided by production sheets, and sometimes from listening to the tape. When catalogues were computerized, the file index cards were replaced by a catalogue database which permitted more effective searches, but the work itself still required specialized cataloguers.

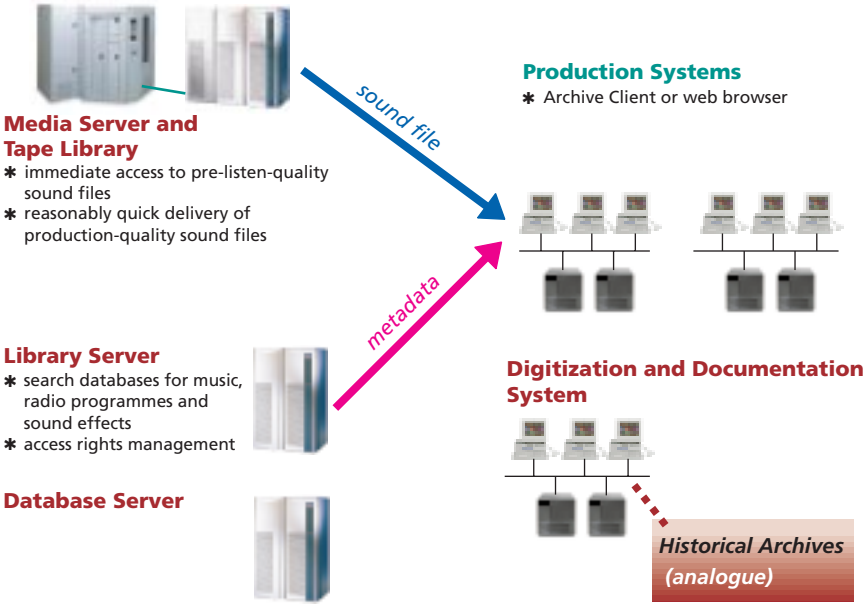

Figure 5
Retrieval.

Today's journalists often type the manuscripts of their programmes on the same workstations used to edit the sounds. Usually they have to fill in a lot of additional data for administrative purposes – the names of persons interviewed, subjects discussed, etc. With a digital archive system, it is possible to retain all this information along with the sound files. But can it be

retained in a format which makes possible the efficient use of this information in a catalogue database?

This will be a key question for future radio archivists. There are good reasons to believe that, with the progress of time, the cost of digital storage capacity will continue to fall. In the future, it will be possible to archive massive amounts of sound permanently. However, such archives will be practically useless, unless we know what they contain and can quickly find the items we require. And the cost of human labour is not likely to go down.

What we hope to achieve in the future is the "semi-automatic" cataloguing of all new programmes deposited in the digital archive. We will encourage journalists to create their manuscripts and all additional data in a format which will enable us to use it as the basis of the entries in our catalogue database. When a producer decides to send a recently-aired programme to the archive for permanent storage, the system will automatically check that certain key facts are also supplied and transferred to the catalogue database. Some supervision from professional cataloguers will always be needed, but such a system should enable us to increase the number of programmes permanently archived, without the need to increase the workforce required to catalogue it.

Preserving the historical archives

The digital archive is also a means of preserving historical sound recordings. Although the storage devices used in the archive, and perhaps the entire system, may need to be replaced within the foreseeable future, digital data as such can be – in fact, has to be – preserved permanently. In this respect, the sound archivist is not alone. Banks, insurance companies, government agencies and other organizations which have the need to preserve digital data permanently have to plan for the future migration of their archives into new systems, as technology advances.

Analogue tape has a maximum life-expectancy of fifty years (although certain brands have proved, in practice, to be far more vulnerable to the ravages of time). Our oldest tapes are from the 1940s, and they have long since been copied. We have for several years been copying tapes from the 1950s onto new carriers – recently DAT or CD-R. The extent of our archiving in the 1960s was still relatively limited but, in the 1970s, our company finally started to archive radio programmes on a large scale. Future archivists will be really busy after 2020, when at least ten thousand tapes a year will have to be transferred to a new format, if they are to be preserved.

All currently-available media for preserving sound have a limited lifetime. Somewhat paradoxically, the shellac 78 rpm disc seems to be one of the most durable, with a proven lifetime of a century – provided that it is protected from shocks and wear. In contrast, the life-expectancy of DAT is uncertain and, even if the tapes last, there is no certainty that working DAT recorders will be available in 10 - 20 years time. Will we ever be able to preserve all the sounds currently in our archives? Probably not, but we are sure that we can preserve a considerable part and, hopefully, the most important part. To achieve this, we need a selection policy.

We have about half a million commercially-issued records in our library. Although they are invaluable for our programmes, many of them are quite common items, and will probably be available commercially even in the future. It is not likely that the permanent preservation of the recorded heritage of Caruso or the Beatles will depend on us. On the other hand, we do

have many rare, perhaps unique, Finnish records which are not known to exist in any other collection, and we have a responsibility to see that they are preserved.

The quarter-million radio programmes in our archives are, for the most part, unique in the sense that no other institution in the world has copies of them. We would like to ensure that they are preserved, but it would take a long time to digitize all of them.

From the viewpoint of a broadcasting archive, first priority must be given to frequently-used items. We know, for instance, that old news broadcasts are frequently used in the production of new programmes. At the moment, we are planning a series of historical programmes which will document the Finnish Winter War, sixty years ago, on a day-by-day basis. Any existing recordings from the war years will certainly be used over and over in the future. In the same way, programmes based on events 50 years, or 25 years, or ten years ago are frequent. To facilitate the reuse of old news material, we are planning to digitize all the newscasts preserved in the radio archives. This represents about five per cent of the volume of the tape collection.

Other frequently-needed items are: the voices of well-known Finnish and international personalities from the worlds of politics, business, sport, culture and entertainment; reports of historical events such as the first moon landing or the fall of the Berlin wall; the Olympic games and other major sports events; popular radio personalities of the past, and so on. In the field of music, we will probably digitize all the records that have been in the Finnish top ten for the past fifty years, as well as other records frequently requested in our phone-in programmes.

On the other hand, digitization will also be a means of saving old recordings which are approaching the end of their lifetime. The amount of Finnish radio programmes preserved from the period before 1960 is so limited that it will probably be best to digitize all of them, without any selection. We will also transfer, as soon as possible, any recordings which still are on acetate discs, compact cassettes or other similar, unstable, media. As we approach the 1970s in the archives, stricter selection criteria may well prove necessary, but there is fortunately still time to plan it.

In the cases described above, digitization will take place systematically, at the initiative of the archivist. But there will always be cases where new needs arise in programme production, and we expect that a considerable part of our work in the future will be digitization on demand. For several years now, we have not lent fragile 78 rpm records to programme producers; anybody needing a copy of a recording not available in other formats has had to order a tape copy. In the future, all such copying will go via the digital archive, and once a copy has been made, it will be kept there for the next user. In the future, our producers will probably no longer be able to use vinyl discs or analogue tapes from the archives, as record players and analogue tape recorders are vanishing from the studios; instead we will have to digitize the material on demand.

Finally, the question of rights also has to be taken into consideration. The rights for news programmes and other similar items produced by regular staff members are owned by the company, and can be freely reused in the production of new radio programmes. But in the case of music, plays, literary programmes and most programmes produced by freelancers, other rights are also involved. The reuse of such material can be expensive, or may demand new negotiations with the rights owners. Should we aim to preserve such recordings, if they are culturally or historically valuable, even if we will probably not be able to reuse them in the foreseeable future?

Pekka Gronow *is the manager of the radio archives at Yleisradio. He has a Ph.D. in ethnomusicology from the University of Tampere. He is also a frequent contributor to journals and reference works on music and recordings, and his work has appeared in Ethnomusicology, Popular Music, IASA Journal, Grove's Dictionary of Jazz etc. His latest publication is "An International History of the Recording Industry" (Cassell, London 1999).*

In addition to planning the digitization of Yleisradio's sound archives, he is involved in a project to present the Finnish national discography to internet users all over the world. At the moment, the discography covers all Finnish records issued between 1901 and 1976. It can be accessed at the home page of the archive, http://www.yle.fi/aanilevysto/firs.

Markku Petäjä *is a project engineer at Yleisradio. He has an M.Sc. degree in computer science from the Raahe technological institute. After working in the computer industry, he joined Yleisadio in 1996 and is now responsible for the technical planning of the company's digital sound archives. He has also been involved in the planning of the company's new studios for DAB broadcasts.*

This brings us to the larger question of opening the radio archives to other users. After digitization, it would be technically possible to open the archives to users outside the broadcasting company – to scholars and students, through terminals in libraries and universities, or even to a larger public. At the moment, we do not have the right to release a large part of the material in our archives to outside users. To open our archives without proper authorization would be out of the question. Should we strive to open our archives to a larger public, beyond the one that we reach through the airwaves, or should this problem be left in the hands of the National Library or other similar institutions? With the new possibilities opened by new technology, many European broadcasters will have to consider this question in the future.