

Technical Challenges and Developments in 21st Century Folk Music Archiving
International Symposium, Budapest ZTI
11–12. 06. 2008.

Pál Richter (ZTI, Budapest)

Endless Job — Past, Present and Future in the Folk Music Archives
of the Institute for Musicology of Hungarian Academy of Sciences

The Folk Archives at the Institute for Musicology was officially established on January 1, 1999. Naturally, however, the Folk Music Research Group of the Hungarian Academy of Sciences has had a technical department since the very beginnings (1951) and continued work under the auspices of the Institute for Musicology which was created through the merging of the Bartók Archives and the Folk Music Research Group. The archive possessing sound recordings of folk music and audiovisual recordings of folk dance has been systematically set up and has won outstanding renown in Europe. The sound recordings of the Archive include 18,000 hours of music, 15,000 hours of which are the Archive's own recordings, that is, the source materials can also be found here, and the remaining 3,000 hours are made up of copies of recordings stored in other archives. The latter include the phonograph cylinders of the Museum of Ethnography, which are available both in analogue and digital form in our Archive.

In the first period, from the late 19th century collection of Béla Vikár to the 1950s, the sound recordings were preserved at the Museum of Ethnography (some 4,500 cylinders with about 200-250 hours of music). The records and metal plates of the Pátria series launched in 1936 are also kept in the Museum. From the 1950s the Folk Music Research Group gradually assumed the role of maintaining a central collection. Until the appearance of portable magnetic tape recorders in 1955, the sound recordings were made on a Webster's wire recorder. After collecting, the recorded material was all transferred on single LPs, called the AP disks (=Academic Pyral), that is, sound recordings were carved onto with a plastic surface covered metal disks which can be played on a record player like traditional vinyl disks. Only the copy was available to researchers for transcription, no one, not even the collector had access to the original. From the sixties, folk music recordings stored in other institutions were also brought into the Archive: the total phonograph collection of the Museum of Ethnography was copied onto tape, similarly to the authentic recordings possessed by the Hungarian State Broadcast, recordings by Hungarian folk music researchers abroad, individual collections, etc. The inventory of 1998 registered 6,100 tape reels, 275 Webster wire reels, 120 DAT cassettes and 150 video cassettes, which add up to some 10,000 hours of music preserved in air conditioned storerooms.

In the 1980s an epochal change took place in technology. Traditional LPs were taken out of circulation and began to be replaced by CDs. AP records stopped being made (we have a total of 9,140 APs) and since 1996, the recordings have all been transferred to CDs.

Since the majority of our recordings are magnetic tapes and cassettes. They have to be preserved on other kinds of data carriers. Such recordings had earlier been copied onto AP (Academic Pyral) disks, some 150 hours invaluable material was copied onto metal disks in the 1970s, however, the costs were extremely high. Apart from being susceptible to rough, external mechanical injury, these disks were made for „eternity” and can always be played with a traditional pickup. In the next phase, in the age of digital technology the digitalization of analogue recordings began and has formed the core work at the Archives in the past 10

years. Unfortunately no subsidy is allocated from the budget for this work, therefore, funds for our most important job must be raised from tenders.

The majority of the sound recordings, about 70% (12,000 hours) have already been digitized. During the process a digitized version of the analogue recording is made and copied onto a hard disk which is then copied onto CDs. Since these recordings require substantial storage capacity they have to be erased from the hard disk so new recordings can be copied onto them. Thus, only the CDs are kept as sources. At least two CDs are made from each recording: a so-called “wear out copy” for researchers and an archival copy. Naturally, analogue data carriers are kept in air conditioned storage rooms as long as possible, and has sense. However, the surface protective film added to CDs has also got a limited durability, and we have had both positive and negative experiences with it, depending on the make. All in all, CDs are more durable than we had previously expected, however, within ten years we will have to begin transferring the whole collection onto a different data carrier. Storing digitized copies on CDs is a temporary solution, therefore, we are searching for new technologies to preserve the recordings for posterity. Audio files are saved in WAV format and are digitized at 44,1 kHz and use 16-bit resolution, which was the highest one for a long time.

Great emphasis has been placed on providing the best possible analogue signals during digitalization. Therefore, devices on which old type data carriers can be played have been safeguarded with utmost care, and several of them, including a studio tape recorder, have been transformed to suit our needs.



Pál Sztanó must be mentioned because he had several creative ideas and archiving techniques which helped the preservation of invaluable recordings. Since his death in 1997 the technical staff has been led by István Németh, who had gone to field collections and learnt his trade under the guidance of Sztanó. The idea to invent a non-mechanical contact mechanism for

playing phonograph cylinders was the brainchild of Sztanó and we have been experimenting on developing such a mechanism for several years jointly with the technological research group of the Hungarian Academy of Sciences. Our experimentation has produced some promising results (see Sándor Manno: *Playing Phonograph cylinders with the technology of the XXI-st century. Contactless recordplayer*).

The sound recordings and field transcriptions (I mean scores and lyrics) in the collection amount to about 250,000 pages. Moreover, there is additional documentation of field work, field notebooks, documentation of collections and photographs. The digitalizing of field transcriptions also began. The present state, the size of the collection and rate of digitalization can be seen in the following table.

Original Medium	SUM (items)	SUM (hours) (estimate)	Digitised (items)	Digitised (hours)	Digitised (%)
Magnetic tape reels	6 631	8 600,0	4 000	4 860,7	55%
Webster magnetic wire reels	297	140,0	277	123,0	88%
Minidisc	255	380,0	251	364,1	96%
Dat	289	500,0	280	484,7	97%
Audio cassette	1 200	1 500,0	510	612,0	40%
X-ray plate & Pyral record	400	100,0	250	34,0	34%
Video cassette	162	400,0	0	0,0	0%
Sound carriers from other places	2 800	3 748,5	2 800	3 748,5	100%
Tapes out of inventory	800	1 200,0	0	0,0	0%
SUM (7 578 CDs in two copies)	12 834	16 500,0	8 368	10 227,0	62%
CDs out of inventory					
Edison phonograph cylinders:					
Museum of Ethnography	4 551	370,0	4 300	350,0	94%
Kodály Zoltán (Kodály Archives)	425	35,0	400	33,0	94%
ZTI	370	30,0	300	25,0	83%
Utolsó óra (Last Moment Project, Hungarian Heritage House) (estimate)	1 000	~1500	1 000	~1500	100%
SUM Total (~9 000 CDs in two copies)	19 180	18 435,0	14 368	12 135,0	67%
Photo	70 000		7 000		10%
Record cards (transcriptions)	250 000		25 000		10%

Non-Hungarian Collections	Amount (hours)	Digitized (%)
Neighbouring peoples (Gypsy, Croatian, German, Romanian, Serbian, Slovakian etc.)	~ 1500	66
Relativ peoples (Finno-Ugrian, Turkish)	~ 500	70
Bru tribe (Vietnam)	252	100
Piaroa tribe (Venezuela)	70	0
Indian folk music (northwest)	50	100
Coptic liturgical music	40	100
Jewish liturgical music	70	100
Other smaller collections: Papua (from the beginning of the 20th century), Korea, Pakistan, Syria, Libanon, Armenia, Angola, Ghana, Ethiopia and Ugurian, Arabic, Lapp, Gypsy (from several countries of Europe)	~ 300	10
SUM	~2782	

Naturally, proper storage and retrievability of data are an integral part of archiving. Thus, another important task is to enter the data about field collections and sound recordings into a computerized database. On Ferenc Sebő's initiative László Dobszay launched a project in the 1990 to set up a computerized catalogue of the collections with the aim of making this vast national heritage available to all. A separation of the electronic data-processing of the closed and the so-called historical systems from the rest of the material took place within the framework of this project. From the mid-'90s István Pávai suggested the unification of the earlier data input methods and their conversion to an up-to-date database. These steps were taken soon. Funded by a tender in 1999 we created an interactive and dynamic Internet website-system that provides a database ([Hungarian Folkmusic Archive](#)) in which the main types of Hungarian folksongs, excluding the so-called new-style melodies, are searchable by their musical attributes and acquisitional and geographical details. The capacity of data-processing systems and the possibilities of computerized immediate display of different media have grown remarkably in the last decade. This extensive technical development makes it possible that earlier independent databases of part-repertoires be joined, and data be supplied with the playing of the media attached to them.

The development that began in the first years of the new century based on István Pávai's plans served this purpose. The database software developed especially for folkloric data is now available and makes our work easier. (You will hear about this as well in detail at the last session). Earlier, data had been entered into Excell tables with each subcollection being assigned to a different sheet. However, linking the subcollections and cross-searches were rather clumsy. Owing to the database software all of the Bartók system is searchable on the Internet and includes the history of the system, [which is a short summary of the introduction of the Complete Collection published in book format (Sándor Kovács: *Evolution of the Bartók system* In *Béla Bartók: Hungarian Folk Songs: Complete Collection Volume 1*, ed Sándor Kovács, Ferenc Sebő, Akadémiai Kiadó, Budapest 1991, 13-31)], and includes Bartók's rhythm charts and an advanced search. Record cards (base sheets) and field transcriptions can be looked at, and any available sound recording can be listened to ([Bartók System](#)).

Making the database accessible to the public does not only seem an endless job because of the extensive amount of data. Historically, the most outstanding scholars of Hungarian

ethnomusicology have also been distinguished composers and cultivated musicians (Kodály, Bartók, László Lajtha, Pál Járdányi, etc.), therefore, musical interpretation and musical systematization of the collected material has always been a priority. From a musicological and ethnomusicological aspect this analytical and contrastive processing has led to several new and significant scientific/musicological discoveries. It has helped in puzzling out the relationship between Hungarian and neighbouring peoples' music as well as that of related peoples and in revealing different styles of Hungarian folk music and has attempted at unveiling the origins and dissemination of various styles. The publication of this vast material has always been planned to be based on musical standards. One of the hermeneutic documents is the Bartók system discussed above, and another is a critical edition of Hungarian folk music currently being published based on musical criteria in the set *Corpus Musicae Popularis Hungaricae*.

The central typology (the main corpus of the collection of transcriptions, the so-called Large System) handled by the Folk Music Research Department of our institute, includes variants of 4,600 tune types, excluding the new style tune types. The computer database contains more than 90 000 items, data of transcribed tunes. The researchers of the Institute participate in the work of systematization by musical types, sorting out the variants and creating new types, subtypes when needed. Few years ago revision and rearrangement of the approximately 2,500 types of new-style folksongs (about 60 000 units) have been carried out and completed by János Bereczky in the Folk Music Research Department. ([tune systems](#))

From the point of view of archiving various kinds of musical classifications done at different times was a challenge that seemed almost impossible to meet. Due to a considerable increase in the amount of collected material after the Second World War, Bartók's and Kodály's practice of having every single data copied for each other had to be abandoned. In the 1960s and 70s, when photocopy machines did not yet exist, copies were made in a very limited number, therefore, each classification affected the whole of the material on each page (record card). In the case of sound recordings each set of field recordings stayed as one because of the data carriers, however, transcripts made during field work without sound recordings were scattered across the collection. Moreover, field transcripts were not assigned inventory numbers as they entered the collection but only later, when they were classified. In the beginning the function of inventory numbers was not clear, and items with sound recordings/audio items were also assigned the same type of inventory numbers. The main problem was that items entering the collection with a sound recording and without a sound recording were treated in two different ways: the archive has always kept proper track of sound carriers (cylinders, magnetic tapes and cassettes) in the inventory books, but items without a sound recording were bulk stored waiting to be catalogued in every system. Reliable retrievability and keeping count of items was unachievable without unified inventory books. Thus, we have still not been able to determine the exact number of items in the collection. The situation is aggravated by the fact that apart from musical classifications which affected strophic vocal items primarily, further thematic subclassification was needed both for scientific reasons and due to the nature of the material. These thematic subcollections (customs, children's games, instrumental music, music from neighbouring countries and ethnic minorities and laments) only partly overlap with the central system since they contain material which is neither strophic nor vocal ([chartmap](#)).

Computerized data entry initially bore all the contradictions it had inherited – structurally it did not differ from the built-up of the collection. The first classification numbers of individual subcollections were system numbers based on musical classification, moreover, the dBase program allowed for the entry of a very limited amount of meta data at that time. With the help of the data processing system developed specifically for our use the whole collection has become easy to survey and a record of all the data has been maintained. The technology is

available. Having learnt from the achievements and mistakes of our predecessors, data entry fields have been created (in Excel format) so that they can be applied universally to any of the different subcollections. After checking these [data tables](#) they are imported into the data processing system. Since work on the digitalization of the sound recordings will probably be complete in two years, we will be able to utilize our scanty financial resources (which can at times be substantially increased through tenders) and our human resources on processing all the data in the collection. Similarly to the Bartók System (see above) we aim at making the whole collection available to our researchers on our internal network. The website will only differ from it in that the restricted parts of the collection will not be accessible from the web. We will be able to provide filtered data from the subcollections and present the material in keeping with the different classification systems. The time when our large-scale plans will be fully realized is off in the distant future. We have made considerable progress in entering the data of the audio collection (AP disks, magnetic tapes and their documentation) (150,000 records) through an integral checking of entered data. However, we will have to face the task of entering the data of all the visual materials, which means that a detailed entry of 250,000 additional data needs to be completed in the Hungarian collection alone, and all of the paper-based collection has to be digitized.