

EÖTVÖS LÓRÁND UNIVERSITY BUDAPEST
DOCTORAL DISSERTATION

LAJOS KÖRMENDY

INFORMATION TECHNOLOGY IN ARCHIVES

(ABRIDGED VERSION)

2007

1 Objectives and method of research

The main objective of my research was to outline the impacts of IT on the archival science and practice. By the beginning of the 21st century both have been reshaped by IT.

For me one of the main questions was whether there is a professional constant what stands both in traditional and electronic environment, and if yes, what it consists in. Revealing this professional constant or „core“ is important in some respects. On the one hand it can prove that archival science has constant rules, and this is an important criterion of scientific quality. The question is how the principle of provenance — that has sometimes been applied by archivists for several centuries, and was formulated in scientific way for more than one hundred years ago — is substantial after the challenge of IT, especially of the electronic records. On the other hand revealing the „core“ may help to solve the greatest dilemma of modern archival science: variability is maybe the main characteristic of IT, but it is against the archives, and for this reason the archivist's community steadily seeks for the (technical) constant in changes. When I wrote this dissertation my main ambition was to underline the firm and fixed points.

Since IT has penetrated into every angle of the profession, it seemed necessary to deal with every important field of archival practice and theory (description, digitisation, electronic records, etc.). Besides outlining a general view, in order to find the „core“, I felt also necessary to analyse in detail the most important points (e.g. archival information), and to do this I had to dig into the deepest layers of archival science.

Nowadays IT in archives has so many aspects and factual data, that there is no professional polyhistor who first-rate expert would be in every respect. In my research I mainly relied on international professional literature, particularly in such fields like electronic records, where I have no personal experiences. Although many works are inaccessible in Budapest, internet assisted me a lot, because due to this I was able to read many missing articles, sometimes in abridged edition or in „second hand“ form. Besides the professional literature I freely drew from my own researches and experiences whose accomplishments I shared with professional public in many publications in the last twenty years.

2 The structure of the dissertation

In the chapter two, which follows the introduction, I endeavour to reveal those societal driving forces which have had direct impact on change of philosophy of archives. As these forces are in dialectic relation with IT, its functioning and impact on archives will be easier to understand.

The short third chapter deals with planning and elaborating archival IT strategy which is of vital importance.

The themes of the fourth and the fifth chapters are tightly related to each other. The former deals with the elements of the archival information and the structure of the archival material. The latter is about the form of presentation of archival information, i.e. the archival description, and sets forth in detail the major international standards: ISADg, ISAAR and EAD.

The sixth chapter is on the one of the most promising fields of archival IT, digitisation, first of all digitisation of images, which makes possible to record facsimile reproduction of archival material.

IT has maybe influenced the most the archives by the internet — this is the subject of the seventh chapter: how the archives are able to exploit the potentialities of the internet, which requirements they have to meet when drafting and operating a home page.

The eight chapter deals with the most acute problem of archival IT: the electronic records. Electronic records cause problems both for agencies and archival institutions because the concept and the form of the record is transformed, and as the electronic environment changes, authenticity and preservability of the records is doubtful. The main question is that how we are able to preserve the electronic records that they are authentic, integrant and accessible after centuries, too.

The last (ninth) chapter is about the archival systems, and I show how such systems work, first by listing general requirements, then by analysing the system called E-Archivum operating in the National Archives of Hungary.

3 The principal conclusions of the dissertation

3.1 Society, archives, IT

In the last decades of the 20th century tremendous mental, societal and economic changes took place in the countries belonging to the European culture, which have considerably transformed the philosophy (expectations, values, way of thinking) of the archivists. The changes of philosophy can be described in a few keywords:

- *Popularisation.* The archives which had been formerly the places of research reserved for academic and university scholars, have become institutions serving a great number of researchers, and making copies in bulk. The new researchers are mostly inexperienced, most of them want to search the history of his/her own family.
- *Transparency and openness.* The closed period of the records has become shorter, and in the same time more and more records have become accessible. The rule of 30 years of restricted access period has become general. The rules of classification of the records have been revised, and radically reduced the number of classified records. On the other hand the protection of personal data has become more severe.

The archives were interested in making acquainted themselves with the general public: they published a number of popular publications, made attractive homepages, organized popular programmes (exhibitions, open days), published regularly their annual reports.

- *Market approach.* An economic market approach has gradually gain ground. Cost effectivity and productiveness has become an important standpoint, the audit of archival institution has become common because the maintainer insists on rational management.
- *Media culture.* The archives, by choosing appropriate themes for their popular publications, homepages and programmes, may arouse wide interest of the general public.
- *Accelerating time and information centricity.* Nowadays the most important archives' communication tool is the internet. In many archival institutions they make retrospective processing works which means that the data of available publications and finding aids are loaded into new IT systems. The retention period of the electronic records has shortened a lot. The 15-20 years period what is common in case of traditional records is already untenable because by that time the IT environment will undergo a complete change, and the electronic records of today will not or hardly readable.

Every archival institution must plan its IT policy in long term. The IT strategy must fit into the general strategy of the archives comprising its other long term policies (preservation, supervising the records of the agencies, finding aids, etc.).

3.2 Archival information, archival material

Archival information consists of the elements as follows:

- *Identifier.* It makes precise identification possible. Identifiers are: reference code and title. The reference codes have an outstanding role in operating an IT system.
- *Content and structure.* The recorded information constitutes the content, the inside structure of the record(s) show(s) how the elements/groups are arranged, and what kind of relation are between them.
- *Context.* It means data and information, as well as legal, historical and societal references relating to the record creator and the record(s).

Archival information has a unique character owing to the structure and the context.

Information and data structure into record, records constitute a file, files form a sub-series, sub-series make a series, series form a sub-fonds, sub-fonds constitute a fonds. According to this we can distinguish (downwards) level of fonds, level of sub-fonds, level of series, etc.

The direction of the relations between the records can be either vertical or horizontal. The relations under the fonds level are called microstructure of the records. There are also relations over the fonds

level (group of fonds, management group of fonds, archival holdings), these constitute the macrostructure. The macro- and the microstructure form together the system of the archival holdings.

The research methods applied in the archives are adapted to the above described system. The contents of the records are mostly reached by vertical relation: the researcher approaches the record looked for step by step localizing narrower and narrower scope of records. He/she moves forward from the general to the specific, and this is called indirect access of information. (In library we can generally reach information in a direct way, by data of catalogue.)

3.3 Archival description

Archival description is an abridged and comprehensive recording of the information captures in the records, as well as information related to the records themselves and the record creator, which can be done in traditional or electronic way. Archivists have made archival descriptions for centuries, but IT has brought about radical changes in this field, too: owing to its integrating force separate finding aids became unitable, the differences between databases, handwritten and published finding aids have disappeared, even more, a passage opened between finding aids and integrated archival registers.

In the last 15 years the international archivist's community elaborated three worldwide accepted description standards, which make possible to describe in a standardized way the data and information of finding aids at any level from the fonds to the record: ISADg, ISAAR and EAD.

The concept of ISADg, according to the principle of provenance, is based on these levels, it shows the vertical and the horizontal relations precisely, i.e. maps the microstructure of the archival material described, but does not mention the macrostructure, although there are much information which are worth describing.

As regards the microstructure in my opinion it is not worth standardizing description below sub-fonds level. On the one hand the data and information captured in archival material are too dissimilar from each other. The data of census records, minutes of an assembly general, a map collection or court records are entirely different, which need different processing (data fields). On the other hand the quantity of data grows at a highly increased rate as we describe archival material at lower and lower level, and excepting some special record units it is useless to give comprehensive description on file or records level, it is absolutely enough to make a list of characteristic or relevant data or a combination of them.

The fields of ISADg precisely cover the elements of archival information (identifier, context, content and structure), and they are balanced. In the history of description (finding aids) the weight of these three elements changed by ages. From the beginning up to the end of the 19th century the content was important, old inventories and lists intended to describe more or less only the content of the records, context did not exist in these finding aids. The role of the identifier was practically constant from the beginning because one always needed it to identify the records. In our information age, especially because of electronic records both identifier and context had to be strengthened: there is no envelope, stamp, authentic signature, all of them must be substituted, and most of the substitutes belong to the category of identifier (e.g. metadata). After the spreading of the principle of provenance, in the 20th century, the importance of the context has gradually, in the last decades dramatically increased. In case of the so called principle of series (see hereafter) the very strong representation of the context is necessary, because the records are torn out of their structure, so their original relations can be reconstructed if a detailed description about the record is available. However, the context reaches an absolute dominance not here but at the theory of post-custodial archives (see also hereafter), whose representatives say that archivist's processing work is only drafting context.

By ISAAR you cannot describe archival material, it does not deal with the elements of archival information like identifier or content and structure, these are left for ISADg. ISAAR only serves for describing the record creator and the circumstances of the record creation, i.e. context. ISADg has a separate context field, however ISAAR does not want to replace it, but recommends to link and use both standards. The reason of this is in the principle of series elaborated by Australian archivists.

In the principle of provenance the organisation of the archival material is based on the fonds, and a fonds is built on one record creator (agency or person). But agencies are often undergo a transformation, terminate, fuse or separate, their function or competence changes, and the current records get to the legal successor if necessary. For this reason archivists are compelled all the time to close fonds, and open new fonds, while the scope of the record creators practically remains the same. In order to

eliminate this inconvenience Australian archivists elaborated the concept of series. It is based on the series, a stream of records, which means records which belong together because they part of a filing system, or they resulted from the same activity, or they are of similar format and relate to a particular function. The record creator does not play an important role in assembling record series. For instance records relating to a given function (e.g. state forest-administration) kept in several fonds are united in a series. Since a series can be a (full or partial) product of several agencies, it is not advisable to treat together the record creators and the records in description, even more, it is more expedient to separate them because they are asynchronized. One cannot describe a series with ISADg, or rather it is possible but with many repeating, because if the records of one agency were distributed between 6 or 8 or 10 series, then the description of this agency should be repeated at these series, and this is redundancy, what is in contradiction with the basic principles of IT. This is why it was necessary to elaborate a separate standard, and unfeasible to expand the context field of ISADg.

EAD is based on SGML/XML, a hardware and software independent descriptive language, developed for the internet. SGML/XML has a separate segment called DTD where one can define tags, which separate the different types of data. EAD is an archival DTD.

EAD approaches the question of archival description from a different point as ISADg or ISAAR. It considers the product of description as electronic record, and wants to give answer to the challenges of electronic records. The descriptions made by EAD are hardware and software independent, long lasting, easy to convert, modularizable, have standardized elements, and make possible various use. With this standard one can describe all kinds of archival material, redescribe the available finding aids, i.e. one can unite all the descriptions into one framework. In the same time the archives can create their own safe electronic archive.

3.4 Digitisation

In case of digitisation and digital copies the most important archival principles to be respected and requirements to be met are as follows: security, arrangement, accessibility, metadata, authenticity and system.

It is hard to meet the *security* requirement — which means that the medium be durable, and the information readable for long time — because on the one hand the media are rather unstable, their life-expectancy is uncertain, on the other hand they are very vulnerable, and the data corruption cannot be detected by naked eye, and finally because there are no long lasting standards in IT, and what is readable now will not sure at all readable ten years later.

For security reasons it is a good solution when the digitised image is made of a proved safe medium, a microfilm. To preserve digital copies it is advisable to apply processes which reduce risks like using a quality medium or converting and recopying at intervals.

In case of digital images orientation among electronic record is much harder than among traditional ones, for this reason the record to be digitised must be definitely *arranged*. First an electronic finding aid must be made of the arranged records, and the digitised images must be linked with the appropriate data of this finding aid, so the images can be searched, listed and browsed by the assistance of these data.

The data of the electronic finding aid can be taken as metadata. (Metadata: data about data, in this case data concerning the digitised material and the file itself.) Metadata are particularly important in case of image and sound files, because contrary to text files one cannot search them by words or terms, so the only tool of searching are metadata. There are three types of metadata:

- Descriptive, which are necessary to identify and explore the data.
- Administrative, which are necessary to use the digitised material (e.g. technical specifications, data concerning the digitisation process, information relating to accessibility).
- Structural, which are data concerning the connections between the components of digital objects (archival or technical).

The best thing to do is to put each image into a separate file, and to attach a metadata record to each file. The latter must be linked with the files because they make sense together. A part of the metadata can be integrated into the file names. in case of image files it is recommended that the names be „talkative”, i.e. contain the exact reference code of the digitised record.

The greatest problem of digital copies is that they are easy to manipulate, and there is no concrete evidence of the intervention. The legal solution will probably follow the model of microfilms which means that authenticity is assured in an indirect way, by regulating the process of creation, and by appropriate and mostly automatic documentation (see more at electronic records).

The *system* requirement means that digital objects and information must be integral part of the archives' whole IT system.

Archives may digitise their finding aids and the records they keep. The great advantage of digitising finding aids is that since all archival researches start with studying finding aids, this means that the majority of the researches concentrates to there, so by digitising relatively few pages one can reach high rate of use. It is worth, first of all, digitising such finding aids as image whose converting (by OCR) is unfeasible.

Although nowadays millions and millions of digital copies are made, it would be an illusion to think that the whole archives' holdings can be digitized within reasonable time. Since making electronic finding aid is a particularly labour intensive professional work, it is worth concentrating the forces to digitisation of heavily used large record units.

Digitisation must be planned carefully, archives have to have a related strategy. First one has to determine the professional aims to be reached (for example: to assure convenient use of heavily searched records, then one has to appraise the tools to be used (investment, personal and financial conditions), and how the records to be copied will be selected (e.g. digitised copies ordered by users will automatically be part of the digital collection), and one must decide who will digitise (e.g. external contractor).

One has to determine beforehand the professional and technical requirements, for instance

- rules concerning handling the records (practically these are the same as in case of microfilming),
- the file formats to be applied (TIFF for preservation, JPEG for communication),
- other technical requirements (sharpness, contrast),
- the system of the file names (the structure of the files, the system of the „talkative“ file names),
- the metadata that must be recorded,
- the media to be used (tape unit or DVD for preservation, DVD or hard disc for everyday use).

3.5 Use of the internet

The internet is as a matter of fact a worldwide web of networks without hierarchy which connects many millions of computers, which can be both data suppliers and data users. The internet makes a two-way relation with the world possible, which means that not only the archives but the users also may take the initiative, communicate data or launch „transactions“. Apart from a simple data supply the archives may open their virtual door before visitors and researchers, who can read electronic data watch digitised images like in the archives' reading or exhibition room.

The homepage, the internet window of the archive, must be planned carefully, and maintained with efforts. In order that the homepage serve well the archive, it has to meet the following requirements.

By the internet the archives send a message of communication to the world in which they give an image about themselves, and determine the aim to be reached. For instance: it is an open public servicing institution where considerable scientific work is going on, and which receives national and foreign visitors and researchers hospitably. The form, the structure and the content of the homepage must be shaped according to the message of communication.

One may use attractive and high colours, but it is not advisable to go too far, because it is against the image of a scientific institution. However, if we apply too many decorative documents as illustration, the homepage will suggest an image that the archives are a place of collection of old and beautiful documents, where serious scientific researches go on — this impression is not compatible with the image of modern archives.

The texts of a homepage must be professional and easy to understand, the general reader does not understand a too scientific text. However, it would be not wise to be too informal, the archives should not take the style of such people who are unable to make any archival research.

The archives must serve different user groups (archivists, genealogists, general public) simultaneously, and for this reason to apply different styles. The documents for the different groups are to separate in the menu system.

The menu system shows the structure of the homepage. It is recommended to divide strongly the content into blocks, but not too much, because the user can get lost easily in a complicated menu system. One must seek after that the user can always see where he/she is.

It is recommended to create the blocks as follows:

- Presentation of the archives (address, telephone, e-mail, fax, where the archives are, how to get to there, opening hours, what the mission of the archives is).
- Information relating to research (where the reading rooms are, opening hours, research rules, rules of copying, price list).
- Information for beginners and inexperienced researchers (methodology mostly for genealogists and researchers studying local history, examples).
- Description of archival records (electronic finding aids, on-line databases).
- Collections of documents, exhibitions (a collection with scientific annotations or studies in popular style and illustrations from the holdings of the archive, as well as material of exhibition).
- Publications (studies, periodicals, finding aids) .
- Information for professionals (rules, methodological works, guidelines and other materials for the staff or the professional community).

3.6 Electronic records

The history of the archives in the last 100-120 years shows that quality changes which had a considerable effect on archival science took place every 50-60 years. At the end of the 19th century the developments of the European bourgeois societies required that the public and the private records be preserved and retrievable. The response of the archivist's community to this challenge was the principle of provenance which made possible that the archives collect and process efficiently the records created in bulk by agencies and private persons. The next great change took place in the middle of the 20th century when — mostly because of the more and more intensive and extensive expansion of the state — the quantity of the records created (and to be processed) increased dramatically. This time we can state a series of new methods in appraisal, disposal and the appearance of microfilming and preservation. As regards the third boundary line, we have reached this now with electronic records.

There are many differences between the traditional and the electronic records:

- Electronic records can be read only by complicated tools (hardware, software).
- Electronic records have much wider spectrum than the traditional ones (word processed texts, spreadsheets, databases, images, sound documents).
- In case of electronic records the information is separated from the medium, and the data are kept dispersedly, only the computer provided with an appropriate software knows what belong to what.
- In electronic records the form is often separated from the content. The form is affected by hardware and software.
- In electronic records the content can also be fragmented. A record can be physically kept in 5-10 separate files, and only the software „put them together” on the screen.
- Electronic records have typically much more complicated structure than the traditional ones (see for instance the role of the links in internet).
- In traditional fonds the records generally reflect the structure of the agency, in electronic records the boundaries are much more indistinct: the staff of different units work wit common documents and datasets (databases).
- In electronic records — due to hardware and software dependency, the fragmentation and the complexity of the records, as well as the difficult orientation — there are metadata which play an outstanding role.

- Electronic records are intellectually very unstable, they change very fast and often, the traditional records have much longer life.
- Electronic records are physically much more unstable and vulnerable than the traditional ones: the discs and tapes lose their magnetism in a few years, the optical media can be destroyed by a scratch.

In case of electronic records the weight of information elements are different compared to the traditional records. The weight of context increases, its concept is wider, because it means not only legal, administrative, economic, historical, etc. information concerning the agency or the person which created the record, but also IT connections.

The weight of the structure element is also increased because one has to register precisely the relations between the records and the units of records (tables, files). The context and the structure element can be strengthened by detailed documentation and metadata.

The major problems of archiving electronic records are as follows:

- Dependence on hardware (because fast changes and various standards the old electronic records are not readable any more with new equipment).
- Dependence on software (the range of software is too large, most of the are incompatible with each other, the change of standards is even faster than in case of hardware, for this reason the software obsolescence is very rapid).
- Reserved codes, proprietary rights (the codes of file formats are mostly not open, thus the outsider software experts are handicapped when writing application programme; most of the software are copyrighted, the agencies must not transfer them with electronic records).
- Authenticity (because of the separation of the form and the content, the easy and undetectable changeability of the content, the necessity of converting due to obsolescence, authenticity of electronic records is rather weak; to keep authenticity of a record one has to preserve its identity, integrity and usability; to meet these requirements one has to comply with a series of rules, and to establish securities such as the regulation of creating and handling records, the appropriate documentation [tracing], access right system, metadata, regular checking and migration).

The records or content management systems, which are widely used in business sector, are not solutions to the above mentioned problems. They include various types of files (records, units of records) into one system, and manage them. They have the viewing and the browsing functions of the most popular applications back to approximately 10 years, but longer period is over the business interest. On the other hand these systems concentrate on simple and single records and tables, but are unable to represent the relations existing in record systems.

Regulating the electronic records created at agencies and individuals was necessary first of all because of the need of authenticity and accessibility. In the countries which have developed electronic government the requirements, models or standards serving standardisation have been elaborated. These generally concern the records management systems like INTERPARES or MOREQ.

INTERPARES sets out from the electronic record itself, examines it, takes it to pieces, and judges everything how it serves as administrative or legal evidence. Although it seeks after covering the whole life cycle of the records — contrary to MOREQ, which does not deal with the archival cycle — the requirement system of INTERPARES is very static, hard to survey, it simplifies because of its unique perspective. MOREQ sets out from functionality, and for this reason it is not static but easy to survey, flexible and versatile. The creators of INTERPARES represent the record paradigm, who consider the record only as evidence, and thus they absolutize the record's context, legal value as well as formal and structural characters.

The so called Dublin Core has become the „world's standard" of metadata, however, archives can apply it only with alterations and contributions. Namely, from archival point of view it has serious insufficiencies because on the one hand the terms used only partly comply with the archives' ideas, on the other hand it does not take into account the most important characteristics of archival records: the hierarchy, the structure, the levels. It does not follow the life cycle of the records, and particularly it does not distinguish the phases at the agency and the archive.

The standardization of the file formats of the electronic records is a key issue for the archives. The formats change as rapidly as software, by the way their use is hardware and software dependent. There is no common position in the international professional literature and practice concerning the standard file formats, we can find only recommendations on general, theoretical level at best:

- they should reproduce all important information and relations;
- they should have open codes recognized by national and international standards;
- they should be long-range and widely used;
- they should be accessible and able to further converting;
- they should be hardware and software independent;
- they should be suitable for reversion into the original system.

According to our knowledge the following formats meet the above listed requirements (more or less, sometimes with compromises):

Simple texts: RTF, DOC, PDF, TXT.

Structured texts: SGML, XML.

Spreadsheets: XLS, CSV, SGML, XML.

Databases: SGML, XML.

Images: TIFF, JPEG.

Videos and sound documents: MPEG-1 and MPEG-2, Quick Time, Windows Media Video, MP3, Apple Audio Interchange Format.

In respect of long term archival preservation there are four strategies that international professional literature recognizes as alternatives (or to be applied parallel with each other): migration, emulation, technological preservation, and the so called post-custodial archives.

Migration. It means that the files are converted from obsolete formats to new, appropriate formats according to the new standards. Migration is a general solution among the most advanced archives in the world. This strategy has a disadvantage: data loss is inevitable. However, it makes difference what kind and what quality of data is lost. One cannot lose at all or in the slightest degree data representing content or important functions. Nevertheless repetitive migrations may result accumulated losses, for this reason it is recommended to apply it as rarely as possible, at long intervals which still do not jeopardize data security.

The widely used, standard, non-proprietary, platform and application independent file formats give strategical advantage. According to our available knowledge, nowadays only one such a format exists: SGML/XML.

Emulation. It means that the original hardware and software (operation system) environment is imitated by a programme made for this purpose, thus one can run the data with the original software (application).

Making emulation needs exceptional IT knowledge because besides the original electronic records as well as the application programmes the archives have to keep and run the original operational system and the files which are necessary to run the application. In addition to this one has to describe the so called emulator specifications, i.e. all the attributes of the original hardware platform (e.g. speed, display, tools and peripheries, multiuser aspects) which will be necessary in the future to programme the emulator of that time. Besides one has to record the metadata and other annotations, too.

Finally, in case of archives applying emulation strategy all the disadvantages subsist which exist in case of transfer of original software:

- Learning how to use many application programmes makes greater and greater difficulty, both the archivists and the researchers, as the number of the programmes increases in the archive.
- As the hardware and the software (operational system) change emulation must be done again and again.
- Proprietary application programmes are often subject to licence charges, which — in case of accumulation — may mean high regular expenditure for the archive.

Technological preservation. It means that the archives keep not only the original data files but software and hardware, too. Nowadays this strategy is driven into the background because maintaining and keeping in operation a steadily growing obsolete equipment park causes inextricable difficulties. However, in certain cases technological preservation can be justified. It is worth keeping some obsolete hardware elements which may be necessary to read some electronic records which will emerge in later times.

Post-custodial archives. It means that the archives do not preserve any electronic record, they remain at the agencies. The archives' duty is only to exercise intellectual control over the records through internet. One can make serious professional and ethical objections to this theory. With this the archives recognize that they are unable to solve the problems of electronic records, and shift the responsibility and the task upon the agencies which create the records. On the other hand the records are in the danger of destroying because the agencies are not interested keeping and maintaining and converting their old records, which are useless for them.

Preservation of the electronic records is one of the most important task of the archives. It has to choose those media on which the electronic records will be archived. When choosing the media the following requirements must be considered: stability, life time, general use (standardization), capacity, manageability, possibility of data change (important because of authenticity). In general way we can state that magnetic tapes (DAT/DDS and DLT/SDLT) or discs are less stable (they tend to lose magnetism) than optical media, and their life time is also shorter. They are hard to handle, and access to data is relatively slow (except for the hard discs), and their data can be easily changed. On the other hand they are widely used, and have a large capacity. The optical discs are more stable and have longer life (it is true, they are very sensitive to mechanical effect), are generally used, their capacity is rather small (but increases rapidly), are easy to handle, and their data are not changeable (except for rewritable CDs and DVDs).

It is recommended to keep electronic records in several copies: one (security) copy in a remote place, another (master copy) in optimal preservation condition, and a working or communication copy can be handled in ordinary conditions. In case of electronic records created by the archives themselves it is important to determine the so called back-up strategy: in what intervals one has to save the fast changing content of the records.

The use of magnetic media (except for hard disc) must be reduced as much as possible, high temperature and relative humidity as well as fluctuations are to be avoided, they must be protected from light, particularly from direct sunlight.

In Hungary archives still do not get electronic records, and for this reason regulation has concentrated on record creation at the agencies. The most important legal rules are as follows: the decree 7/2005 IHM on digital archiving of non governmental agencies, the decree 12/2005 IHM on the formats of the electronic records created in administrative procedure, the decree 13/2005 IHM on authenticity of digital copies, and the cabinet decree 335/2005 on the records management of public agencies.

The common feature of the listed decrees is that they do not consider and regulate electronic records in system, and for this reason they have to assure authenticity by electronic signature. The greatest imperfection of the format decree is that the format requirements for structured texts, spreadsheets, databases, videos and sound documents are missing, which practically means that the agencies have free hand to use any format.

3.7 Archival systems

An archival system must be planned carefully, it is advisable to integrate into it as much datasets and functions as possible. Every element (image, metadata, record, etc.) must have its predetermined place, thus in case of accession everything automatically gets its place. It is recommended to plan the system in that manner that new tasks and functions (management, registry, digitised records, etc.) can be easily inserted. The greatest advantage of a well planned system is that there is no redundancy, and every information is up to date. The system should be built around a database.

The system called E-Archivum, which was created by the National Archives of Hungary according to the ministerial decree 10/2002 NKÖM, meets only partially the IT and archives' requirements. Surprisingly some basic IT functions like printing, export or import are missing or very primitive. Besides there are serious archival insufficiencies in the system.

Under the name Central Registry we get a mixture of the registry and the summary of the records. We can see „registry pages” under sub-fonds level what is absolutely against the professional rules, on the other hand we do not know what was the last registry number. The data that we get in the Comprehensive Registry Structure do not meet the requirements of a summary of records, first because the data of year range and extent are missing, secondly because the records under sub-fonds do not suit here either.

The record levels (sub-fonds, fonds, group of fonds) are sometimes arbitrary and confused, there are unknown terms like „sub-fonds middle group”, and these confuse the whole system. For instance

between the management group of fonds and the group of fonds there are two additional levels, the so called „archives main group” and the „archival institutions of microfilms”.

It is a great problem that both the collection of the photos and microfilms were put into one table with the original records. Since these copies keep the level of the original records (which they were made of) the hierarchical order is confused, for example under the level of management group of fonds is the level of country or archival institution. The integration of the separate datasets (e.g. census records) into the same table causes similar problems, furthermore they have different data field which do not fit into the scope of data of a general registry, thus the programmer had to put them into the field called general description.

But E-Archivum is unable to search in the field of general description, thus several thousands of pages of information is practically hidden for the user. The search hits are arranged in order of reference codes (identifiers) what is gross error because the result is chaotic. To orientate well the user one should see the hierarchy at least.

Listing function is missing in the system, although it is a basic function in archival systems.

The description of the content can be done at „Description of fonds”, although it is applicable at other levels, too. The description can be made at two screens, the first one offers separate fields, the second one only one general field. The first screen’s fields looks like, to a certain extent, the fields of ISADg. Separate fields make a structure and make possible standardization, but only in case when the system supports it by appropriate searching, filtering, listing, exporting and importing functions, in order to make the most of the possibilities. Since this is not the case with E-Archivum, the descriptions are dead inclusions in the system.

The registry of agencies is an important module of E-Archivum. The ministerial decree ordains to keep data of predecessors, too. E-Archivum goes beyond this, and contains data concerning successors, too. This extra service is problematic both in legal and professional respect. The decree says to keep the data of existing agencies because this registry is a tool of survey of records management. If one also keeps the data of successors this means that the data of defunct agencies are the equal parts of the system what is against the intention of the legislator. A series of data (address, telephone, fax, contact person, etc.) is nonsense in case of defunct agencies. No doubt that keeping the data of defunct agencies is necessary, but the two things should have been separated, it would have been a simple task of programming. For the time being this is not the case, the two categories are not separated, and we can only deduct the status of the agencies (if it has a successor it is defunct).