Olivera Porubović-Vidović
(Archives of Serbia and Montenegro)

DATA MIGRATION IN ARCHIVES
OF SERBIA AND MONTENEGRO – CONCEPT AND EXAMPLE

Abstract. We present organization of database used to store metadata about backup i.e., about electronic files that make backup, and describe problems that occurred in the process of migrating data before database has been designed and filled.

Key words. backup, data migration, electronic document, database, metadata

1. Introduction

Apart from the possibility to be overwritten or physically lost, one of main threat for the integrity of electronic file is to become inaccessible because of technological change. The remedy for technological gap which occurs from time to time is data migration. On the other side, without metadata that describe files and help us to find the right moment for migration a whole process can be less usable than it is expected, with more human involvement than it is actually need.

Based on experience obtained through the data migration project realization, this paper intends to describe concept applied, and to report on the problems and findings in the process.

Remaining part of the paper is organized as follows. Section 2 describes backup storing concepts in Archives of Serbia and Montenegro. Section 3 presents experiences obtained from data migration, section 4 gives basic information on the migration data database concept. In the section 5, some conclusion remarks are given.

2. Backup storing concepts in Archives of Serbia and Montenegro

Computers have been using in Archives of Serbia and Montenegro from 1989. For a long time they have been used mainly as sophisticated typewriter machines; i.e. for business correspondence and the preparation of publications in The Department for publishing activity Therefore, text processor[1] was the main software that was used till 1997 and a great percentage of documents created in Archives before 1997. contained pure text only.

---

[1] Only a few texts have been generated by program package Ventura and all others used some version of MS Word starting with version 5.5.

During this period, backup storing procedure may be described as follows:

- final files' versions were kept on hard disks were initially they had been generated;
- one backup copy was stored on adequate floppy set;
- floppy sets were moved to safe place and were accessed only occasionally[1];
- no metadata for backup was defined.

As it is well known, key objective of any backup is in preserving a copy of original in such a way that copy can be used when original is damaged. Of course, in that case it is expected that a backup copy can be find easily and in short time. However, mentioned backups created in nineteen's did not fit these assumptions. Hence, other approach to backup organization had to be found. Apart from this initial reason, necessity of having metadata, i.e. descriptive information that facilitates management of and access to files that are kept, has become more obvious when mass document scanning has started in Archives.

In the first half of the year 2001, a special Act named: "Instructions for protecting data in information system in Archives of Yugoslavia"[2] was proclaimed. This act defines[3] obligations that everyone who is engaged in generating electronic document has, as well as basic metadata should follow those documents.

In the Archives, the first practical response to proclaimed Act was creation of backup files with associated descriptive data:
- type of file,
- software used,
- date of creation,
- name of person who created the file,
- short description about contents,
- document size.

Note that backup files were made through three generations, and mentioned descriptive data "covered" the last – third generation. Later, after analyzing our needs and current state, and after investigating information on the concept of metadata in other institutions in the domain of culture, mentioned structure of annotated data used in Archives has been enlarged. In the database created meanwhile, which is described in section 5, following metadata are kept:
- unique file identifier,
- file name,
- file format,
- software environment,
- hardware environment,
- short description of file contents,

---

[1] Having in mind that the basic use was to print what has been generated by computer, old backup files have rarely been referenced and when such situation occurred, the reason was additional printing

[2] The base for this act was Regulation about measures for data and information protect in information systems in federal organisations (Official Gazette of Federal Republic of Yugoslavia no 23 from June, 2 2000.)

[3] By Instructions it was determined that during operational use of file its creator is obliged to take care of it, and afterwards it is delivered to IT group for final preserving.

- key words,
- belonging to a certain group of files,
- date of creation,
- name of creator,
- organizational unit within Archives,
- storage medium,
- storage medium label,
- migration date,
- short explanation about migration,
- date when backup file was used,
- reason that backup file was used,
- the name of person who used backup files.

3. Experiences from data migration: 5.25" floppies to CD-R[1]

Recently, Archives of Serbia and Montenegro realized project of data migration to the newer data storing media. For data migration existing copies on floppies were used, and copies on hard disks where files were initially created were used only in a few cases when data could not be read from floppies. Data on floppies were compressed, and because of that program ARJ had to be used as decompression tool.

Number of floppies revised was 105, and only several, actually 5 floppies, were unreadable. Namely, content from one floppy was reconstructed from hard disk, and unfortunately, contents of another 4 were lost.

During data migration following jobs were done:
- each floppy was checked for viruses
- each archive was unpacked to one hard disk dedicated for that purpose
- content of every file was identified
- key words and short explanation were added for each file
- links between files, usually under one archive, were noted
- files were saved in new format

Being the first migration in Archives, this was the first practical experience concerning the metrics about work that was done. On the other side, data migration project has provided us with practical experience about problems that may arise as a consequence of having no additional data about files were kept. As all files were created in DOS and Windows 3.1 and 3.11 maximum file name length was 8 characters, which was insufficient to give a name that can be explanatory enough concerning the content of a file. Therefore, every text file had to be read and then, for every file a short explanation about contents was made.

Having in mind that till 1998 Archives has a modest number of computers in its ownership and that only a few employees were familiar with computers result that total amount of data did not exceed 271 MB is not surprising. Also, during the process it became obvious that only few backup files were referenced, and that there was no evidence about any other purpose of the files except for printing document. Actually,

---

[1] Till 1995 5,25" floppies were used, from 1998 the main medium were zip diskettes and in 2001 first CD writer was bought

this first data migration was realized after so many years because both software and hardware were available for a long time, without need to change anything.[1].

Another "problem" arose from a fact that documents were written in different fonts with Serbian characters made for different code pages. Hence, after opening files in new program version, local characters š đ č ć and ž were unreadable in almost all of them Furthermore, problem with code pages was even more complex regarding to the most common behaviour till 1997 when each new computer brought new OS and forced implementation of new code page. From nowadays perspective it seems that these small steps in changing, actually upgrading OS and application software, didn't bring any benefit to backup management,. Prolonging these upgrade steps to every fifth year was assumed to be better strategy.

From 1997 till 2000 we stick to Windows 95/98 and Office 97 and a year ago we passed to Windows 2000 and Office 2000. Two years ago we ultimately stopped using specific codepages and Unicode is used wherever it is possible. Normally, to convert documents made in old codepages into Unicode we use word macro converters for MS Word.

Finally, it should be noted that the first data migration did not give absolutely identical text documents, that is the content is preserved but as for fonts only font group (serif and no serif) is preserved.

The attempt to use Adobe Acrobat, directly from some old Word versions, left us without result because even we embedded fonts previously used some characters still stayed unreadable. For better identification, all documents are going to be classified by theme or contents and to any particular document brief explanation will be added.

4. Archives' database of metadata concept

Nowadays, Archives' backup consists of: text files, image files, web site, all printed publications prepared by computers and a few databases together with application software[2]. Database of metadata used for searching and defining the right time for new migration and backup of existed databases' content as well as contents of all other files defined for long term preservation are carried separately. Database was designed and defined in program package MS Access 97 and Visual Basic 5.0 was used as programming language and their main structure is shown in the Fig. 1.

5. Conclusion remarks

A so called "information era" we live in, brings us both new possibilities and challenges. Concerning mass creation of electronic documents, ensuring their longevity is one of the main challenges one has. Everyone has to be aware that rapid technological evolution means rapid obsolesce, both hardware and software, and adequate steps have to be done in making and carrying strategy that will help to override all threats of losing electronic documents. Not to mention loosing time and money that have been spent and human efforts that have been put.

---

[1] Nowadays 5 PCs with processor 486 are used and it was last year that we stopped using PCs with processors 386 and 286.
[2] Following applications are active: Registry of equipment, Reading room work, Inventory – finding aid, digital exhibition of archival holdings – exhibitions organised by Archives
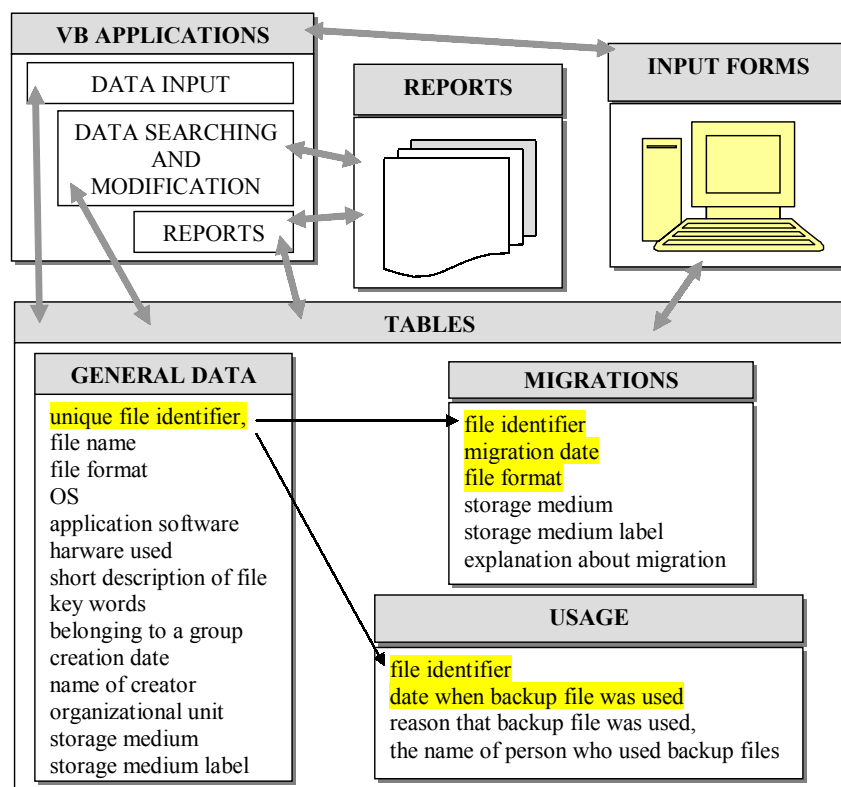
Fig. 1 – Simplified database concept

One of possibilities to provide what have been said is conveying periodical data transfer from one to another software/hardware generation. Availability of adequate metadata surely simplifies this process, and among our first steps, in building strategy for ensuring document longevity, was defining metadata we want to keep and designing database.

References

1. *Instructions for protecting data in information system in Archives of Yugoslavia*, 08.02.2001, internal document

2. Minnesota Recordkeeping Metadata Standard
http://www.mnhs.org/preserve/records/metadatastandard.html (accessed 27.05.2003.)

3. National Archives of Australia: Preserving Electronic Records through Migration
http://www.mnhs.org/preserve/records/metadatastandard.html (accessed 14.01.2004.)
4. PRO : Requirements for Electronic Records Management System
http://www.pro.gov.uk/recordsmanagement/erecords/2002reqs/2002metadatafinal.pdf (accessed 31.08.2003.)

5. Recordkeeping Metadata Standard for Commonwealth Agencies
http://www.naa.gov.au/recordkeeping/control/summary.htm(accessed 31.08.2003.)

6. Regulation about measures for data and information protect in information systems in federal organisations (Official Gazette of Federal Republic of Yugoslavia no 23 from June, 2 2000.)

7. Kenneth Thibodeau, Building the Archives of the Future
http://www.dlib.org/dlib/february01/ thibodeau/02thibodeau.html (accessed 20.06.2003.)

8. Woodyard, Deborah: Farewell my Floppy: a strategy for migration of digital information
http://www.nla.gov.au/nla/staffpaper/valadw.html (accessed 31.08.2003.)

mailto:admarhiv@gov.yu