

Theoretical and Technical Solutions for Preservation of Electronic Records in Finland

Outsourcing as a temporary solution

The National Archives of Finland faced a severe problem at the end of the 1980's. Electronic records had replaced traditional paper records in many cases. Though the National Archives preferred - and still prefers - paper records to electronic ones, it was impossible to generate hard copy records from electronic records in all cases.

The reason is simple. The capacity of a computer enables it to handle records which are unmanageable in paper or microfilm. It is obvious that appraisal can't be based on the physical format of a record, but on the information the record contains.

Our first attempt was to outsource the preservation function to in the main computer centre of Finland in 1989. Information was stored in 'traditional' way, by 1/2" magnetic tape in mainframe environment. It was, without a doubt, a safe solution. But outsourcing had its obvious disadvantages, too.

Preservation was costly. The charge was 50 FIM/reel each month, including backup. This meant that the cost of preservation was calculated in a different way for electronic than for paper records. This fact had an influence in appraisal practices whereby preserving data was considered as 'expensive' while paper was regarded as 'cheap'. It meant that a great amount of electronic records was not preserved because the solution was considered too expensive.

Access to electronic records was difficult. It was costly to load a tape and run queries in a mainframe. Mainframe serves well as a multi-user database server, but it is too robust a tool for archived material. The basic problem of archiving data is to provide access to a very large volume of very little used data. Only in very rare cases do you have more than one simultaneous user for the same file. You don't need to worry about the speed of transactions - so you don't need a mainframe.

One main point was whether it was wise for an organisation to outsource its key functions. It was clear, that the use of a computer centre had to be a temporary solution.

by *Matti Pulkkinen* *

Solutions in middle of an economic crisis

In the beginning of the 1990's the economy of Finland suffered a major setback. In 1991, the gross domestic product diminished 7,1% and continued to diminish until 1994. Government consumption expenditure sank, too.

It was a crisis. For the National Archives it wasn't possible to invest in a tape repository, although the need for a preservation system of its own became evident. It was, also impossible to have more personnel.

That made us think about what we actually did need.

Who should be able to preserve more.. If we don't have data, then nobody asks for it, and we may think that nobody needs data archives. This circular argument was sometimes made, when we considered our task.

Safety is a major problem

Secondly, we should do it in a safe way. Safety means safety in every respect, and it means both safety of material and safety of citizens, too.

When we think of the long-term preservation of a record, we must take into account the ageing of the material and other such threats. Also, we mustn't forget that the society tends to change as well.

Let us take an example. In the year 1897 a protocol of the Överstyrelse för pressärendena - the Supreme office of press affairs - written and then preserved. It is there, in the repository even today. In one hundred years it has 'lived' through a coup de état in 1899, a general strike and rioting in 1905, abolition of Överstyrelse för pressärendena and rioting in 1917, a civil war in 1918 and series of bombardements in 1939 - 1940 and in 1941 - 1944. It has been preserved through economic crises in 1973 - 74 and 1992 - 94.

If we think of the permanent preservation of electronic records, we must accept, that the future is uncertain. We hope, of course, that the next century is happier than current one, but we can't be sure of it.

In the preservation of electronic records you always need

some manpower and supplies. You can't just forget your disks or cassettes in a repository. There is a risk, because a severe economic crisis may make the performance of your routine duties impossible. If you run out of money, you may lose your data.

The nature of an electronic record makes its preservation during a severe crisis a difficult task. If we think of threats like a hostile occupation or a totalitarian coup de état, the main benefits of electronic records - , easy access and rapid altering - become threats. The horrors of Rwanda were enabled partially by misuse of census records. .

Safety by replication

We selected a file in logical sense as the object of preservation. A physical media can't be preserved readable long term. File formats and physical formats become technically obsolete. A record is a semantically coherent set of information, a semantical entity, and not a physical one. It is, however, true that a record can't exist without some physical container for the information. As a corollary we can say, that you can use different types of containers for a record. This means in practical level that we have freedom to use any media we want to preserve a record as long the semantical integrity of a record can be guaranteed.

We can replicate electronic records. Indeed replication is a key technique in our concept of security. We always use two different media in preservation. We take a master copy in DAT and a backup in 8 mm, and if we want, a CD R-copy, too. DAT copy is preserved in National Archives, backup in high security deposit in another place.

The choice of storage media was dictated by the lack of money - we couldn't possibly construct a traditional repository for 1/2" tape. By selecting a more compact media, we lost perhaps some of proven security of traditional 1/2" tape. We can compensate for it by a faster cycle of recopying of records - once in five years. For data cassettes, it is sufficient to use a safe, secure storage vault as a repository. It is difficult to gain illegal access and it will keep the temperature and relative humidity on a stable level. Inside the vault the data cassettes can be stored in fireproof cupboards to secure them from fire, water and magnetic interference. There has been, alas, a difficulty with our use of cupboards. Relative humidity has been too high, probably because the insulation material contained wasn't perfectly dry before we started to use them.

However, these cupboards are easy to evacuate, if necessary. Changes of temperature are very slow inside the cupboard, and we can transport material for example in an open lorry in wintertime.

Downsizing with Unix

As a computer environment we use IBM RS/6000 C20 machines. The operating system is AIX 4. There are 1/2"

tape-, QIC-, 8-mm Exabyte-, CD R- and 2 DAT devices in the same machine, and we are about to connect a 3480-device in this machine, too. For security reasons, this machine works as stand alone, but we can temporarily connect it to our network, if we want.

Unix has been an optimal platform for archival preservation. It contains powerful script language, c-compiler (unfortunately an option in AIX 4) and all the device drivers we need. AIX is easy to manage.

You can do a lot with Unix programs like cut and grep; you don't need to load your data for example in a relational database to query it.

The cost of investment for the construction of the vault and the purchase of storage cupboards and computer equipment is the equal of the cost of outsourcing the electronic records preservation function over a three-year period.

The value of records

The values that inhere in basically any records are of two kinds: primary values for the originating agency itself and secondary values for other agencies and private users. Records are, of course, originally created to fulfil the first of these: they are created to accomplish those purposes for which an agency has been created - administrative, fiscal, legal and operating, if we are referring to public records. But in archival perspective, more important is the second meaning - records are, after all, preserved in an archival institution because they have values that will exist long after they cease to be of current use, and because their values will be for others than the creators.

These secondary values of records can be furthermore divided into two kinds:

1. The evidence they contain, i.e. evidential value, referring to the value that depends on the character and importance of the matter evidenced.
2. The information that they contain, i.e. informational value, which may relate, in a general way, either to persons, or things, or phenomena. In modern archival science the evidential value is regarded as the principal value of a record.

There has been, however, little or no discussion about the nature of these two values. A closer study of both these aspects of preservation is important, and especially when we are discussing the preservation of electronic records.

Informational value

The semantics of informational value is quite clear. We can here adopt the so called 'the naive theory of truth' on the semantic issue. From an epistemological point of view this theory in itself is not sufficient, but altogether it is 'good

enough' to be used in logic. This theory of truth implies, that, for example, the sentence 'there is an A' is true only and only if 'A' is a really existing entity. If there is no entity called 'A' in the world, the sentence is obviously untrue.

According to what we said earlier, we can conclude, that when any record includes true and meaningful statements, it also contains informational value. So if a record states, that there exist - or has existed in the time that record in question was created - an 'A', and if this statement is true - it can be verified to be true -, the particular record in hand contains unquestionable informational value.

Evidential value and speech act theory

We can't, however, adopt this theory of truth used above to the matter of the evidential value of a record. While a receipt can be genuine or not, but semantically it can't be said to be either true or untrue. In fact, the evidential value is always bound in the use of the particular record - the record does not have evidential value per se.

The semantics of evidential value has to be derived from so called 'speech act theory'. This theory states, that a performative act of speech is neither true nor untrue. A more coherent way to make the distinction is to use the terms successful and unsuccessful. A judge, for example, can impose a penalty on a criminal in the court. He has the authority to do so. In this context his sentence 'the court fires you the sum of 1000 pounds' is successful, when and if the judge follows formalities stated by the law. The judge can't use the power given by the law to him outside of the court. So his sentence of punishment will be unsuccessful when imposed outside of the court.

Some conclusions

We can now argue that the very nature of evidential value is based on semantics like this. It implies that we can use binary logic as a tool of falsification of evidential value. If and when we accept the argument generally shaped in this paper, we can furthermore define some formal criteria for the evaluation of evidential value.

Some essential - although maybe not all - questions of evidential value can be reduced in the two following concepts of authenticity and integrity.

The first of these, the problem of authenticity, is one of the key problems concerning the preservation of electronic records. This question can't be solved with those techniques used in data transfer. For example public key encryption is highly software-bound and sometimes also hardware-dependent. In archival preservation we cannot and we should not rely on public key encryption techniques. The other key issue in preservation, i.e. integrity, can also be easily violated, for example by losing referential integrity or transactional integrity.

So we can state, that - in the strict sense - evidential value is lost when authenticity or integrity is lost. Many preservation programs for electronic records seem to be unaware of these problems mentioned, and to which should be given more investigation.

* Paper presented at IASSIST/IFDO '97, Odense, Denmark, May 6-9,1997. Matti Pulkkinen, KANSALLISARKISTO.