The Information Life Cycle – Issues in Long-term Digital Preservation

Runardotter, Mari; Quisbert, Hugo; Nilsson, Jörgen; Hägerfors, Ann; Mirijamdotter, Anita.
Depatment of Systems Science, Luleå University of Technology
{Mari.Runardotter; Hugo.Quisbert; Jorgen.Nilsson; Ann.Hagerfors; Anita.Mirijamdotter}@ltu.se

Abstract. In this paper a new view of the life cycle of information is described. Information never dies. The constituentes for information are: the context in which information is born and flows, internal attributes which enrich and gives better understanding of information, and the tools used for managing distribution, dissemination, access and availability of information.

Introduction

In this paper we discuss the concept of information in relation to long-term digital preservation. This is a most acute question for all sorts of “preservation institutions”, such as archives, libraries etc. The society is today facing a problem that might stem from a momentary loss of awareness. The problem concerns the fact that if digital records produced today shall be accessible in the future we must secure this today. As Brand describes it:

> Due to the relentless obsolescence of digital formats and platforms, along with the ten-year life spans of digital storage media such as magnetic tape and CD-ROMs, there has never been a time of such drastic and irretrievable information loss as right now. If that claim seems extravagant, consider the number of literate people in the world and how much work is "knowledge" work, which increasingly means computer work. The world economy itself has become digital. This is a civilizational issue. (Brand, 1999)

Just as Brand states, this issue is a concern for the society as a whole, especially for organisations and institutions set out, or assigned, to preserve digital material. Amongst them are archives and libraries, but also governmental and local authorities and administrations. However, even enterprises that develop information systems will need to be aware of, and prepared to handle, this problem since preservation measures need to be built in from the start (Bearman 1994). This problem is today identified and acknowledged. Several projects and institutions around the world have taken on the challenge and are working on finding solutions to the problem[1]. These projects are mostly concentrated on research and developments activities or education.

The Swedish National Archives and Luleå University of Technology are running the Long-term Digital Preservation (LDB[2]) research and development project with the aim to develop methods and models for long-term digital preservation. Key concepts are openness and trustworthiness since the electronic records should be physically and technically preserved in order to be possible to bring back and made available over time and thereby used in various citizen oriented solutions.

Beginning our research, within the LDB project, naturally involved a literature review. Since our intention is to complement each other, in order to provide a broad view on the subject, the initial literature review made it clear that we needed a common view on what is to be preserved, data, information or knowledge. The reason for this is that it will have consequences for our further research as well as for what methods, models and techniques that will be developed and used. Hence, this paper, which is a result of the literature review, will describe our common view on what to be preserved in a long-term perspective. This paper therefore discusses the concept of information as central for long-term digital preservation. Furthermore, we approach the issue from three different angles that are essential for digital information accessibility; context, embedded attributes, and information processing tools.

The paper is divided in four main parts: the first part discuss the notion of information, that is, what should or could be regarded as data, information and knowledge, respectively. We continue with implications of this in the archival world, and discuss consequences of this from the archivists' viewpoint. Thereafter we give an account of the embedded attributes of information in an Archival Information System (AIS) and finally we describe the role of information processing tools in the archival context.

The Information Life Cycle

The term information is used with different meanings by different groups and in different contexts (Langefors, 1993) and there is no well-defined definition of the terms 'data' and 'information' (Checkland & Howell, 1998). A common and short definition is that information is interpreted data (Bratteteig & Verne, 2000). Hence, data is signs used to represent information (Langefors, 1993), or signs

---

[1] E.g. Library and Information Science at University of Stockholm, Computer and Systems Science at the Royal Institute of Technology. InterPARES, CAMiLEON, CEDARS, the Digital Archives Research Project at Edinburgh University Library and the UK National Archives in Great Britain, NEDLIB in Holland, the Danish National Archives, the EU project MINERVA, the National Archives of Australia, and OAIS, NARA, ERA and San Diego Supercomputer Center at the University of California San Diego in USA (Ruusalepp, 2005).

[2] In Swedish: Långsiktigt digitalt bevarande.

which carries with them the possibility to compose and/or transfer back the text, photo, music etc. from which it is derived. Thus, “bringing back” the file provides access to the information and we can start reading/viewing/listening but also interpreting and analysing the information existing in the file. This results in the creation of knowledge, that is, we learn from information.

The central idea is that data become information by a process of interpretation (Langefors, 1993). The infological equation says that $I = i(D, S, t)$ where I represent the information we can get through the interpretation process, i, which is operating on data, D, together with our pre-knowledge, S, during a certain time, t (Langefors, 1993). Langefors states that “One of the central insights from infology is that data or texts do not “contain” information (knowledge) but will only, at best, represent the information to those who have the requisite “pre-knowledge”” (Langefors, 1993, p. 30) and he also means that the infological equation is a representation of this.

Dealing with information systems require a broad view of information, since using a computer implicitly involves information services of some kind, that is, we get service by being informed by data. According to Langefors it is necessary to define information as knowledge, since “information is knowledge and not physical signs” (Langefors, 1993, p. 113). He regards information as knowledge structured in such a way that it is communicated. Because of this it can also be stored, which leads to information being stored knowledge.

Checkland & Holwell (1998) provides the view that information is a service that supports decision making within organisations. Data are facts and a starting point for mental processes. They introduce the concept capta, which is the result of selection of certain data (we pay attention to, create some new category or get surprised by the data). In other words, data that catches our interest transforms to capta, a consciousness of something. This signifies that turning data into information is done through a mental process and during that process the data changes form, and will ultimately lead to knowledge. Once we have selected, paid attention to or created some data, or turned it into capta, we relate it to other things or put it in a context, we attribute meaning to it, and by this we once again convert it, this time to information. This can be done individually or collectively and this process, selection and conversion of data into meaningful information can lead to larger structures of related information, or what we call knowledge. The interpretation process Langefors (1993) emphasizes is implicit in this reasoning.

This is line with our claim – that information is not equated with knowledge. Just like data must be interpreted in order to become information, so must information be interpreted and analysed in order to be knowledge. Information is never the knowledge itself, the knowledge is within the knower (Goldkuhl, 1995). We agree with Langefors with his notion that in order to understand information people must have pre-knowledge. This we see as knowledges that consist of the ability to assimilate the information, but also to have an understanding of the information content, or the subject in focus. If these abilities are found, new knowledge is created. This can, in turn, be written down, or transferred to information again, and so the cycle evolves, in a never ending manner.

Finally, when thinking of information it is necessary to be aware of the communication aspect. Based on Langefors, Göran Goldkuhl (1995) has investigated a communicative action view of information, and he claims that there is a producer of information whose action is to create information, which later is to be achieved by the information user through interpretation. Goldkuhl (1995) perceives that the infological equation stresses the information user, but never explicitly refer to the information producer.

Information is thus created the moment someone (individually or in a group) thinks of an issue. When people’s ideas, or knowledge, are being shaped and transferred to a media they are made available for others to share. Hence, the knowledge has now once again become information and exists physically or digitally, available for people who need it in order to learn and increase their knowledge of something, for example in an organisation. This implies a very important phenomenon to observe – information never dies provided that it is preserved and used. It is constantly shaped and reshaped, evolving through time, picking up further information or loosing some parts but information is all the time under reconstruction. Information is a living concept, and forms a part in an ongoing process.

When information is created, stored and retrieved digitally new problems arise. It is pertinent that digital information is available to the future society for e.g. legal, historical and democratic reasons. But how to achieve this? Numerous issues need to be addressed. Some of the important ones are: the context of digital information in relation to the provenance principle and appraisal of public records, the attributes of digital information and its interpretation, the limitations of digital information processing tools and storage media. This leads us to a new view of the information life cycle in a long term digital preservation context as shown in Figure 1.

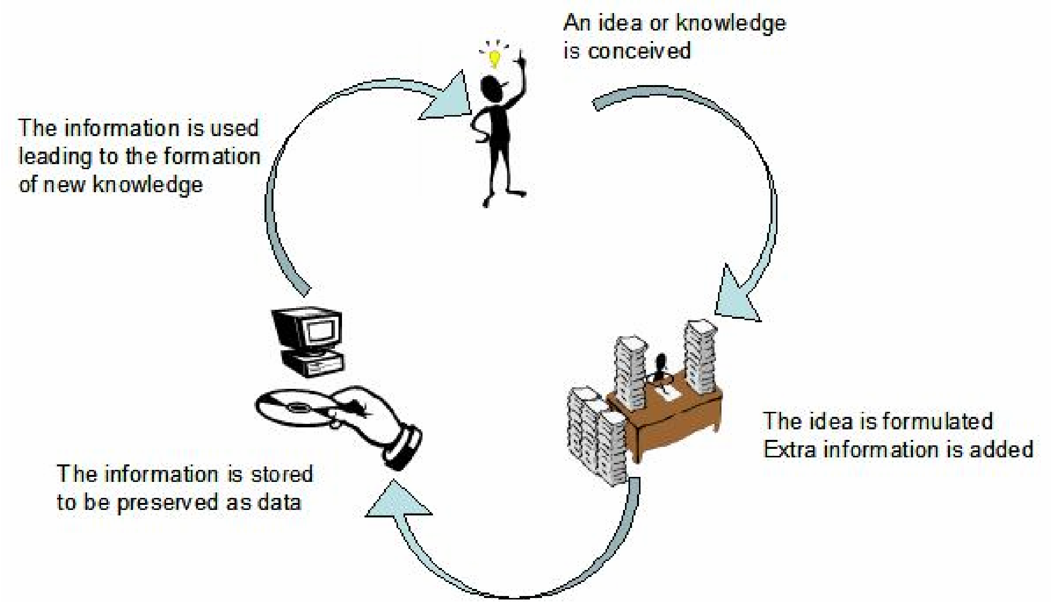

Figure 1: Our model of The Information Life Cycle in a long term digital preservation context

An idea is born in the mind/s of individual/s. The idea is then formulated as information and extra content information, that is embedded attributes, such as layout and colouring, is added. The information is then stored as a digital object where more information, preservation description information, is needed in order to preserve the context of the information. This is where provenance and appraisal is crucial. When the preserved information is used, with the assistance of processing tools, it is once again possible to transform the data into information which ultimately leads to new knowledge in the user/s. This is a constantly ongoing process and the responsibility for keeping this process alive lies on the archival community.

The Context of Information in the Archival World

In the archival world we can notice that, based on the discussion above, what reaches the archives is information, in the form of records created by information producers. Archival sciences is the scientific study of the nature of archives, their importance and development, and the study of methods needed in order to preserve, make appraisals and to make archives functional. The nucleus is systematisation and the principles of order (Ulfsparre, 2005). The archival profession rests on two pillars, appraisal and to arrange and describe the content and the context of archives. The fundamental idea with archives is that they should be houses of memory that holds the keys to the collective memory of a society. In these houses archivists were seen as ‘keepers’ partly because nothing in the archives should be changed after the records were archived (Cook, 1997). We want to challenge this view and means that archives should be regarded as a living and participatory part of the society where the information is made accessible and useful to citizens. The archivists’ role would then be changed from keeper to facilitator.

What, then, is a memory? It is to recall the past and its information, since a memory that only holds data would not offer us anything to remember if the means to interpret was not available. Furthermore, memories do, besides providing us the opportunity to remember, also plays the trick of forgetting. Until this day what is preserved in the archives is selections of information, they do not hold a complete and total storage of information.

The Provenance Principle

The provenance principle states that the original order or context should be preserved and explained, thus provenance stands for the origin of records and implies that these matters should be clear when records are being archived. The principle can be regarded as an apprehension of how archives are created, or as a method for how they should be managed when it comes to questions such as how to deal with making lists and appraisal, or in other words, what information is needed in order to make the archives functional and meaningful. The principle of provenance has in Sweden a twofold meaning, as work method but also as a guiding principle who will secure the preventive archival care (Backhaus, 2005). We regard the principle of provenance to state that it is information that is to be preserved since the context matters. Data without context is not possible to interpret and can never transform into information. Moreover, the principle of provenance inherits a conceptual interpretation – the societal perspective, which could be regarded as more appropriate in the digital age, but also a physical interpretation, which has to do with research. The organic context is however regarded as the most important to mirror (Cook, 1997).

Records have both primary and secondary values, according to Cook (1997). Primary value reflects their importance to their creator, while secondary value concerns their use for researchers. The latter is further divided to evidential value – which reflects the importance for researchers when it comes to documenting the functions, programmes, policies and procedures of the creator – and informational values. Informational value concerns the content of the records and relates to persons, corporate bodies, things, problems, conditions etc.

The moment the records are to be archived is interesting. As mentioned earlier it is information – not data – that should be preserved. We now have a situation where the interpretation vanishes unless written down as information. Furthermore, there must be added information of the information in order to secure the preservation of the context in which the original information first was produced. On the other end we have the situation when the information is asked for – that is, it should be brought back in order to once again be interpreted.

To arrange and describe is to provide the devices necessary in order to be able to interpret the records, which is to secure that information is preserved. This puts specific demands on archivists, since despite all guidelines and regulations on how to conduct this, it ends up as an individual judgment. This shows the importance of analysis and planning of archives in advance of the actual archiving activity (Cook, 1997). In other words, it is better to take on a pro-active role as an archivist, than it is to afterwards try to fix things that could have been done in a better way.

Appraisal

Appraisal of public records is by the Swedish National Archives defined as destruction of public records or information in public records. Destruction of information in connection with transformation to another data carrier is regarded as appraisal if the transformation leads to loss of information, loss of possible ways to gather the information, loss of possible ways to search for the information or loss of possibility to confirm the authenticity of the information (Gränström, 2005). As we can see has IT influenced this definition, new data carriers and new media are used that challenge the traditional way of performing the appraisal act. There is a need, according to Gränström (2005), to analyse in what way appraisal should be conducted within the IT community. Most essential to remember is that the use of the records, their potential, should never be jeopardised. In computer based information systems (IS) it is the principles for withdrawals- and work up of material, or the integrity and context of the archive, that are important to preserve, not the original order, since there does not exist one fixed order in computer based IS (Gränström, 2005). When it comes to appraisal the question is, what is sorted out – data or information? From which perspective should appraisal be conducted? Is it functions rather than individual records? And which way is the most suitable in the digital world? This implies that archivists should research and understand the functions and activities of archives. Archivists are active interveners and auditors in the archival document continuum (Cook, 1997).

Information Structuring in Archival Science and Information Systems Sciences

Functionality appraisal and systematisation and principles of order, we argue, show that the borders between archival science and information systems sciences are not easily detected. Instead there are a lot of similarities. Both archival science and information systems sciences deals with structuring information where the latter is especially concerned with the structure and order in computer based systems, whereas archival science takes on a more general grip.

However, Mats Burell (2005) argues that archival science might belong to the group of sciences that is especially influenced by IT. Similarities are found in the fact that the information system, within information systems sciences, is regarded to mirror the organisational activities, and thereby there exist linkages in the relation between information system – activities and archival creation – activities.

The differences are that in information systems sciences are systems considered to consist of parts that together constitute the whole and in order to understand the whole it must be broken down in these parts (Burell, 2005). In archival science is the premise that archives should be viewed and shown/presented as one total system, and only occasionally be presented or separated into a partial archive. The reason for this difference is, according to Burell (2005) that the main purpose in system science is not to describe existing structures, but to create new information structures that are not influenced by the traditional archival classification. This implies that the concepts that are fundamental in archival science, such as record, document or archival record, has not the same dignity in systems science.

Another difference, states Burell (2005), is that for the archivist it might be obvious that the original structure is important to preserve, since this traditionally has been important. For a systems designer it is the abstract model that holds the structure of the information that is in focus, such as routines, functionality, event or occurrences, and data or object. In other words, information systems do not contain fixed archival records or archival series, which implies that a re-formulation of the Swedish concepts archive and archival record could be, or maybe should be made (Burell, 2005). Furthermore, in information systems sciences there are not methods that in the same degree as in archival science deals with long term, “for ever” perspectives. This has not been detected as an issue in systems science until recently.

There is a huge need for development of common platforms, according to Burell (2005), in order to secure the long term digital preservation of records. Standardisation is used in order to decrease problems with transference from one organisation to another, which implies that possibilities and limits of digital long term preservation because of archival reasons is not separated from IT development in general (Burell, 2005). It should be a concern for all authorities that are expected to keep digital files to develop and maintain IT platforms that can unite the need for preserving older technical solutions and at the same time safeguard the use of new techniques.

The Archivist in the Digital Ages

What are the implications of this, for the archivists’? Contextual information is the archivist profession’s legacy and provenance-based retrieval of information, focused on form and function of records, and the context of creation, could be regarded as superior to subject- and content-based methods. The rediscovery of provenance and the richer understanding of creator contextuality are regarded to transform information to knowledge, and the relevance of provenance should be interpreted liberally rather than literally, conceptually rather than physically, if the principle continuously will revitalise the archival profession in the electronic environment (Cook, 1997).

Moreover, the traditional assumption of a one-to-one relationship between records and their creating administration is no longer valid – records are as changing and dynamic as their creating administrations. Cook (1997) also claims that IT professionals too often are concerned with efficient access and use of information and because of this loses sight of the essential qualities of integrity, completeness, accuracy and reliability, which is needed if the information will serve as evidence of actions.

> Unless institutions can thus be held accountable, which includes being accountable for ensuring that these qualities of "recordness" are present in their records-keeping systems, then any efficient access gained to information will be meaningless, for current and archival users alike (Cook, 1997, p. 40).

According to Cook (1997) there has been a strong emphasis on information content over provenancial context, library cataloguing over archival description and also of treating electronic data files as discrete and isolated items rather than as part of the comprehensive, multimedia information universe of the record creator. There are challenges if we are to maintain the traditional view of records when all new electronic formats change the foundation for the traditional scenario. But even if the information and communication technology rearranges the record matters, we still have a world of information relationships, interconnections, context, evidence and provenance. Hence, the essence of the archivists' task of comprehending and elucidating contextual linkages remains the same (Cook, 1997).

The new model of the information life cycle, described earlier, implies that archivists need new awareness, new skills, routines and work methods because their role will change in the future. What are the most adequate skills, routines and work methods in order to secure long term digital preservation? The question is still to be answered. However, in order to secure that the information is accessible and possible to interpret, archivists need to adopt new work practices. One of them is how to preserve the embedded attributes of digital information. This leads us to our second approach on the subject.

The Embedded Attributes of Information

As already mentioned, information can be a number of things. Allegedly one of the shortest, if not the shortest, correspondence between two parts was when Victor Hugo wrote a telegram containing only "?" to his publisher Hurst & Blackett, and got "!" in return. The correspondence concerned the publishing of his new novel "Les Misérables" and how it was doing.

Getting the Most Out of Data

Data needs to be combined in some manner to make information. The letters in this sentence is combined in a specific way to make words. The words are in a way just data in a more structured form. The meaning of the sentence is not inherent in its words or letters; it will need to be put in context to make sense.

So a sentence is more or less meaningful when put into context. But there could be more information hidden than the sequence and combination of words (or letters) presents. The text may be formatted in special ways by making it bold or italic which leads to a slightly different interpretation of the actual text. There may also be differences in size or colour of the text, as well as different fonts. All of those visual features of the text make us interpret the data slightly different than if there had been no differences in the visual appearance. So even if this kind of information usually is regarded as not being part of the information to be preserved, the ability to include such information must exist. This has not been a problem earlier since a paper document also carries layout and formatting besides the actual information. With information that is created digitally, and only exists in digital form, the problem to preserve layout and other visual attributes of information becomes more complex. Paper is a self contained carrier of information but digitally created information is only a stream of bits that need to be interpreted to a higher degree of information to be understandable to a human. This first step of interpretation is usually handled by a combination of software and hardware.

Take for example a form in a web based information system. This form probably contains a heading and some labels. It certainly has a background colour and input fields are positioned in a specific manner, and some text might be bold to emphasise its value. If a user of this form fills in data and stores them, they are probably stored in some kind of database, and when the user wants to open the same form again, the data is presented in its corresponding fields. This is the view that the user is used to see, and for authenticity reasons this is the view that should be preserved. The performance model (Figure 2) developed by researchers at The National Archives of Australia shows this concept in a rather simple way (Heslop, Davis & Wilson 2002).

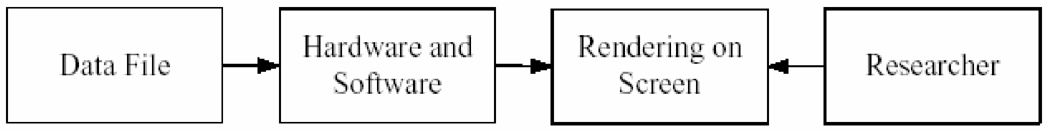

Figure 2: The Performance Model for a digital object (Heslop et al. 2002)

As shown in Figure 2, the researcher is actually dealing with a rendering of data. This rendering is produced through a process involving hardware and software that acts upon a data file. The performance (in this case the rendition on the screen) is what is worth preserving as close to the original performance as possible. Both the data file and the hardware and software are replaceable as long as the essential parts of the performance remain the same. This can be compared to an audiovisual record (a film) where the nitrate film is not the prime interest for archiving, instead the performance of the film that is of interest. (Heslop et al. 2002)

With regards to what is essential in a record, some non-essential characteristics are mentioned, for example the ordering of bytes in a data file or the data format of the document, whilst essential characteristics, in for example a word processing document, might be the actual content, formatting, colour, tables and figures. Determining which characteristics that are essential is a task for the archivist. (Heslop et al. 2002)

Performance and Context

Even if the performance is accurate, the context of the performance is of essence as well. Take a look at the classic example in Figure 3:

Figure 3: The importance of context (Preece et al., 1994).

Most, if not all, of you probably read the content of the figure as "the cat" although the figure in the middle of each word actually looks exactly the same. This is due to the (presumed) context and your previous knowledge of the English language. In another context (or another language) perhaps the "correct" interpretation would have been *the cht* or *tae cat*.

Let us say that the context of Figure 3 is *construction machines*. Now the meaning of “the cat” has changed to a vague reference to a caterpillar or a bulldozer instead of perhaps a pet, if the reader has background knowledge on the subject *construction machines*. This exemplifies the importance of context information, information about the information for the information to be understood and perceived in a reliable and trustworthy way, i.e. enacting the provenance principle. Even if it is archivists that are expected to judge what to be preserved and what is not to be, they will always benefit from tools that assist them in their task. In other words, what information processing tools that are available will be of great importance. Thus, we move to our third approach on the subject.

The Role of Information Processing Tools in the Archival Context

The latest year’s development of computers and software has been very fast and to a large extent unforeseeable. This has caused a number of conditions to change in society as we use the new artefacts (computers, PDAs, cell phones, tablet PCs, etc) for information exchange and dissemination. Large amounts of digital information have been created. At the same time, the technical evolution has created a number of problems. The problems of software and hardware dependency are pointed out by Rothenberg (1999), the longevity of storage media by ZD Net Australia (2002) and the ultimate goals of accessibility and availability by Dollar (2000).

Hardware and Software Dependency

There is a clear difference between dealing with digital material and paper based material. Traditionally data has been printed on paper according to some format making the data readable at first sight. Shifting to another kind of data carrier (magnetic optical, disc, etc.) reduces readability because the following two reasons. Firstly, in a conventional hard disk data is stored by “magnetic” points on the surface of the disk. The distance from point to point is so small, that it is impossible for the human eye to read the stream of magnetic points on the disk. Hardware is needed to facilitate the mechanical access to the stored data. Secondly, even if we could read the optic stream on the disk (by using a very strong microscope) we are unable to understand the meaning the optical (bit) stream[3]. A bit has a single binary value, either 0 or 1. A bit stream is a contiguous

[3] Bit is short for binary digit and makes up the smallest unit of data in a computer. A bit is a magnetic point on the hard disk.

sequence of bits, representing a stream of data. But what does a bit stream represent? A bit stream could represent different things, as described in Figure 4:

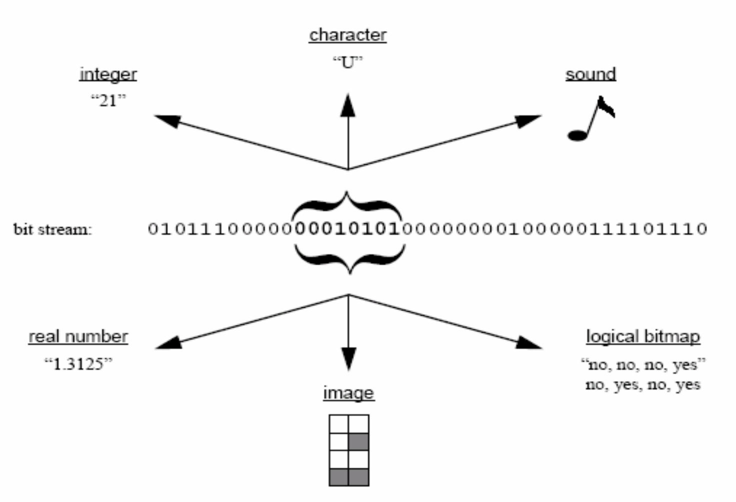

Figure 4: What a bit stream represents. (Rothenberg, 1999)

Making an analogy to the use of alphabetical symbols like the symbol "A" is readable and (at least in western cultures) understandable, but this is not the case for a bit stream. An interpretation facilitator (software) is needed in order to render the "meaning" of bit streams.

Storage Media Longevity

There are several kinds of storage media, such as hard disks, compact disks (CD, DVD), magnetic tapes, magnetic disks and static memory (Flash memory). Every kind of these media has an expected lifespan (ZD Net Australia, 2002).

Media	**Lifespan**
Hard disk	3-6 years
Magnetic tape	10-20 years
Optical disk	10-100 years
Magnetic disk	1-5 years
Static memory	50-100 years

Table I. Storage media longevity

From the table above it is easy to conclude that all data must be transferred to another media approximately in 3 to 100 years, at least once.

Information Processing – Threats and Limitations

The rapid and unforeseen development of computing has contributed to solve and support several ways of dealing with information processing. Still, there are several weaknesses in computing. Some of the major problems and issues are summarized by Walker (1994):

- *Specifications*. The specifications for a problem might not be complete and consistent.
- *Problem unsolvability*. Some problems have no solutions.
- *Problem nonfeasibility*. Solutions to some problems might not be delivered in a feasible amount of time.
- *Hardware reliability*. Malfunctions might always occur.
- *Software*. Error-free software is not achievable for complex systems. Errors might be introduced, for instance, in the algorithm design or/and in the maintenance of the software.
- *Human errors*. Can be expected when entering data or/and when operating a computer system.
- *Security*. Data and/or hardware might be altered, corrupted or even destroyed by malicious software such as viruses, trojans, etc.

These threats and limitations will surely remain in the future, since some of them are not connected to the technical aspects, but instead of a social or human character. However, as Walker (1994) points out, not even hardware will ever be one hundred percent reliable.

Accessibility and Availability

The improvements of computers, the development of software, the implementation of different versions of the same software, and the existence of different platforms and operating systems, have caused the creation of different formats. There is format incompatibility in versions and platforms. A large amount of existing digital material, produced just some years ago, is not readable with today's software. Thus, there already exists a lack of accessibility. This problem should be handled by an approach that is aimed at forward-compatibility. By this we mean that new information systems should, yet during their first phases of design, deal with the matter of long-term preservation of the information to be produced by the system.

The paradigm shift between paper based mediums to digital media was not preceded by a strategy for long-term preservation. This phenomenon can be explained in terms of lack of knowledge and experience when handling digital material. Besides, no one could foresee the speed and complexity of the evolution of computers and software.

Digital material that is produced on a certain platform, with a certain application, a certain version and in a certain format, needs to be converted to other platforms, applications and/or formats in order to gain access to the information over time. Another way of retrieving stored information is to emulate the original software (and platform) in which digital material has been produced. In either way, information might be lost and the information retrieved may not be fully reliable.

Digital Processing and Archival Information Systems

To achieve a high level of accessibility and availability to trustworthy digital information in the future, developed forms of digital processing and Archival Information Systems (AIS) are needed. Information in the digital era demands several kinds of tools for its processing. Nevertheless there is a lowest common denominator: digital processing. We begin on the position that we do not want to give a definition of digital processing, we only want to show our perspective of this phenomena, as a mediator for digital information.

Traditionally we understand digital processing as the result of the execution of a computer program in a digital environment. A simplistic model of this process contain an input, a process (that processes the input and generates) an output (Figure 5). We consider that this sequence as a part of an information system.

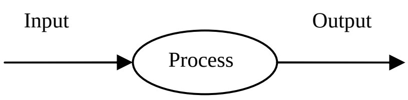

Figure 5: A simplistic model for a computing model

When dealing with long-term preservation of information, it is necessary to define the term AIS. The ISO standard Open Archival Information System (OAIS) Reference Model defines an archive as: “An organization that intends to preserve information for access and use by a Designated Community” (CCSDS 2002). A system built according to this model should contain six high level processes:

1. *Ingest*: Receives the information and prepares it for preservation.
2. *Archival Storage*: storages, maintains and retrieves the preserved information.
3. *Data Management*: coordinates description information from an archival object and system information which is used to support archival operations.
4. *Access*: helps the consumer to identify and retrieve information.

5. *Administration*: plans the ubiquitous archival activities. Monitors the environment and assess the development of new standards and policies.
6. *Preservation Planning*: provides recommendations for conversion, migration, monitoring changes in technology.

In order to develop well-functioning AIS we need useful definitions of what they might be perceived as. A well-known definition is that an information system contains of two linked subsystems. The system that serve and the system which is served, which consists of people taking action (Checkland & Holwell, 1999). This idea is depicted in Figure 6.

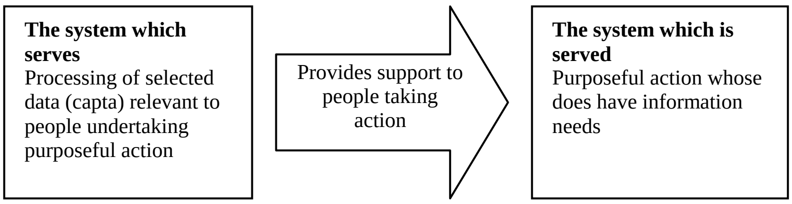

Figure 6: An Information System (Checkland & Holwell, 1999)

Our position is that information is alive when used and useful for the designated community; otherwise the information is just stored and therefore only data. To keep information alive demands a set of ongoing activities, both organisational and technical. When dealing with organisational activities of information preservation Checkland & Holwell (1999), Buckingham et al. (1987) and the broader sense definition of Verrijn-Stuart (1989) are suitable. Concerning the technical activities of information preservation the narrower sense definition of Verrijn-Stuart (1989) and also Langefors' (1993) infological view of information systems are appropriate. Therefore an AIS could be defined as an information processing system that serves an archival organisation which has the responsibility of preserving and making archival information accessible to a designated community i.e. the society itself.

Conclusions

In this paper we started by discussing the information life cycle by providing a view where information is regarded to evolve. That is, information lives longer than people, organisations and tools. Implicit in the Information Life Cycle Model (see Figure 1) is the fact that people, organisations and tools are replaced over time. This result in consequences for the three areas addressed: archival context, embedded attributes of information, and the use of processing tools. The problems in these three areas must continuously be addressed in order to ensure that archiving of digital information is an ongoing activity.

The first issue, the archival context, will be effected, not least will the role of the archivists change due to IT and it could be regarded adequate to educate information managers as well as archivists, if we are to cope with this challenge. IT and its influence on modern archival creation put expectations and demands on the archival science. Archivists need to move from being keepers to act as auditors “if they hope to preserve provenance and protect the evidential accountability of archival electronic records” (Cook, 1997, p. 43). A keeper can handle the records as data – an auditor must deal with the question of securing the information. This leads us to the conclusion that archivists need new awareness, skills, routines and work methods that can deal with long term preservation strategies and management. When this is achieved, the archivists can have a pro-active role and the archives can move from being houses of memory to a more active participation in the information life cycle and thereby providing the society with information. How these skills, routines and methods will be designed is to be investigated in our research.

The second issue, embedded attributes of information, implies that the context of information must be preserved to enrich the understanding of the archived information in the future. Since information can be more than just the words and sentences written down, visual attributes may be necessary to preserve as well. In general, it is the performance of an archival object that is essential, not that its bit stream is intact. The conclusion to be drawn is that both visual content information and context information need to be preserved to ensure future interpretation and understanding of the archived information. This issue is also to be dealt with in our forthcoming research.

The issue on use of processing tools, states that new information systems should ensure the longevity of digitally created information. This means pro-actively set up forward-compatibility. Today information lifetime is prolonged by a set of activities when information passes from its native environment to an AIS. These activities include making the data readable (compatible) for the new environment. Forward-compatibility should be addressed in collaboration with soft ware vendors.

It is inevitable that any kind of information system must rely upon a specific platform and be supplied by one or more suppliers, even an AIS. Platform dependency sometimes forces developers to platform specific technical solutions. This could imply deviations from standards. Supplier dependency implicates not only expertise from a third part organisation but also people involved in the process of long-term preservation such as archivists.

What needs to be addressed in the issue of information processing tools? We believe that an important area to be investigated is the development of new models for the development of AIS which ensure long-term preservation of information. This new models should rely on monitoring, and quality assurance of preservation activities due to the limitations and threats of computing.

Acknowledgements

We would like to thank everbody involved in the LDB Project for valuable discussions and support.

References

Backhaus, H (2005). Arkivteori i internationell belysning. In Ulfsparre, A C, ed. *Arkivvetenskap.* Studentlitteratur, Lund.

Bearman, D (1994). *Electronic Evidence. Strategies for Managing Records in Contemporary Organizations.* Archives & Museum Informatics, Pittsburgh.

Brand, S (1999). Escaping the Digital Dark Age. *Library Journal.* Vol. 124. Issue 2, pp. 46-49. Available at http://library.colstate.edu/libr1105/kramer/cadeau/deadm2.html [2005-04-04]

Bratteteig, T and Verne, G (2000). Feminist, or merely critical? In search of Gender Perspectives in Informatics. In Mörtberg, C, ed. *Where do we go from here? Feminist Challenges of Information Technology.* Division gender and Technology. Luleå University of Technology, Luleå. Pp. 39-56.

Buckingham R., Hirschheim R., Land L. and Tully C. eds. (1987). *Information Systems Education: Recommendations and Implementation*, Cambridge University Press. Cambridge.

Burell, M (2005). Arkiven och informationsteknologin. In Ulfsparre, A C (ed.). *Arkivvetenskap.* Studentlitteratur, Lund.

Checkland, P and Holwell, S (1998). *Information, Systems and Information Systems – making sense of the field*. Wiley & Sons Ltd, Chichester, England.

Cook, T (1997). What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift. *Archivaria 43.* pp. 17-63.

CCSDS (2002). *Reference Model for an Open Archival Information System (OAIS).* Consultative Committee for Space Data Systems. Available at http://www.ccsds.org/CCSDS/documents/650x0b1.pdf [2005-04-04]

Dahlbom, B, ed. (1995). *The Infological Equation. Essays in Honor of Börje Langefors*. Gothenburg Studies in Information Systems, Report 6. Göteborg University, Göteborg.

Dollar, C M (2000) *Authentic Electronic Records: Strategies for long-term access*. Cohasset Associates Inc. USA.

Goldkuhl, G (1995). Information as Action and Communication. In Dahlbom, B, ed. *The Infological Equation. Essays in Honor of Börje Langefors.* Gothenburg Studies in Information Systems, Report 6, pp 63-79. Göteborg University, Göteborg.

Gränström, C (2005). Arkivteori. In Ulfsparre, A C, ed. *Arkivvetenskap.* Studentlitteratur, Lund (In Swedish).

Heslop, H; Davis, S and Wilson, A (2002), *An Approach to the Preservation of Digital Records,* National Archives of Australia, Canberra. Available at: http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf [2004-03-03]

Langefors, B (1993). *Essays on Infology.* Dept. of Information Systems. Göteborg University, Göteborg.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. (1994). *Human Computer Interaction*, Addison Wesley, Harlow.

Rothenberg, J (1999) *Ensuring the Longevity of Digital Information*. Available at: http://www.clir.org/pubs/archives/ensuring.pdf [2005-04-04]

Ruusalepp, R (2005). *Digital preservation in archives: An overview of Current Research and Practices.* Riksarkivet, Sweden. Forthcoming.

Ulfsparre, A C, ed. (2005). *Arkivvetenskap.* Studentlitteratur, Lund.

Verrijn-Stuart, A (1989). Information Systems Concepts: An In-Depth Analysis. In Falkenberg E. and Lindgreen P., eds. *Some reflections on the Namur conference on information system concepts.* IFIP.

Walker, H M (1994). *The Limits of Computing.* Jones and Bartlett Series in Computer Science.

ZD Net Australia. (2002). *Teach Guide: Storage media lifespan*. Available at: http://www.zdnet.com.au/reviews/hardware/storage/0,39023427,20269043,00.htm [2005-04-04]