The value dimensions of Digital Preservation

José Borbinha
IST / INESC-ID – Information Systems Group
Portugal
jlb@ist.utl.pt

ABSTRACT

In this position paper, I intend to motivate the digital preservation community to give one more step ahead, and engage in what I call the new age of value, a third age of a possible process of evolution that already left behind the age of reaction and is closing the age of awareness. The age of reaction was focused on acquiring and preserving data. The age of awareness was focused on preserving information, as data in richer context. The age of value should be focused on producing knowledge as new value, in the sense that knowledge it to make it able to act in the presence of information is knowledge. In case of success the implication of this step would be to move the issue of digital preservation from a source of concern and costs, to a potential source of new value, making it relevant to new audiences, especially in the corporate world.

1. ON POSSIBLE MOTIVATIONS FOR DIGITAL PRESERVATION

The motivations for ones concern on digital preservation can vary, depending of the stakeholders (or the called "communities of interest", or "designated communities" according to OAIS). In a simple analysis those can be classified in views as follows.

1.1 Memory

The most classic view of digital preservation is memory. It means to acquire the ability of access and rendering of information independently of the time, which can be addressed in several areas, such as:

- Private memory - me and my groups (family, friends, tribes, etc.);
- Institutional memory – organizations, public or private (organizational memory), which also might include the motivation for recordkeeping;
- Community memory - public interest (neighboroughs, towns, regions, countries, etc.).

In general, the challenges of this view can be characterized by technical challenges related with the information objects (data formats, etc.).

1.2 Operationally

Another relevant view is operationally, which means the ability to preserve information in order to assure business continuity. This is especially relevant on organizations, private or public.

The challenges of this view can be characterized by technical challenges related with the information objects (data formats, etc.) but also with their context of creation and reuse.

1.3 Strategy

Finally, we can identify possible strategic motivations for digital preservation. This might be related with the need for supporting to processes of decision taking, also eventually related with the concept that the "long-tail" of information also might hide new sources of value and potential knowledge (the importance of the past and of the past experiences).

2. ON THE MATURITY OF THE DIGITAL PRESERVATION

Let us consider the following possible ages for the digital preservation subject (that eventually also can work as a simplified metamodel for a possible capability maturity model):

- Reactive: Preservation of data/file.
- Aware: Preservation of information.
- Value: Preservation of Knowledge.

2.1 Reactive: Preservation of data

We can realize the digital preservation community started addressing the problem in a reactive state of mind (concerns with stable long-term storage technology; standards for image and text formats; etc.).

In this state of evolution we understood digital preservation as a necessary reaction to imminent perceived threats. The lemma was: "For the sake of preservation, I must have to do something proactive and set up a system to collect the information and preserve it; I cannot influence the processes of production of information, but at least I can control the depositing process: in the day I receive the information, I must take immediate action or I might face problems in the future; if possible, I even must impose my deposit rules, so information producers can make it easier for me, independently of their original motivations".

The focus was the preservation of files (image files, office files, etc.) and the lifetime of media (magnetic tapes, CD-ROM, WORM, etc.).

The designated community was mainly the libraries, and the results were the first digital preservation repositories.

Digital preservation was understood here as a cost to fix something potentially wrong, not foreseen as a product or any source of new value at all.

2.2 Aware: Preservation of information

After a moment of reaction, we realized we had to move to the state of awareness. At this stage, we believe that anyone should be caught in fault by lack of digital preservation awareness. To be

competent means to be prepared in advance to face the beast when it showed up.

At this stage digital preservation is perceived as a precaution for potential threats (mitigation to risks), and the lemma was: "For the sake of preservation, I still have to do something proactive and set up a system to collect the information and preserve it; however, in order to be really effective I must influence the producer to assure that I receive the information completed, and not just its data part; I must take prior action before the day I'll receive the deposit, or it might be too late to complement the data I receive with its credible context and make it information."

At this stage we learned that in order to preserve information, we need to put it in a context. Information already is data in a context, but that is the original context meant for the data. For preservation we need to consider one more level of context, which must be now the context of the context of the data. We called it metadata, and then, for example, defined PREMIS for that, and considered bundled it all together according to OAIS, governed by principles as expressed by TRAC, DRAMBORA, etc.

The community of interest enlarged, and archives also become more involved, with their recordkeeping motivations, but they were mainly the public archives (still missing the corporate archives). The vision was of deposits of information, as keepers of objects of public interest, as information resources and also as records of acts and evidences. The results of this stage were the new digital preservation repositories claimed to be OAIS compliant.

At this stage also the eScience community became aware of the issue, with the emerging of the concept of Data Management Plan, which is something we all agree is important, but we are still learning on how really to do it.

Anyway, in this stage digital preservation is still understood as a cost, but with some potential value associated with eventual future needs of data reuse (e.g. in eScience) or proof of acts (e.g. in recordkeeping). Therefore, the result of the digital preservation already might be defined an this level as a "by product".

2.3 Value: Preservation of Knowledge

Finally, I believe we should be in a status ready to drive us to the next level, where we can use digital preservation not only as a reaction to mitigate technological risks or extra measures to assure continuity, but as a new technique contributing to the competitive advantage of the organizations where it is taken in consideration. For that, the concern must be not only to preserve the data and its contexts, but also to be able to align it all with the ability to act on the top of that.

Acquiring and preserving data is fundamental to record facts. Putting data in context produces information. Being able to act in the presence of information is knowledge!

Organizations able to make any necessary use of the information, at any moment can gain an important strategic advantage. But for that they must be able to access and explore not only the data and its original context, but all the contexts where it has been relevant and also all the actions or processes that can be apply to that data. This is strategic thinking.

If we assume the value of this thinking, digital preservation can be an important contribution to the promoting of knowledge! But for that organizations must be able to measure its net value. They must be able to understand digital preservation not only as a cost or a byproduct, but as a capability relevant to produce first order products with positive added value (I mean, the final value will be higher than the cost).

The new stakeholders that might enrich the community of interest are now the corporate archives. But also these have to be put out of their traditional scope of passive entities, and challenged now to be sources of added value.

3. RESEARCH CHALLENGES

In Table 1 we can show a cross analysis of the two higher level views we exposed before.

Table 1. Motivations according to the levels

Levels / Motivations	**Reactive**	**Aware**	**Value**
Memory	Is Focus	Is Focus	Can be Focus
Operationally	Has Potential	Has Potential	Is Focus
Strategy	Low Potential	Has Potential	Is Focus

It is not clear if we already fulfilled the level of aware, but in a looking forward perspective, from, the analysis carried out we can claim the opportunity for research in order to support the transition of the digital preservation, as an area of research and knowledge, to the age of value.

3.1 Why is this topic relevant?

There are no doubts that the corporate world will soon or latter address digital preservation, when business will realize they'll have to do it for compliance (I mean, due to external requirements for recordkeeping), but they also can be motivated to do it proactively, proving it can be shown to be profitable.

However, there is a high risk of having all this done from a perspective and with actors than the actual digital preservation community, which this way might lose an important momentum to expand its area of actuation and the relevance of its body of knowledge (and thus, on the same way, loose the opportunity of getting back new inputs, experiences and knowledge).

Also, this new move will put digital preservation in the general agenda of the "information technology" policies, eventually making it a stronger mainstream issue instead of a niche of concern as it has been until now.

3.2 Why has it not been addressed yet?

Digital preservation has been so far understood as mainly a motivation from the digital cultural sector, with a more recent interest from the eScience community and also some signs of interest the engineering sector (e.g., the airplane industry).

However, the way all this related knowledge has been structured and exploited has been not consolidated and made attractive. The community has been closed in a culture "looking back", considering too much the memory motivations, eventually inspiring the operationally view, but too far from the practices of the corporate world that might be advised to address the strategy views. IT governance is a very complex issue in the corporate world, so we should try to align the digital preservation with the

established practices and techniques at that level, and not expect it the other way around.

Concluding, the digital preservation community has to change its state of mind in order to address the value age, thinking on itself as one more source of potential added value and innovation for general problems, and not only as a source of cost solutions to its own native problems.

3.3 What are the problems involved?

Considering that digital preservation can have a positive return when addressed in the corporate world, the first order problems involved are:

- How can we classify the potential scenarios when and where that can be the case? What would be the main relevant attributes to characterize those classes of scenarios (e.g. central public administration; health care, civil protection; engineering processes in general; supply chains; strategic analysis, etc.)
- What (digital preservation) principles and techniques for analysis do we require in order to really assess the relevance and potential value of these scenarios?
- What knowledge already exists in the actual body of knowledge of the digital preservation community that can be shown as of potential value for the addressing of those scenarios? How can we complement that knowledge with the one already existing in other areas (e.g., supply chain management, risk management, data mining, etc.)
- As opposite of the previous point, what new knowledge do we need to promote in order to reach our new proposed objectives? What of that new knowledge must be responsibility of the digital preservation community, and what of it must be stimulated in other complementary communities?

3.4 What is the potential impact if this is successfully addressed?

As said in the previous point, the problems previously listed are, in my understanding, only first order problems. They can make a first agenda for the digital preservation community, but they don't make yet a relevant research agenda. Such research agenda should result from the findings resulted, in a first place, of the addressing of the problems listed ahead.

Anyway, just doing this might represent a very important refreshing for the digital preservation community, by exposing the actual members to new opportunities, and by attracting new potential stakeholders. It also as an important potential for widening our scope of research and of impact…

4. ACKNOWLEDGMENTS

This work was supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds and by the project TIMBUS, partially funded by the EC under the contract 269940.