Building Digital Preservation Practices, Tools and Services on Quicksand

Bram van der Werf
Open Planets Foundation
The British Library, Boston Spa
Wheterby, West Yorkshire
+31 6 424 06 775
bram@openplanetsfoundation.org

QUICKSAND: THE FUNDING MODEL

The existence of today's digital preservation tools and services within the cultural heritage sector proves there is a need for them and that financial resources can be made available to develop functionalities that address specific digital preservation issues. However, what is really lacking is a solid business case for maintaining these tools over time.

It is best to look at the IT industry to see how the maintenance of products and services works. The life-cycle of a major operating system, for example, is four to six years, with two to three major releases during this period. The releases are part of the maintenance effort and consist of bug fixes and implementation of cost saving and/or innovative improvements. The release and maintenance strategy of operating system vendors has in its turn a strong impact on the cost model for the services and applications that run on top of these systems. These services and applications need to follow the pace of maintenance of the underlying software layer and at the same time they also follow their own bug fixing/improvement cycles. These accumulated maintenance cycles add up exponentially. In general it can be stated that approximately 20% (or less) of the total cost of almost any application or service goes into its development and over 80% (or more) of the total cost is required for its maintenance.

Project-funded software development does not consider maintenance costs. If the aim of a project is solely to produce a software product with specific functionality, and it does not consider the responsibility to maintain this functionality during the life-cycle of the product, then it is very unlikely that this product will be sustainable. Not only maintenance issues are at stacks, even worse, project planning and financial constraints inhibit proper software development. Indeed, in order to deliver software which meets the criteria defined in the project plan, project owners and developers will often be tempted to use approaches that do not really take deployability, long-term maintainability and supportability into consideration. This applies to both the technology and the content back-end which feeds the system with up-to-date information. The content back-end is a crucial source of information (usually a database filled by human intelligence and maintained by human effort) for the application. Typical examples of such applications that are dependent on content back-end are DNS-services and persistent identifier resolver mechanisms. In addition, software technologies are often based on niche and weakly supported technologies, whereas the information backend lacks a competent community to feed it with up to date information.

What we see in the cultural heritage sector happening in the past two decades, is a flourishing R&D activity, based on generous project funding and hardly any serious commitment for deployment, maintenance and sustainability. The sector operates a digital preservation and long-term access business process with tools, prototypes and services that lack any appropriate long-term business planning. The global community of heritage institutions seems unable to secure appropriate structural funding for long-term commitment to digital preservation. Most efforts heavily depend on projects funding and are because of that under the heavy scrutiny of political and financial climates.

THE REAL PROBLEM

Many years of R&D effort have been invested in digital preservation, and more importantly, in developing tools and services to aid long-term access to digital material. In practice, however, those responsible for preservation and long-term access are confronted with the urgency to take ad hoc actions that respond to 1) content production and distribution trends and 2) the needs of users accessing the content. There is still a large gap between R&D and practice. This has been the same for ages, regardless of digital or analogue content. It has always been important to preserve media, whether clay tablets, papyrus, paper or more recently bits and bytes on digital carriers. Understanding context and content (e.g the rosetta stone),in other words keeping the contextual and content metadata, is equally important. The challenge of preserving digital information requires the same tiered conceptual thinking. Bit preservation and media obsolescence remains a risk and though this in itself is not a trivial challenge, understanding context and content of digital objects is a far greater challenge in terms of its risk and mitigation level. Perhaps it helps to articulate long-term access as the real challenge, rather than bit preservation. Bit preservation and longevity of bit quality and integrity requires resources and often implies buying power, whereas long-term access requires a range of reliable and stable information sources and availability of competent and dedicated human resources supported by analysis and decision-making tools, allowing organizations to test, verify and decide on actions for managing accessibility to digital objects over time.

iPRES2011, Nov. 1–4, 2011, Singapore.

DIGITAL AS PANACEA

The majority of society assumes that preserving digital information is similar to preserving paper copies or should at least be cheaper and easier – certainly not more costly and complex. The use of digital media has been marketed to users as utilities offering them higher volumes of information with easier and faster access, but marketing has almost never focused on the associated risks. Where analogue media can be accessed and preserved autonomously, digital objects have a strong dependency on rendering software. The wider community lacks an understanding of technology and does not realize that rendering digital information also requires the right software. Society is also used to the longevity of analogue media (paper, vinyl) and therefore lacks a sense of urgency to take timely action to preserve digital media. Access to information held on CD, DVD, USB sticks etc. is only guaranteed during the commercial life-span of such carriers, which is deceptively short.

For software and hardware vendors, their primary objective is to run a healthy and profitable business. As long as their products generate revenue, they have neither business objective nor commercial interest to guarantee long-term access to digital objects produced for non-supported software versions - unless it generates profit. Where the academic and cultural heritage sectors think in terms of centuries for preservation, technology vendors and format producers often don't even think in terms of decades.

Beyond the cultural heritage and academic sector, awareness of the importance of long-term access to digital objects and the complexity of it, is low to non-existent. Over the past few decades, technology vendors have managed to convince society of "digital" being the panacea for our insatiable need for lots of open, easily accessible and user-friendly information.

Societal naivety and commercial interests are not helpful to convey the necessary sense of urgency and to enforce structural financial commitment for long-term access solutions

NO FREE RIDE

But even in the cultural heritage and academic sector, explaining the consequences of long-term access and taking adequate measures accordingly proves difficult. By contrast, explaining the importance of immediate access and of building information portals is an easy win when needing to convince decision-makers and funders. Open, reliable and user-friendly experiences as we perceive today can be well articulated and marketed as products. As immediate results, they are a good fit for commercial exploitation or fancy demos.

Maintenance services that are necessary to manage these experiences over time are however much harder to market and reactions to such services are more similar to how society reacts to maintenance, insurance and other less tangible, not directly product related activities. When we think about flexible transportation for example, we experience the idea of buying a car as a very positive thing. The price of the car and even the accessories are perceived as fully acceptable, but the pain comes with insurance, maintenance, repairs, taxes, tires etc. It is the big bad world that wants to spoil our joy and only when things go wrong do we appreciate the value of some of these services.

And often during tendering and other types of project or program negotiations, it is on maintenance costs that cost cutting takes place. On an operational level long-term access is a technology challenge with a typical products vs service equation in terms of financial consequences. A 20/80 % cost consideration, 20% for developing a software solution, versus 80% for keeping this solution well maintained and up to date over a period of 3-4 years.

Most of today's information services are being developed with project funding or sponsored by single organizations. Commercial digital preservation solutions actually use information from these services due to the non-existence of business models for sustainable information services with an SLA option and the ability to provide reliable information required for assuring long-term access. This is actually where the term "no free ride" gets put into practice as well. In both the private and public sector, no sustainable service survives without a decent business plan, so therefore tools and services that support digital preservation and long-term access can and will never be a "free ride"

WHAT NEXT?

R&D, innovation and organizational initiatives over the last 12-15 years have produced several long-term access products and services. Many of them have a functional potential to become of significant value and to be mechanical to long-term access. Most of them originated from project funding (Mellon, EC, JISC, NDIIP) or incidental program or project funding coming from individual organizations. Looking back, while appreciating the initial functionality and quality of many of these services, most of them lack long-term sustainability when it comes to maintaining the initial information quality and when it comes to its capacity to adapt to change. Often this is caused by the fact that it takes a relatively big economical effort to monitor, edit and maintain all the potential information resources.

Information coming from these services or architectures is almost trivial compared to the fact that in reality it takes a network of competent, dedicated people, their input and above all a business model with a structural funding, to maintain the quality of the content of such a service over time.

An Open Source approach where a community is actively supported and financed by stakeholders willing to dedicate competent resources is a feasible and realistic option, as business model. But this type of community also needs moderation, leadership and long-term funding to be able to steward, sustain, maintain and manage the solution in the interest of the same stakeholders. While being a great option in theory, practice is more complicated. Sense of community is based on accepting commonalities and the biggest players in ALM sector still tend to foster differences. Another complicating factor is mistrust towards technology service providers while there is a high need for some of their core competences such as engineering skills.

Other business models can be services by subscription, API's with license keys connected to payments, Software as a Service (SaaS) with contracts based on volume or size of organization, or a community model financed by stakeholders and users similar to the DOI business model for persistent identifiers.

Subscription, license, SaaS and API type of business models all assume that core knowledge will reside outside organizations, it is for this very reason that a strong community endorsing and sharing Open Source solutions will bring most value and long term sustainability of solutions to Libraries and Archives. But this has to be treated as a funded business model and not a "free ride".