From Hobbyist to Industrialist. Challenging the DP Community

Peter McKinney
National Library of New Zealand
PO Box 1467
Wellington 6140

+64 (0)4 462 3931

Peter.McKinney@ dia.govt.nz

NDHA Team
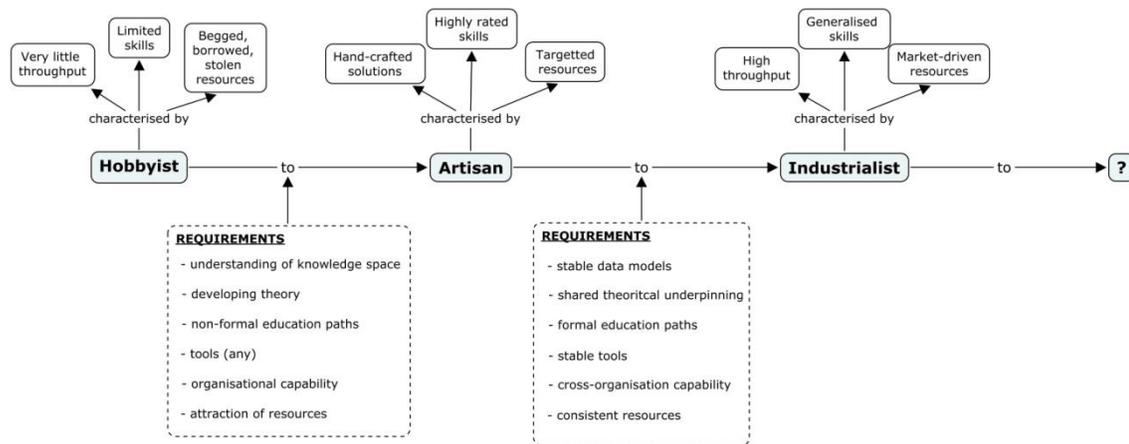National Library of New Zealand
PO Box 1467
Wellington 6140

Figure 1. Digital Preservation Epochs

1. INTRODUCTION

This paper is offered as an introduction to what we would like to present in the Open Research Challenges in Digital Preservation workshop at iPRES 2012.

We discuss what we see as some of the major challenges faced by the digital preservation community. We do so through an epochal framework of digital preservation.

2. Background: Hobbyist to Artisan to Industrialist

Figure 1 offers a view of the epochs of digital preservation. The epochs will not surprise many readers as they are relatively common constructs that describe the stages and abilities of those working in digital preservation.[1]

The stages as defined in Figure 1 are identified by the type of practitioner in the digital preservation field. These are: the hobbyist, the artisan and the industrialist.

The hobbyist is the beginning of the digital preservation journey. While becoming rarer, hobbyists are altruists, driven by an awareness of loss that no-one in their organisation has understood. They are individuals in their organisations, undertaking work that is not within their job descriptions. They have recognized that information is, or has been lost and are trying to do "something", anything, about it for the good of future generations. Generally, they have no official resources and have neither the tools nor skills to be successful.

In next period appears the artisan. Highly skilled, they specialise in hand-crafted work. They are supported by tools that may be off-the-shelf, but more often are bespoke. Within an organisation, artisanal work is recognized as a requirement and supported with a degree of resources. Artisans have acquired their knowledge and skills through non-formal education routes.

The industrialist is encountered next. Digital preservation has now moved onto the "factory floor", with resources culled from offering services rather than fulfilling mandates. These services are well-specified and have massive throughputs. However, this requires that areas requiring hand-crafted solutions are either ignored or destroyed. Specialist skills are lost, but many more people have a base-level of ability, gained through formal education. The move from artisan to industrialist is a painful one. It requires the generification of workflows and practices, which are antithetical to the practices of the artisan.

---

[1] It is hard to pin down an originating source for the framework however. Discussions dating back over ten years with Seamus Ross have certainly shaped our thinking on the framework. Professor Ross utilises the term 'boutique' when describing what we have called the artisanal stage.

The framework in Figure 1 is described in a linear fashion. It is however, far more complex than the diagram suggests. Firstly, we do not know where the end point is. It seems unlikely that industrialisation is the end, so we have left the next stage open for debate.

Secondly, the stages can co-exist. During the real industrial revolution, hobbyists and artisans, despite their declining numbers, were still in place (as indeed all three still exist in this digital age). The Hobbyist is the least desirable epoch to be in: it is the emergent phase that should not be shared by everyone, only the very early adopters. We believe that artisans will always be required, but as the ages progress the percentage of digital content that is managed in this way will diminish.

Which stage is the community at currently? The era of the artisan.

3. The challenges: transition between epochs

For the purpose of this workshop, the challenges are found in the movement between the epochs, the transitioning from one phase to the other.

There are four principal areas of challenge. These are:

- Economic
- Educational
- Procedural
- Cultural

3.1 Economic challenges

Economically, the move between the epochs is a major challenge. We would argue that the battle for the first transition has mostly been won. There are numerous organisations around the world that have teams working on the preservation of digital content. Some have consistent, base-lined resources, others are on less certain grounding. But generally, in areas where long-term value is understood, digital preservation is attracting funding.

The movement to industrialisation requires a large jump in resources, both in terms of size and consistency. This cannot be borne by a few public-body organisations, rather it is likely that resources will come from the sale of services.

There are three critical aspects here: is the market big enough to sustain the industrial phase; where does the venture capital come from; and finally, is the community willing to come under the thrall of market forces?

3.2 Educational challenges

The challenge of education is not a difficult, nor novel one, and is being addressed through various initiatives. The question here though, is what are the courses preparing the students for: boutique preservation or mass-preservation?

In addition, should the students graduate as digital preservation specialists, or rather, should digital preservation be grafted on to other disciplines (as a specialism within IT, records management, information management, librarianship?).

The move to become an industrialist will require a thorough rethink of the curriculum. Their educational needs will be different from the artisans. Is the artisan a step further in the education process or a parallel stage?

3.3 Procedural challenges

Procedural challenges are those encountered in areas such as workflow, tools, standards and knowledge bases. In order to make the jump to industrialised preservation, major changes must be made in all of these areas.

Current workflows are predicated on intimate knowledge of the content being handled. Generally speaking this is because the digital preservation community lacks a shared understanding of preservation management. Examples abound: what puts an object at risk, what is the boundary that moves that risk into the "red" (actionable) zone, what is the best way to mitigate that risk, what are the standards that preservation actions must meet, and what, finally, are the proofs required? There are no community-agreed answers.

The artisan can cope with these preservation management issues because they are hand-crafting solutions for discrete pieces of content. In the industrial age this is not possible. Individualistic (and often emotional) solutions to preservation management cannot be tolerated. Monolithic workflows must be created and used consistently across content types as well as political and professional jurisdictions. These flows will, by necessity, generalise the detailed and complex flows that the artisan has built up. This is no small task and requires:

- enterprise-class tools (at both the system and sub-system level);
- rigorously defined and enforced standards;
- complete knowledge-bases containing proven information;
- shared workflows; and,
- acceptable change embraced as a concept [1] both in terms of skill loss and content loss.

Not one of these items is currently in place.

However, the step towards having the necessary industrial procedures must be taken firstly in the cultural challenges.

3.4 Cultural challenges

Principally, this is the shift required within the DP community itself. We do not believe that people would be offended if we suggested that there is a certain pride in being a digital preservation artisan. We have worked hard to attain to this degree of skill and understanding.

However, industrialisation requires a de-skilling, a generification. What are the steps needed to move the community to this space? Indeed, if industrialisation is achieved, will there be such a thing as a community? Will digital preservation experts have moved on, or become something else?

The community must reassess its own boundaries. Are we a specialists, entire of ourselves, or are we a piece of something else? What is that something else? What are the skills we require and what group does that place us in? Do we have to be system administrators, or specialists in the minutiae of software, hardware and formats? Or are we only the guiding hands behind these technical fields, ensuring that other disciplines take note of the requirements in our field? The community must assess and define before we can move onward.

As has been suggested already, both the artisan and industrialist must co-exist. The nature of this co-existence needs to described and all participants become comfortable with the differences.

In addition, consideration must be given to the shifts that must happen across organisations to allow digital preservation to move forward. This includes shifting decision-maker's predilection of splitting "digital" into discrete packages (with the shiniest winning). We have to change their view to be a holistic one, one that sees the dependencies between all aspects of the digital environment.

4. Wargames

Part of moving forward is understanding exactly where we are. We propose that part of the research challenge is to set up and manage an international 'wargames'. These games will carry out preservation actions on a shared, "massive" set of files.

The participating organisations will decide on a set of content to undertake a preservation action on. The set will be large (10,000+). A scenario will be created which places content at risk making it a candidate for preservation action. The organisations will then use their internal processes and flows to create an end point where the content is no longer at risk. This may mean the migration of the content, or creation of an emulation environment.

The wargames will be a powerful piece of work, generating many positive outcomes. The work will allow participants to assess their own workflows. It will generate meaningful comparisons between organisations and different procedures. The games will expose the many different flows, expectations and assumptions built into preservation programmes. These will include, but will not be limited to; how the organisation identifies content, the data models used, the preservation planning framework, the information required by decision-makers to sign-off actions, the types of actors employed in preservation work.

Essentially, the wargames will:

- Provide a stop point to look back on previous work;
- Be a meaningful, large-scale international collaboration;
- Offer data that can be used to inform:
    - business-as-usual work
    - requirements for future work.

5. Final Words

The key outcomes of looking at these challenges will be an understanding of where the community is currently (we believe it is far from being homogenous) and where the community believes it should be trying to get to. Are we looking to achieve a post-industrialist nirvana (can there be such a thing)? How do we overcome the economic, educational, procedural and cultural challenges? The wargames will help us approach a path of getting solutions to the challenges. At the very least, it will offer the chance for us to embrace true international collaboration.

Keywords

Digital preservation, artisan, hobbyist, industrialist, challenges

6. References

[1] McKinney, Peter. 2010. You did what to it? The concept of acceptable change in a preservation repository. Presentation given at *Future Perfect Digital Continuity Conference, Wellington, 3-5 May 2010*. http://ndha-wiki.natlib.govt.nz/ndha/attach/Presentations/McKinney_Future_Perfect_new.ppt