Co-operative and distributed infrastructures for digital preservation[*]

Michael Day
Digital Curation Centre
UKOLN, University of Bath

http://www.ukoln.ac.uk/preservation/dcc/networks/

The wide range of different stakeholders with interests in the preservation of digital information means that the "most effective and affordable strategy of developing a system of digital archives" would be to "assume a distributed, rather than centralized, structure for collecting digital objects, protecting their integrity over the long term, and retaining them for future use" (Garrett & Waters, 1996, p. 21)

Introduction

It has long been acknowledged, e.g. by the influential Task Force on Archiving of Digital Information, that successful approaches to solving the digital preservation problem will depend to a large extent upon co-operation, i.e. on the development of organisational and social structures that enable the sharing of preservation knowledge, expertise and infrastructures across traditional domain boundaries.

Contexts

The OAIS Reference Model (ISO 14721:2003) saw the potential value of co-operative infrastructures which could help distribute preservation functions amongst two or more partner organisations. This recognised that organisations could co-operate in different ways, e.g. by the adoption of common standards that would facilitate the exchange of objects between preservation services, or through deeper forms of federation whereby the OAIS functions (Figure 1) could be distributed among different partners.
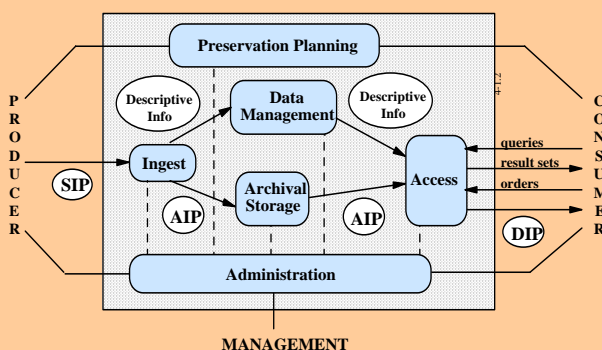

Figure 1: OAIS Functional Entities (ISO 14721:2003)

Furthermore, the nature of the OAIS information model means that there are further benefits in co-operation, e.g. in the development and maintenance of shared services like registries of information about file formats, metadata schemas and other types of Representation Information. A number of such registries are already under development, e.g. the PRONOM registry developed by The National Archives, and the Global Digital Format Registry (GDFR) initiative currently being taken forward by a project funded by the Andrew W. Mellon Foundation. The nature and development of registries of Representation Information is also being investigated in related research being undertaken by the Digital Curation Centre (e.g., Giaretta, *et al.*, 2005).

Within distributed networks, preservation repositories will need technical ways of interacting both with each other and with third party services. One practical example of this might be the modular technical architecture proposed for the US National Digital Information Infrastructure and Preservation Program (NDIIPP). Recognising the undesirability of centralised (or monolithic) approaches, NDIIPP has developed a modular architecture that defined three different functional layers: a lower level for storage, verification and retrieval, a middle layer for management functions like ingest and the management of objects and metadata, and an upper level focused on providing access (NDIIPP, 2003). As Smith (2003) comments, the architecture takes into account the different motivations for repository development, and does not assume that all will need to assume responsibility for long-term preservation. Version 0.2 of the NDIIPP architecture takes into account the need for objects (or collections of objects) to be exchanged between institutions.

The NDIIPP technical architecture "does not envision that every collecting institution would assume the burden of building and maintaining digital preservation repositories; rather, it foresees that a handful of trusted repositories in higher education ... will be certified through some means to assume a national responsibility for preservation" (Smith, 2003, p. 20)

The recent rise of the institutional repository paradigm has also reinforced the need for co-operation on preservation. In this model, it is assumed that many repositories will not be able to undertake long-term preservation activities themselves, so many will rely on third parties, e.g. the preservation services offered by companies like OCLC or by national and research libraries. One example of this is the role of the National Library of the Netherlands within the DARE (Digital Academic Repositories) initiative. Similar issues arise in e-journal contexts where services like Portico and the National Center for Biotechnology Information's PubMed Central are beginning to take responsibility for the long-term preservation of journal articles (Day, 2007).

Organisational challenges and co-operation

Digital preservation can often be viewed as a set of technical problems that need to be solved. However, while developing technical approaches to the problem will remain important (e.g., Day, 2006), organisational problems may in fact be much more difficult to resolve. For example, David has already argued that the social and legal contexts of *cyberinfrastructure* pose more challenges than the the technical ones. He has written (David, 2004, p. 8) that "by comparison with the pace of engineering advances, progress has been slow in constructing social and legal agreements enabling individuals, groups, and organizations to arrive at reliable and transparent agreements for the governance of collaborative work, and especially to do so in a dependably speedy fashion at affordably low transactions costs." These same issues will also have to be resolved in digital preservation contexts.

"The fact that digital preservation is expensive, funding is scarce, and preservation responsibilities are diffused suggest that digital preservation activities would benefit from cooperation. Cooperation can enhance the productive capacity of a limited supply of digital preservation funds, by building shared resources, eliminating redundancies, and exploiting economies of scale" (Lavoie & Dempsey, 2004)

Co-operation on preservation can take place on a wide range of different levels. The strategic level, for example, may often be best dealt with at a national or supra-national level. Successful co-operative initiatives like the UK's Digital Preservation Coalition (DPC) exemplify what can be achieved in raising awareness of digital preservation issues nationally and in fostering joint action to address them. On the more practical level, international initiatives like the Preservation Metadata: Implementation Strategies (PREMIS) working group, the International Internet Preservation Consortium (IIPC) and the InterPARES projects have done much to focus effort on particular problems.

In 1996, the Task Force on Archiving of Digital Information already had indicated some of the other areas where co-operation might be possible.

> "Both informal collaborations (associations and alliances) and formal partnerships among contractors and subcontractors will also surely arise, in which responsibilities for archiving are allocated among various other interests in digital information. Moreover, shared interests in, for example, intellectual discipline, in type of information, in function, such as storage or cataloging, and even interests in the output of information within national boundaries will all form a varied and rich basis for the kinds of formal and informal interactions that lead to the design of particular archival organisations" (Garrett & Waters, 1996, p. 21).

The research described here has a particular interest in the nature of inter-organisational co-operation, but the remainder of this will briefly focus on two key issues: the need for trusted preservation services (and infrastructures) and co-operation on collection development.

Repositories and trust

The concept of trusted repositories is one that has received a lot of attention recently. In 2005, a working group sponsored by RLG and the US National Archives and Records Administration published a draft audit checklist that could be used to evaluate repositories and other preservation services. This list is currently undergoing evaluation, an activity in which the Digital Curation Centre is taking an active role.

Other potentially useful certification frameworks exist. For example, the DINI-Zertifikat has been developed in Germany to support the self-evaluation of institutional repositories. Also, while they have not (as yet) been widely applied to digital services (or repositories), it *may* also be worth investigating the potential of library service quality frameworks like LibQUAL+(TM) in the evaluation of preservation services (Cook, *et al.*, 2001) .

Furthermore, the exact nature of trust in preservation contexts has yet to be fully explored. Clifford Lynch has recently commented that what we are really talking about in this context is competence (Day and Hockx-Yu, 2006). That said, there is a vast literature on trust in the management science literature that may be worth exploring further.

Co-operation on collection development

Collection development is a loosely defined term in library science that is generally used to refer to the selection and acquisition of materials, sometimes also to their relegation and disposal. The concept of organisations - especially libraries - co-operating on collection development has been around a long time. In the United States, for example, there has been an interest in co-operative collection development for research libraries at least since the development of the Farmington Plan in the United States during 1940s. Such co-

operation has not always been successful (e.g., Line, 1997), although more recent attempts to create shared collections in the University of California libraries suggests that it can work in certain kinds of context (Greenstein, 2004). More successful has been the development by libraries and other custodial institutions of co-operative approaches to cataloguing, conservation activities and preservation reformatting. Examples of the last include unified registers of microform and digital masters.

In the digital era, organisations with responsibilities for long-term preservation (not just libraries) will need to consider in detail the benefits of co-operation. The key to this will be maintaining a level of redundancy (e.g., with regard to different technical approaches to preservation or geographical distribution) while avoiding excessive (and unnecessary) duplication of effort. So, for example, it remains to be seen how many national or research libraries will need to preserve electronic journal content, given that Portico, PubMed Central and the National Library of the Netherlands e-Depot are beginning to maintain a growing number of titles (Day, 2007). This is a logical area for co-operation, but may face potential challenges from the problem of 'free riding' (e.g., Day & Hockx-Yu, 2006).

References

Cook, C., *et al.*, 2001, "LibQUAL+: service quality assessment in research libraries." *IFLA Journal*, 27(4), 264-268.

David, P. A., 2004, "Towards a cyberinfrastructure for enhanced scientific collaboration: providing its 'soft' foundations may be the hardest part." Oxford Internet Institute Research Report no. 4. Retrieved from http://www.oii.ox.ac.uk/resources/publications/RR4.pdf

Day, M., 2006, "The long-term preservation of Web content." In *Web archiving*, ed. J. Masanès. Berlin: Springer, pp. 177-199.

Day, M., 2007, "Preserving the outputs of scholarly communication for the long-term." In: *Ejournals Management and Access.* , ed. W. Jones. Binghamton, N.Y.: Haworth Information Press (forthcoming).

Day, M., & Hockx-Yu, H., 2006, "Joint US-UK Digital Preservation Workshop, Washington, D.C., May 7-9, 2006." *International Journal of Digital Curation*, 1(1). Retrieved from http://www.ijdc.net/ijdc/article/view/10/9

Garrett, J., & Waters, D., 1996, *Preserving digital Information: report of the Task Force on Archiving of Digital Information*. Washington, D.C.: Commission on Preservation and Access.

Giaretta, D., *et al.*, 2005, "Representation Information for interoperability now and with the future." In *Proceedings of Local to Global Data Interoperability - Challenges and Technologies* (Sardinia, Italy, June 20-24, 2005), pp. 42-46.

Greenstein, D., 2004, "Library stewardship in a networked age." In: *Access in the future tense.* Washington, D.C.: Council on Library and Information Resources, pp. 9-23.

Lavoie, B., & Dempsey, L., 2004, "Thirteen ways of looking at ... digital preservation." *D-Lib Magazine*, 10(7/8). Retrieved from http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Line, M. B., 1997, "Co-operation: the triumph of hope over experience." *Interlending & Document Supply*, 25(2), 64-72.

National Digital Information Infrastructure and Preservation Program , 2003, *Update to the NDIIPP architecture*. Retrieved from http://www.digitalpreservation.gov/technical/NDIIPP_v02.pdf

Smith, A., 2003, *New-model scholarship: how will it survive?* Washington, D.C.: Council on Library and Information Resources.

> "Library co-operation has always been assumed to be a good thing, but much thinking has focused on the means of co-operation rather than on the ends that co-operation is intended to serve ..." (Line, 1997, p. 64)