

RAND EUROPE RESEARCH AREAS

ARTS & CULTURE
 DEFENCE & SECURITY
 EVALUATION & AUDIT
 HEALTH & HEALTHCARE
 INFORMATION POLICY
 & ECONOMICS
 MODELLING
 POPULATION & AGEING
 SCIENCE & TECHNOLOGY
 SOCIAL JUSTICE

Digital preservation

The uncertain future of saving the past

In recent decades, the core processes of organisations, companies, governments and individuals have become increasingly dependent on digital objects of all kinds—including documents, images, videos, sound files, spreadsheets, datasets (such as patient records), executable programs and integrated combinations thereof. But the pace of technological change and the nature of digital objects make them obsolete and unusable in just a few years. In particular, ‘inherently digital’ artefacts that are not simply page images—and are characterised by complex dynamic, executable and interactive behaviour—are likely to become increasingly important over time. If preservation methods cannot perpetuate their full range of behaviour, the future scholarly record will at best bear only static, snapshot representations of the first generation of inherently digital objects; at worst it may bear no meaningful trace of them at all. At a time when digital objects are being generated everywhere and technology is changing at unprecedented speed, developing an archiving strategy for these objects is like chasing a moving train.

A case in point is academic research. Today’s researchers need to be able to ‘stand on the shoulders of giants’ through access to the work of the existing body of knowledge.¹ Academic progress is marked by articles published in peer-reviewed journals. Tens of thousands of scholarly journals published worldwide distribute an estimated 1.4 million articles a year. Libraries and other institutions have traditionally archived these journals, building massive physical collections that provide researchers with originals or facsimiles (such as microfilm copies) of published work. Paper and photographic page images have so far proved a durable and accurate way of obtaining access to the past.

¹ Isaac Newton’s modest reflection on his achievements is famously expressed in a letter to Robert Hooke in February 1676: ‘Pigmaei gigantium humens impositi plusquam ipsi gigantes vident.’ (If I have seen further it is by standing on the shoulders of giants.)

Abstract

Digital objects are becoming a critical component of scholarly research, but stakeholders show an alarming lack of concern about preserving digital data accurately and sustainably; those charged with archiving information have not yet developed strategies that will enable future generations to build their knowledge on what has gone before. How will research findings be communicated in the future, and how true to the original look, feel and behaviour of these publications will digital archives need to be? Scenario planning can address the challenge of developing strategies in the face of these uncertainties to help librarians and archivists maintain the time-honoured tradition of preserving the past to inform the future.

But it is uncertain whether this traditional preservation approach will be sufficiently robust for the ways in which research results will be circulated, reviewed, accessed and maintained in the future. Archiving and preserving digital objects is fundamentally different from archiving and preserving print objects. Digital objects must be rendered into human-readable form by

It’s not just libraries—examples of digital objects that may be endangered include:

- Spreadsheets
- Patient records
- Meteorological data
- Seismological data
- Multimedia works
- Models and simulations
- Geographical Information Systems data
- Web content

This product is part of the RAND Europe research brief series. RAND research briefs present policy-oriented summaries of an individual peer-reviewed document or a body of published work; current or completed work; and innovative research methodologies.

RAND Europe
 Westbrook Centre
 Milton Road
 Cambridge CB4 1YG
 United Kingdom
 TEL +44.1223.353.329
 FAX +44.1223.358.845

Stijn Hoorens
 hoorens@rand.org

Jeff Rothenberg
 jeff@rand.org

© RAND 2008

[www.rand.org/
randeurope](http://www.rand.org/randeurope)

executable software, and they may become unreadable in just a few years because their formats have become obsolete. Preserving these objects for the long term in a form that ensures future access to their original intellectual content requires a substantial investment in infrastructure, equipment, skills and expertise.

RAND interviewed more than 50 experts and stakeholders from national libraries, universities, governments and major publishers to build a picture of the state of digital preservation today, using scenario planning to explore how they might develop strategies to address the major challenges they face.

Key driving forces

Many industry thinkers regard the archiving and preservation business model—who provides access to whom, in what form, at what time, and who pays—as the biggest challenge for publishers, libraries and other stakeholders. This business model is beset with uncertainties that make the development of archiving strategies extremely difficult. However, behind these uncertainties we find a number of identifiable forces driving the future of digital preservation. Four of these are discussed below.

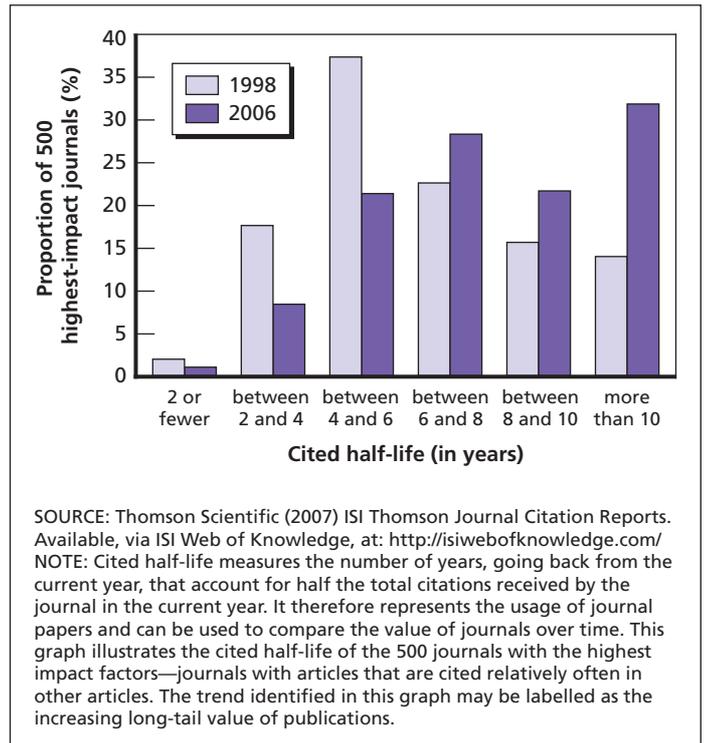
How will research be communicated in the future?

Scholarly communication is fragmenting as it evolves from the traditional model of publication in peer-reviewed journals to wider access through electronic media and information hubs. A number of factors influence this, not least the cost of subscriptions to journals and the fact that paper-based publications cannot support the active nature and pace of new forms of academic research. But while alternative dissemination channels are emerging, existing tenure and promotion mechanisms are still an incentive for scholars to publish papers in traditional peer-reviewed journals. Additionally, published articles are developing an increasingly ‘long tail’ of economic value, spurred by new digital opportunities for repackaging, reselling and improving access. This trend, illustrated in Figure 1, means that academic articles may now be seen as long-term assets rather than as items to be sold once only.

Who preserves what?

As journals are increasingly sold in the form of licences that give access to online content, libraries are no longer in control of their archives. When a journal ceases publication or its publisher goes out of business, its online content may be lost if the long-tail value of the articles is not commercially attractive. Traditionally, national libraries have stepped in to guarantee that access to national research output is guaranteed in perpetuity. However, with the increasing internationalisation of research and the vanishing boundaries of the digital environment, international initiatives will be required.

Figure 1
Cited half-life has increased over the last few years



Who pays for what?

Stakeholders in digital objects have different needs and perspectives, and the optimum funding model for preservation has yet to be determined. One obvious answer might be government funding, but a private approach may be equally—or even more—effective. Archiving and preservation may be regarded as an insurance policy for guaranteed access to digital content, but who will be willing to pay the premium?

What do research libraries and universities demand of archiving and preservation services?

As yet there is not wide concern in the scholarly community about preserving the original behaviour of digital artefacts; the focus is mainly on guaranteeing access to future ‘vernacular’ versions of page-image documents. Most current digital archiving initiatives do not therefore attempt to preserve the full functionality of digital objects. However, that functionality may be a crucial part of a digital object’s intellectual content. Furthermore, historians, the general public, governments and nonprofit institutions may require the preservation of such functionality, whether to ensure intellectual validity, recreate historical perspective, enable aesthetic appreciation, enforce legal and ethical accountability or simply venerate important artefacts. For all of these reasons it may be necessary to retain the original behaviour (as well as the look and feel) of digital objects, particularly those that are inherently digital.

Understanding uncertainty

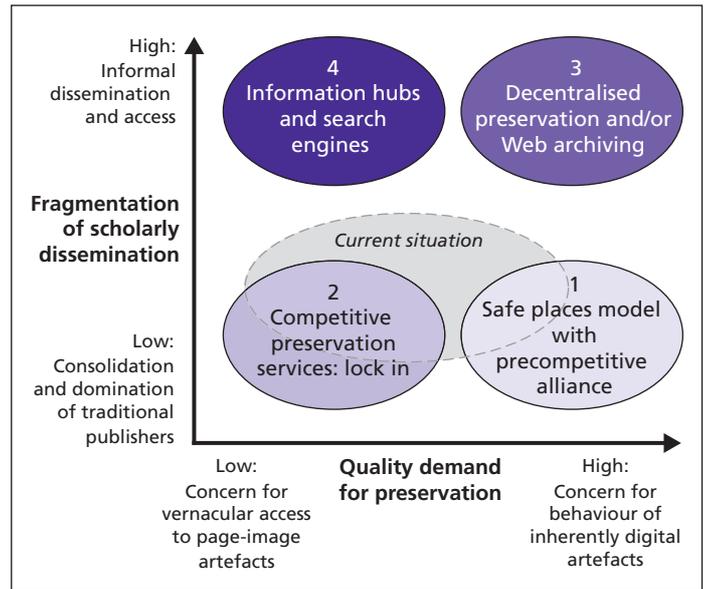
In a changing environment it is very difficult to make firm predictions. Scenario planning may be used to envision potential futures and provide a mechanism for testing strategic assumptions in order to determine their robustness. We have selected two driving forces that are characterised by high uncertainty and have potentially important consequences. First, if traditional publishers further consolidate their dominance of the sector, scholarly communication may become increasingly either open and informal or confined. Second, the lack of concern about long-term preservation of the full functionality of digital objects may prevail or stakeholders may become aware of the importance of such preservation. These drivers span a two-dimensional framework representing the plausible range of futures in this area (see Figure 2).

With an understanding of the drivers that could lead to each extreme, we may build detailed descriptions of the different scenarios and test strategic options against them. Scenarios may also be developed around projected funding and business models, giving strategists a greater depth of information as they look to the future.

Policy implications

Many of the current strategies for digital preservation are evasive and insufficient. We have identified a number of key drivers behind the uncertainties that endanger the preservation of digital objects. Strategies must anticipate these

Figure 2
Scenario framework for the future of digital archiving and preservation



uncertainties and build in resilience to respond to future developments that are unpredictable but not inconceivable. If these strategies fail to be robust under these uncertain circumstances, we may fail to provide access to the 'shoulders of giants' for tomorrow's researchers. ■

Further RAND reading (available on www.rand.org/randeurope/publications):

Hoorens, Stijn, et al., (2007) Addressing the Uncertain Future of Preserving the Past: Towards a Robust Strategy for Digital Archiving and Preservation. RAND Europe, TR-510-KB

Rothenberg, Jeff (1995) Ensuring the Longevity of Digital Documents. *Scientific American* 272(1): 42-7

Rothenberg, Jeff and Tora Bikson (1999) Digital Preservation: Carrying Authentic, Understandable and Usable Digital Records Through Time. RAND Europe. Available at: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/final-report_4.pdf

Rothenberg, Jeff and Jacqueline Slats (2003) Emulation: Context and Current Status. RAND Europe. Available at: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_emulatie_EN.pdf

Rothenberg, Jeff and Stijn Hoorens (2008) Enabling Long-Term Access to Science, Technology and Medical Data Collections. RAND Europe, TR-567, forthcoming

Rothenberg, Jeff, Irma Graafland-Essers, Harry Kranenkamp, Abigail Lierens, Constantijn van Oranje-Nassau and Rob van Schaik (2005) Designing a National Standard for Discovery Metadata: Improving Access to Digital Information in the Dutch Government. RAND Europe, TR-185

Rothenberg, Jeff, Maarten Botterman and Constantijn van Oranje-Nassau (2008) Towards a Dutch Interoperability Framework: Recommendations to the Forum Standaardisatie. RAND Europe, TR-552

RAND Europe is a not-for-profit research organisation providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND Europe's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.

RAND Offices

Santa Monica, CA • Washington, DC • Pittsburgh, PA • Jackson, MS / New Orleans, LA • Cambridge, UK • Doha, QA

RB-9331-RE (2008)



EUROPE

THE ARTS
CHILD POLICY
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND
HOMELAND SECURITY
TRANSPORTATION AND
INFRASTRUCTURE
WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of individual published, peer-reviewed documents or of a body of published work.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Europe](#)

View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).