

Report from 'Trust and E-Journals', London 31st January 2012, Pat Hadley, University of York

1. Introduction

This document reports a DPC briefing day entitled 'Trust and E-Journals' held at the Wellcome Library Conference Centre, London on 31st January, 2012.

Tools and services for digital preservation have been slow to develop and are hard to embed within organisational workflows. Many agencies and sectors report significant gaps in the infrastructure necessary to deliver lasting impact from highly prized and valuable digital resources and those charged with preservation often face complex and highly specialised issues in relative isolation. Arguably, the e-publishing sector is the exception that proves the rule. Perhaps the most advanced part of the digital preservation community, this sector has growing experience in fixing technical challenges and is supported by a well-developed - if complicated and at times dysfunctional - value chain that connects authors, publishers, sellers, purchasers and consumers. A range of service providers and tools now aim to secure this supply chain with digital preservation. Outsourcing - specifically knowing how to trust services that claim to provide digital preservation - has been one of the key barriers to preservation being adopted more widely so the experience of the E-Journal community is of much wider relevance than just the library and academic community.

If the E-Journal market has genuinely solved the 'trust question' then everyone needs to know about it. If it has not, then consideration of the issues will at least enable a more nuanced reflection on how the wider community might want develop trust in the preservation of more esoteric or challenging content types. Therefore, this DPC briefing day examined:

- perceptions and procurement of preservation services for E-journals
- technical architectures for existing services for preservation of E-journals and what they can tell us
- lessons learned, problems solved, experiences to pass on
- trust: how it is established and maintained
- emerging trends for e-journal and e-book preservation

This report serves to assist DPC in measuring the success of its events programme, help shape future work in the field and provide a commentary for those DPC members unable to attend. It includes a narrative report drafted by Pat Hadley, who has also compiled evaluation from participants. The narrative report was produced using an online multi-user service called etherpad which allowed multiple users to contribute to the report: consequently it includes contributions from a range of partners.

2. Narrative report

Attendees from: British Library, JISC Collections, Open University, Charles Beagrie, EDINA, University of Oxford, LSE Library, Tate, CLOCKSS, Royal Society of Chemistry, English Heritage, ALPSP, Wiley-Blackwell, KB (Netherlands), Stanford University Libraries, National Library of Ireland, UK LOCKSS, Wellcome Library, University of Leicester, Archaeology Data Service, Portico, SOAS

1045 Welcome and introductions William Kilbride, DPC

William opening the day with a summary of the event's aims.

What strengths and weaknesses have the e-journal sector got and what have they got to share with others working in the digital preservation sector.

Technology report on this day's topic produced by Neil Beagrie. This is in the packs

1050 The Nature of the Problem: an introduction to e-journals and their preservation, Neil Beagrie, Charles Beagrie Ltd

Issues include trust, continuing access and technological stability
Decade of lessons from the e-journal sector. We will summarise 5 years. The progress made since 2007.

A very big topic to introduce, hence a very dense talk, addressing large amounts of content!

Definitions:

- Continuing access
 - is most commonly associated with e-journal production and
- Long-term preservation
 - Preservation is a simulcra of paper deposition and access in a library that continues in spite of the journal's survival.

Benefits of e-journals:

- convenience (24/7)
- potential space saving
- improved access to larger numbers
- increasing expectations of content formats

Issues

- Different business model
- Concerns over long term access: (eg servers beyond the control of the library)
-

Emerging solutions?

- CLOCKSS
- UK Legal deposit libraries
- KB e-depot

Other services

- LOCKSS
- Portico
-

Note that national library collections, such as created by legal deposit, are subject to draconian access restrictions and thus do not address post-cancellation access. Their usefulness in preserving the record is similarly limited. See for an example from the

KB:<http://blog.dshr.org/2011/01/memento-marketplace-for-archiving.html>

Library concerns

Content coverage - does it cover the interests

Access -

Progress since 2007

E-journal preservation studies

Metes and bounds: A survey of the landscape (Kenney 2006)

Recommendations:

Press publishers to enter into archiving relationships

Share best practice

Join DP initiatives

Recommendations for publishers

Recommendations for e-journal archiving problems

Emerging ISO standard for repositories

Recommendations for JISC

2004-2011 26 recommendations made by Charles Beagrie and others.

Many of these issues will be addressed by other speakers

DPC Tech Watch report "Preservation, Trust and Continuing Access for e-journals"

Scope of the report is to deal with practical and emerging issues for the community involved in the preservation of e-journals.

Issues:

Economic considerations, e-only, print

The fundamental problem is the economic sustainability of e-journal preservation services (see the Blue Ribbon Task Force report <http://brtf.sdsc.edu/>), Related to this is the fact that e-journal preservation, like many technology markets, has increasing returns to scale (see W. Brian Arthur "Increasing Returns and Path Dependence in the Economy") also called "network effects". Thus economic sustainability is possible for at most a very small number of solutions, if for any at all. For details see Brian Arthur's book, or for a simple explanation see: <http://blog.dshr.org/2009/04/spring-cni-plenary-remix.html>

Standards

PDF, PDF/A

XML

DOI

As regards the NLM DTD, note that it is being extended, formally standardized, and renamed JATS. See: <http://www.niso.org/workrooms/journalmarkup>

Any discussion of access to preserved e-journal content should acknowledge that the Web now has a standard for access to preserved content, called Memento. For details, and some discussion of looming problems, see: <http://www.mementoweb.org/>, <http://blog.dshr.org/2010/12/importance-of-discovery-in-memento.html>, <http://blog.dshr.org/2011/01/memento-marketplace-for->

[archiving.html](#)

The Web is evolving from a document model to a programming environment model, from static to dynamic content. Preserving dynamic content is much more difficult than preserving static content, yet the more the intellectual content of research communication is expressed in dynamic, executable form (as for example in workflows or interactive media), the less useful preserving a "frozen" form (such as PDF/a) becomes to future readers. The standards of interest here include AJAX and HTML5. For an explanation, see: <http://blog.dshr.org/2011/08/moonalice-plays-palo-alto.html>

- OpenURL and KBART - see:

http://www.lockss.org/locksswiki/files/Link_Resolver_Integration_White_Paper.pdf

- Semantic Web Publishing standards (See

<http://www.dlib.org/dlib/january12/shotton/01shotton.html>). Note in particular the importance of explicit licenses in this context.

- Standards for publishing supplementary material, see the David Shotton article and

<http://dx.doi.org/10.3789/isqv22n3.2010.04>

Use cases

Cancellation of subscription

Journal discontinued

Note that Post-cancellation access is a problem only for subscription e-journals that do not use a moving wall and are not subject to deposit mandates such as PubMed Central.

Note that, from the future reader's perspective, there are only two use cases, since the last three are equivalent:

- The content is available from the publisher but the reader is not authorized to access it (post-cancellation access).

- The content is not available from the publisher (preserving the record). This applies, for example, to the case where a publisher abandons a journals (examples available at http://www.clockss.org/clockss/Triggered_Content - note that they are under Creative Commons licenses) as well as to publishers ceasing operation or catastrophic failure.

Note that the implementation of access is different in the two cases. In the first, the reader's finding aids (bookmarks, links, DOIs, search, etc) will direct them to the publisher, who will refuse access and may (or may not) redirect to a separate provider of post-cancellation access. In the second, the reader's finding aids must know to where redirection should point (see, for example, multiple resolution of the DOIs of abandoned journals by CrossRef), none of which will be the publisher. Thus the distinction between the last three use cases is mainly about the interaction between the service(s) and the finding aids, and how seamless this is from the reader's perspective.

Trust

A Brief History of E-Journal Preservation at:

<http://blog.dshr.org/2011/08/brief-history-of-e-journal-preservation.html>

Question about repositories:

Peter B: Archiving of author's final copy of an article vs publisher's final copy. Which has the preservation priority?

Granularity is an issues, some repositories (eg, University repositories) keep a collection by article

rather than whole journals.

Vicki: The notion of an academic article is changing and including data inclusion.

David Rosenthal: The notion of the article is being rebelled against. Scientists are publishing workflows rather than results. Emphasis on publishing active elements of research (eg 3D). PloS Currents brings scientific publication closer to blogging and it makes it much more challenging to preserve.

William K: Trust in services are important rather than 'repositories.'.

1120 Licensing E-Journal Content , Liam Earney (JISC Collections)

A view from JISC Collections

Areas of interest:

Preservation, post-cancellation access, perpetual licence agreements
These issues overlap and are complicated problems for JISC

Rights are complicated! Perpetual licencing is sometimes complicated by finding rights-holders or issues with revocation after the fact.

JISC has a 'dark archive' of backed-up content that duplicates other archives or publisher's own material.

Licensing activity

- Changes over time

Nesli2 (<http://www.jisc-collections.ac.uk/nesli2/>)

Difficult principle to follow up on. Transfer of publisher often has practical issues that are different to cover in a licence.

JISC were passing information outwards but began asking questions of institutions and finding that there were things missing from the licence. Can we align the needs of end-users with publisher's business models and their technical capacity.

A bottom up approach has helped facilitate a stronger licence

- The Nesli2 licence today

Covers:

Preservation - and where?

Post-cancellation access - and how? to what?

Journal transfer - what arrangements are in place in the event of transfer to another publisher

Important not only to make sure that the licence is effective and watertight but is understandable to the institutions concerned

Entitlement Registry Scoping Study

Issues and challenges:

Availability of the data

- incomplete publication information
- Ability to output subscription information

Publisher's struggle to know what their catalogue consists of in a consistent way

Workload

Who does the work? How is it recorded and disseminated?

Granularity?

Is it enough to know what was in a deal in a given year? Rather than try and map the entire collection to which an institution had access

Attitudes:

This is not a core issue for all institutions

'Not an issue'

Many institutions will deal with problems as they occur

Discrepancies between what institutions have practical access too and what publisher's say they have access to.

Implications of cancellation as a part of 'big deals' is a major issue.

Questions:

When institutions lose access have there been problems with publisher's attitudes?

No, mostly publishers have been apologetic and fixed things.

David R: In the paper world this was the process of "claiming the serials", and it is a major function of the LOCKSS model of e-journal preservation - i.e. the library actually taking custody of the content that the library paid for.

William K: When material has gone missing where is the driver for recovering it?

David R: When things go wrong at Stanford (crawler issues) the journals turn off access for the whole of Stanford normally the researchers are the first to notice (particularly with current content) and the important thing is to build in machinery to keep checking that material is accessible constantly.

Kieth M: What happened in the paper world when copies were lost? There were licencing agreements to cover replacements and check systems were still human based. Electronic systems enable us to know where the problems are sooner and see how big they are.

Randy: Readers Authors are often first to pickup on gaps in electronic access.

Alena: The move to electronic has decimated the teams that used to check paper journal collections. In an electronic world there is a role for a central body (like JISC) to generate machine solutions to these checking issues.

Alan R: EDINA looked at tools for constructing registries to access and catalogue these through metadata.

1140 Service Providers' Forum,

William K: The workflow from the e-journal world has got examples of services that provide a variety of solutions to DP. The e-journal market has 'cracked' the problem of getting outsourcing to work.

Portico (Kate Wittenberg),

Used to run an e-publication group at Columbia and looked at born-digital collections management.

Requirements for Trust in DP?

- Clear preservation definition and mission

Needs to communicate what DP means in their services. The governance and operation needs to back this up - preservation is a priority not secondary to access.

- Transparency and auditing

Archives must be transparent about what has been preserved and how. Status must always be visible. Machine tools must facilitate this and it must be done in consultation with stakeholders. This must extend to openness to auditing by third parties and have this presented to stakeholders. Does the community really value these audits?

- Reliable delivery of service over time and effective response to events requiring a preservation solution

Say what your aims are and follow up! This means responding to content in jeopardy and being able to demonstrate speedy effectiveness in a crisis.

- Commitment to and capacity for research into future needs of the community

Demonstrate effective and sympathetic solutions to new sorts of content and new ways in which research is done.

This needs to take place through collaboration with research institutions

Adam Rusbridge (UK LOCKSS Alliance),

Goal of LOCKSS: help libraries build local archives of web-published content

Trust in their own capacity to respond

Lots Of Copies Keeps Stuff Safe!

Distributed solution to DP: Sustainable access to scholarly work

LOCKSS uses web archiving technology and prevents accidental or intentional damage to archives and single-point failure

LOCKSS integrates into OPAC library catalogues, and local ownership of content puts access in the control of the library. This access will reinforce institutional support for the approach (enhancing trust)

Benefits include the retention of publisher branding and allows publishers to access use stats.

2008: 300 publishers

2010: 400 publishers

2012: 500 publishers and e-books

Collaborative infrastructure enables openness to stakeholder needs.

LOCKSS software enables the coordination of security, search and access.

Randy Kieffer (CLOCKSS),

C in front of LOCKSS = Control

CLOCKSS is a 'dark archive' founded by libraries and publishers

Community governed: Board - 50% libraries and 50% publishers

12 CLOCKSS boxes around the world - protect from geographical/political issues.

Maintain modest costs and keep resource needs low

Open Source software

Triggered content is made available to the whole community

Allows abandoned/orphaned content to be kept by the whole community

Board is spread globally and libraries are those that have CLOCKSS boxes

Publisher's comments:

Impressive capture of the full web-published content. Rich content and dynamic elements
eg, old branding will be maintained!

For OA publishers the content will have an OA future.
These born-digital publishers are the most at risk

The big guns are there! Solid list of publishers and important list of libraries that get listened to.

Commitment to diligence on owner and access licences before triggering
The biggest challenge is technology. The second is always rights and ensuring publisher's business is protected if necessary.
Board always requires 75% quora on any issues including triggering.

Marcel Ras (eDepot/ KB)

All the issues have been set out! Why archive and various methods.

The Koninklijke Bibliotheek is the national library of the Netherlands

Deposit library but not legislated: This works well for print but not for digital.

Will merge with the national archives in 2013

System is OAIS and integrated with the catalogue and other library modules

Also archiving e-books, digitised materials and websites.

The KB is building a 2nd generation DP system
International e-depot as a European service

Stakeholders:

Researchers, Libraries and Publishers

Co-operation with

Service providers

Registries

Safe places network

Next steps for eDepot

Develop a business model

Implement the infrastructure

Engage with partners

Continue research

1240 Discussion

Pat H: How do we facilitate machine readable archives that enable datamining, topic modelling etc?

Kate W: This is similar to enabling new data formats internally

David R: licenses for e.g. linked open data essential but mostly missing. Publishers want to make money from data-mining. Format matters - see the work of Peter Murray Rust.

Vicki R: CC helps!

Marcel R: We need to preserve in anticipation of new archival methods but preserve FIRST

Alena P: Publishers look down on datamining but that is what researchers want. This presents a l

Peter B: Author's final copy is important to this query but is often less consistently machine-readable

William K: Designated community needs to be able to recycle research

Randy: Preservation and datamining are different businesses. Money is there to be had enabling datamining in major publication archives. Publishers will want to protect their commercial interests

Peter B: Author's rights must be protected too!

Keith M: What is dynamic content?

Kate: Material that is not a static argument - a GIS a 3D model or interactive content. It is no longer confined to the hard sciences. It is content that facilitates new kinds of questions

Keith M: Migration issues quickly kick in

Vicki: Video games are the canonical example. We cannot preserve someones experience of a game. Material can be executable (software) rather than static.

Neil B: Scope for the Tech Watch report should be 3-5 years. How quickly are things changing? Kate's concept of their being an R+D element to a service's offer.

William K: Staff training and knowledge are one component of keeping up to date.

Audit and certification in Kate's presentation: What kinds of audits are appropriate? What is the methodology? Who should do it? Who decides what needs to be asked?

Peter B: Some of this will come up after lunch. This gets to the heart of the trust issue.

Randy: We need to look at two aspects: How good is the preservation technology? How sustainable is the organisation as a body - accounts etc. ?

? Role of copyright libraries - no legislation for statutory deposit of e-content. This needs to be borne in mind in the Tech Watch as voluntary deposit doesn't work as well for digital as for paper.(e.g. BL has a voluntary model but low take up)

BL Andy: Voluntary deposit of e-content from journals is taking place

William K: Draft regulations are in the pipeline but taking a long time

David R: Generic problem with copyright deposit is the restrictions on access. Physical access is needed and makes it less useful as anything more than a dark archive.

Peter B: National libraries should have an ongoing commitment to being last resort if necessary.

Research/university libraries have a different emphasis.

William K: What is the best methodology? How can we decide what to judge our preservation solutions against? Topics to think about over lunch: Who is the appropriate body? Is tot the National Library?

1300 Lunch

Busy! Lots of lively discussion and meetings of archaeologists! We're everywhere!

1400 The Keepers Registry: Enabling Trust in e-Journal Preservation, Peter Burnhill (Edina)

Trust, e-journals and preservation

Who does things right? New Zealand

Trust takes place between agencies, information must also be trusted to be authentic and reliable.

Life cycle models

These illustrate how material is created, curated, made accessible and preserved

In whom or what should we put our trust?

We need more than faith in systems for 'eternal' preservation

Content is definitively different in the digital realm: Content can be copied - there is no original?

Trust implies risk

We need to trust people and procedures - we need archival fire drills

Systems need to be designed to cover significant loss.

We need to trust in:

Methodology

Competence

Commitment

Disclosure of what is done and audit of what is said and done

The Keeper's Registry is about disclosure of archival action

Aims to be a global online facility that:

Reports on who is looking after what ejournal

A showcase for archival intent

EDINA and ISSN became partners in the JISC funded PEPRS project

Piloting and E-journal Preservation Registry Service

Six agencies: BL, CLOCKSS, e-Depot, LOCKSS and Portico

What it looks like: <http://thekeepers.org/thekeepers/keepers.asp>

Soon to add National Science Library of China and subsequently Canada, UK and USA

Eg Royal Soc Chem. 74 Journals being archived and keepers can be seen and ID'd

Currently: 20,024 titles and 16,558 preserved.

This must be international development - a global facility.

How to assist DPC members with other works?

What lessons apply to other areas of DP?

Self-disclosure and enables stakeholders to see what needs to be done.

Abstract data model requires high quality metadata

In the e-journal world there was a pre-existing lessons:

Taxonomy (ISSNs), many copies (standard for print), this record must be integrated to be effective.

It is difficult to get below the serial level to the volume level (let alone the article level) - Granularity

The key thing is copies! This is easier with the digital than it was with paper.

Stewardship is the action that engenders trust.

Digital accidents can lead to those not used to acting as archives becoming so (eg publishers due to libraries not keeping backups of e-journals like their paper copies)

Ease of access is the ultimate aim.

Trust is about avoiding disaster by sharing the task

1 Methodology of copies and enable different approaches

2 Competence to do what is required

3 Commitment to do it in terms of sustainable funding

4 Disclosure of what is being done

This is the role of the 'Registry of Keepers'

5 Audit to ensure the above is done!

Questions:

William K: ISSN is a useful tool it may have parallels?

Peter B: Yes, it is problematic that it doesn't go below the serial level. Others may use similar taxonomies to register their material with another 'back-up' agency

Keith M: Can we map ISSNs to URIs? Do we need to map down this to individual articles?

Peter B: Yes we can do either but for registry purposes mapping ISSNs to URIs is sufficient. There are too many articles!

1430 Publishers' Perspectives, Fiona Murphy (Wiley)

Climate-gate led to issues for e-publishing. The politics of datasharing, research-cycle and authorship, version control.

If best practice had been followed the crisis may have been averted.

Wiley Blackwell:
Scientific, Technical, Medical, Scholarly Content
1500+ Journals
800+ Society Partners
4 million articles 10k online books

Wiley-Blackwell engages with Portico and CLOCKSS
This involves a significant investment.
This includes deposition of supplementary information

Is it worth it?
Wiley examines issues with their long-term goals
DP with these bodies demonstrates WB's commitment to their own customers: Researchers

E-science requires a DP solution to innovative research methods and new dissemination technologies

Data
Textured articles
Cross-linking
Funder mandating
Long-term curation

Case study
Geoscience Data Journal

British Atmospheric Data Centre and The Royal Meteorological Society.

This new journal will link articles to repositored (DOI'd) datasets on servers not run by Wiley.
Making these links perpetual is crucial.

The model for this was the ICSU World Data System

These have four levels of membership:
Regular - Deal directly with data curation and analysis services
Network - Groups of regular members
Partner

Questions:
William K: This is a bigger challenge than anyone can take on alone. I like to hear that Wiley thinks that too! You seem so big!

Fiona: Wiley has many partners and they are needed for many of our projects

Jonathan Tedds (Leicester)? The recent RIN report

<http://www.jisc.ac.uk/news/stories/2011/09/~media/Data%20Centres-Updated.ashx> highlights that Coordinating with datacentres is very important and researchers value the expertise of those running these services.

Neil B: DRYAD <http://datadryad.org/> is an excellent example of data centres engaging with small publishers for supplementary data.

Peter B: CLOCKSS led me to think of things from a publisher's point of view. Though the big publishers have lots of resources putting something visibly with another archive engenders trust.

Fiona: The more one learns about DP more one realises its importance!

Judith W: At Internet Archaeology linking articles with data archives has strengthened DP and scholarship

Fiona:

Neil G: Who has the role of engaging with other bodies working with DP? Is it a PR role or is there a dedicated DP at Wiley gap?

Fiona: There's no-one engaged specifically at the moment but there are many of us that engage with it in various ways, particularly the technical people.

1500 Publishers' Perspectives, Richard Kidd (Royal Society of Chemistry)

Informatics manager at RSC

A view from ALPSP

Access is perpetual

RSC signs up to Portico, LOCKS and CLOCKSS

Since 2000

SGML to XML

XML to XML v2

XMLv2 from DTD to schema

and 2 platform upgrades!

Data must be reused in order to be regularly checked and upgraded.

What not to do?

Supplementary information moved from Word to PDF

Stable but not usable

Protect the PDFs to keep the 'master copy' but this makes them impossible to use!

Personal perspective

Services are about reassurance and insurance

how likely is catastrophe?

betting on who stays in business

publishers tend to keep stuff available

People trust what they can understand

All parties understand the costs and what they're buying

The easy stuff is done

Evolution to linked data and services will be more challenging

Semantic enrichment is difficult and brings more problems

Database rot

In biology particularly academics retirement and funding running out causes loss of huge amounts of data.

It is easy for data to disappear without people realising.
For example datasets can end up co-referring to one another and not to a definitive source

Internet Journal of Chemistry

Disappeared and the originators are making it impossible to resurrect or enable the back catalogue to reappear

Questions:

Peter B: New trends are toward the publication of rich datasets that are dynamic. The publication world is dependant on fixity and permanence. How do we steward constantly updated scholarship?

Richard K: Malleability of publication is incredibly difficult to plan for. It's changing very rapidly

William K: Memento style web-archiving technologies may be more appropriate.

Peter B: But the problem is provenance and attribution. Fixity is easy to read!

David R: The separation between journals and the web is gradually going to disappear. Blogs and wikipedia both have mechanisms for preserving history. Because production is more weblike we need new solutions. We've done the easy bit.

Vicki: RSC is heavily AJAX and more innovative at bringing data to the end-user in a useful way.

Keith M: Does better interlinking lead to higher readership and returns on e-publications? Are there any metrics?

David R: Yes if links are not behind a paywall. Making e-journals open to google searches has massively improved readership

Keith M: But does making better linking lead to better revenue?

Richard K: Very hard to eliminate all the variables?

Jon T: Linked data has found to increase citations independent of article cost.

see <http://researchremix.org> Heather Piwowar

Neil B: Time dimension is important to our risk management. Like compound interest, Low risks over very long periods compounded = high risk and are not acceptable.

William K: Article half-lives in various fields demonstrate the lengths of time we need to be aware and anticipate

David R: Some resources don't follow this "half-life" model with a gradual tail off. They are "Sleeping Beauties" which suddenly wake up

1530 Coffee

1600 Panel session and discussion, led by Neil Grindley, JISC

Trust underpins a lot of the discussion though is seldom mentioned explicitly. Makes is a useful topic and one which allows us to look afresh at some familiar issues.

Neil G: Frustration that some aspects of e-journal DP isn't moving as fast as it should. Trust is definitely a contributor to this frustration.

Definitions of DP, access and other key terms are important and difference between Preservation and Access is not necessarily clear beyond the sector.

Randy: Many librarians are confused about untangling post-cancellation access and long-term access because they tend to view the issue only from their own perspective..

As CLOCKSS is a dark archive they do not do post-cancellation access. LOCKSS has 14 years of helping with that.

Kate: Seconding Randy's point that libraries are sometimes conflating the issues but are unsure when they are participating in long-term preservation and when they are just ensuring post-

cancellation access

Lisa C: Librarians want everything!

Kate: Publishers can see long-term preservation strategies as competing with their offer of digital access.

Peter B: Smaller Universities have described their missions as 'access libraries' rather than 'holding libraries'. Some bigger libraries now behave in this way and ignore any stewardship responsibility that they have.

Post-cancellation access is ultimately the libraries responsibility to their users.

David R: It's easier to maintain access to subscription content because it has a business case and it therefore attracts more funding. . But with OA or small journals it is far more challenging to preserve material and this material is at far more risk. There are no easy solutions for this project.

Work like KEEPERS helps with the allocation of resources to content and how it relates to the risk of loss of access to that content. Note that customer demand for post-cancellation access means that the vast majority of e-journal preservation resources are devoted to preserving subscription content. Because subscription content has a presumptively viable business model, it is inherently at lower risk than open access content, whose business model is tenuous. It is difficult to fund preservation of open access content, because doing so addresses only preserving the record, whereas preservation of subscription content addresses both functions.

Further, even within these two classes of content, resources are preferentially devoted to preserving the content of larger publishers, whose content is at lower risk of loss. From practical experience in the LOCKSS program, the cost per byte or cost per article of preserving content from a large, technically sophisticated publisher such as Elsevier is much lower than the cost of preserving content from the typical smaller publisher. Thus we can say that the vast majority of e-journal preservation resources go into preserving content that is not merely at low risk of being lost from the record (since Elsevier is not going away) but also at low risk of having its subscription canceled (since Elsevier is the last subscription that would be canceled, and post-cancellation access would be from Elsevier's platform anyway).

On the other hand, the vast majority of important open access content is being preserved, if at all, only by the Internet Archive.

For the proportion of open access publication see <http://oacs.shh.fi/publications/elpub-2008.pdf> - for some insight on the growth of open access publication see <http://blog.dshr.org/2011/10/plos-is-not-as-lucrative-as-elsevier.html>

Neil B: It may be a painting the 'Forth Bridge' problem. Education of the newest generations of librarians and archivists is important.

William K: DPC is becoming more specialist due to increasing sophistication and the pace of change. What message should we be taking to students that need to learn quickly, the basics of DP best practice.

Neil G: People need to know the achievements and milestones that have been reached. We are not simply reinventing the wheel. There are lots of lessons that have been learned in the sector.

Marcel R: Modular approaches are now far more robust, extensible and sophisticated than 10 years ago. They can link articles, e-books and data sets based on persistent identifiers. These are still challenges but we can transfer our lessons into new infrastructure

Peter B: In some cases its impossible to begin with user requirements. Vaporware and reports are not sufficient. One might have to build something and then generate feedback.

Marcel R: User requirements used to be very loose and now we understand them in more detail.

Kate: Maybe DPC has a potential role as an educator for those engaged on the fringes that need to learn quickly through webinars or workshops.

Randy: There is a great deal of commonality that we can communicate, vocabulary and expertise can be shared without being tied to sales pitches. It's important to recognise that some practitioners have better facilities or skills for certain jobs.

Adam: There is definitely a role for education. Some users treat LOCKSS as a short term back-up system and believe their material will be deleted if they unsubscribe. Perhaps there's a risk model that needs applying to various areas so that we can target the best areas to put resources.

Vicki: It's ironic that the DP community are spending our limited resources preserving the content of wealthy publishers. which is at relatively low risk

Fiona: Wiley and other publishers need to engage with supplementary information and underlying data. The remit for what needs preserving is difficult to define. Publishers get their money from libraries so the pot is smaller for everyone. With the resource shortages it is worth thinking about efficiency, economies of scale and putting the cash into that which is most at risk.

Peter B: Our digital book shelves are growing and are coming from born-digital and digitised legacy journals. e.g. recent addition of about 8K titles from 250,000 digitised journals held by HATII Trust <http://www.hatii.arts.gla.ac.uk/research/publications.html> but it is difficult to engage with other producers of content that do not have the same business model as e-journal publishers. Priorities include creating an alert system for what is at risk. Crowd source an "At Risk" Register? Archiving agencies need to be able to respond to this. Often only small proportions of a title's output is digitised and back catalogues are incomplete and poorly protected.

Neil G: Back to trust!

How do we deal with innovative methodologies (crowdsourcing archiving!!) ??

If journals look like blogs and we have bigger datasets are we solving a dead problem?

William K: Digital material grows on 3 axes:

- Scale trillions of files
- Complexity
- Expectation

We need workflows that can cope with trillions of diverse files and understand how they are created. Trust becomes more important as an individual's interaction with these huge ranges of data is mediated by many of the agencies involved (libraries, publishers etc)

Richard K: It is currently impossible to cover the diversity of where scientific communication might go. We may have to just identify and protect the lowest common denominator. The next 10 years is impossible to predict. We cannot even scope systems to mitigate the preservation of material we are unable to anticipate.

David R: We do know some of what will stick: Web standards have a 10 year life cycle. We can understand some of the fundamentals if not the appearances of the final products. It will be difficult but should not be impossible and work has begun with those looking at scientific workflows and elsewhere. We need to invest in looking at infrastructure. The copyright law for the static stuff is well understood, but the licensing of the executable stuff is much less clear, making the legal aspects more challenging than some of the technical aspects. Only 20% of the Linked-open-data online can be attributed to a set licence.

Fiona: Publishers do not have sufficient understanding of these new formats. They are used to working at very macro (top-down) levels and publishers need to learn more about the nitty-gritty of strategies and methodologies for new research and dissemination.

Peter B: It seems that the learned societies have not kept up with Open Access in the last 10 years but they potentially have a niche to help researchers, we need to raise our expectations of them to engage with new approaches to scholarly communication. DP archivists and publishers engage with one another.

Richard K: Scientific research is about preserving what you have completed and funding should stipulate this.

Kate: It is a challenge to do R+D while running an effective operational service in business. Need to find partners (publishers, libraries, archives) we should have these run in parallel.

Keith M: What percentage of a project budget is needed to ensure decent DP? In archaeology it seems 2-3% This is good news insofar as it is not 20%! That means the costs of archiving are not necessarily so substantial as to completely inhibit the inclusion of preservation costs in research grants.

Neil B: Data centres shows similar figures for preservation

Randy: Learning societies can be unaware of their role. The economic crisis should be seen as a wake up call: Some have most of their revenue come from publishing but think they exist to run meetings. It is not wise to target publisher's at risk as this might open you to litigation. There is never going to be one group that can capture everything! We need to spread ourselves sensibly.

Peter B: We only represent a portion of the sector.

Adam: Helping titles move between publishers or have contact points or methodologies for protection.

Peter B: Publishers want to reassure their customers by demonstrating that their material is protected by 3rd parties.

Neil G: Who engenders the least trust? How can we tackle this?

Peter B: We must not get hung up on auditing.

William K: We must not get stuck reinventing the wheel because we don't trust others in the sector. We need to engage positive feedback loops that help with scale problems and shared complex issues.

Adam: We need to trust services (active) rather than repositories (passive)

William K: OAIS repository certification tools which check a small part can lead to unwarranted assumptions about the health of the whole operation. demonstrates that all the different parts of the machine need to work together and build upon one another's efforts.

Keith M: Access still hangs on terminology and taxonomies. Both in terms of researcher's terms and those used by DP sector.

Peter B: Publshers need to declare what their exit strategy is. Do the National Libraries need to act as a safety net for crises that we cannot anticipate?

1650 Wrapup, William Kilbride, DPC

Thanks to speakers, biscuits and the Wellcome trust!

Reminder for feedback for the Tech Watch and the day's feedback forms.

Also worth noting:

<http://f1000research.com/2012/01/30/f1000-research-join-us-and-shape-the-future-of-scholarly-communication-2/>

http://www.frontiersin.org/Computational_Neuroscience/10.3389/fncom.2011.00055/abstract

3. Compiled feedback

Attendance:

37 people booked for the session and 6 completed an evaluation form. There was 1 no show on the day.

Type of organisation:

Library	Archive	Museum	Other (specify)	No response
2			3 Publisher University Consultancy	1

Role:

Librarian	Archivist	Conservator	Other	No response
2			2	2

			Research liason manager Information strategy manager	
--	--	--	---	--

What are your reasons for attending this event?

Understanding the issues

To learn more about digital preservation and the issues to consider re: long-term pres and access

Learn more about digital publication and the world of publishing and current issues in preservation

Awareness of issues around digital preservation and journals. Knowledge of different service providers

Invited by Fiona Murphy (Wiley-Blackwell) - interest in linking researchers to latest developments in this area

Speaker

On a scale from 1-5 how would you rate today's event (actual numbers):

	Not satisfied				Very satisfied	No answer
	1	2	3	4	5	
Relevance to you	0	0	1	4	1	0
Presenter(s)	0	0	0	3	3	0
Level of information	0	0	0	4	2	0
Venue and facilities	0	0	0	2	4	0
Value for money	0	0	0	3	3	0
Overall satisfaction	0	0	0	4	2	0

On a scale from 1-5 how would you rate today's event (%):

	Not satisfied				Very satisfied	No answer
	1	2	3	4	5	
Relevance to you	0	0	17	67	17	0
Presenter(s)	0	0	0	50	50	0
Level of information	0	0	0	67	33	0
Venue and facilities	0	0	0	33	67	0
Value for money	0	0	0	50	50	0
Overall satisfaction	0	0	0	67	33	0

Which sessions did you find most useful?

The nature of the problem: an introduction to e-journals..	1
Licensing e-journals in perpetuity	
Service providers forum	2
Discussion	
Emerging trends in e-journal preservation	2
Publishers' perspectives	1
Panel session	

Thinking about everything you've heard today, what do you think are the most important requirements which Trust and e-Journals ought to address?

Not necessarily in this order: 1. Complexity of new dynamic publishing models, 2. Post-cancellation access - perhaps need to simplify the levels of access and standardise them. 3. An 'at risk' register would imply an ultimate 'in care' state for those that fall through the net. Peter B suggests this should be the National Libraries.

Further education both of students and people working in this area, of issues explored today, Many are of relevance outside e-journals and form a fundamental base of dp

Shared terminologies and vocabularies for resource discovery. No point preserving stuff if people can't find it!

Partnering between researchers, their trusted partners (including national data centres eg NERC) and preservation practitioners

Transparency, audit, reliability, safe places network, sharing the task avoids disaster. Librarians service providers, publishers

What did you learn? What will you do (differently) as a result of attending this event?

Loads, but I am fairly new to the subject

Updates on current concerns in publishing world

Good networking and links to further info

More things around LOCKSS-CLOCKSS

Knowledge of service providers. Emerging trends - KEEPER, etc

Was there anything else you would have liked us to have included on the course?

No answers

What did we do well?

Very good at keeping a good pace and focus

Good presentations. Liked the 'rolling agenda with comments' (etherpad) even though it did not work fully on my iPad

Presentations, discussions all good. Conference location very good

Where did you hear about this event?

Archives-NRA	
Digital preservation list	2
DPC discussion	
Twitter	1
Preservation Advisory Centre website	

Other:

Via colleague

Any further comments

none