

The Keepers Registry: Enabling Trust in e-Journal Preservation

Peter Burnhill

EDINA, University of Edinburgh

31 January 2012

Wellcome Collection Conference Centre, London



Re-thinking Trust [New Zealand Archives]

- “X has **entrusted** Y with ensuring that important [] digital information is not lost and instead remains discoverable, reusable, secure and accessible over-time
- “needs to be tamper-proof and free of technological digital rights restrictions. It must be **trusted** to be authentic and reliable”.
- If no action is taken, [] digital information will be lost forever.

<http://archives.govt.nz/advice/government-digital-archive-programme>

Life Cycle Models, Provenance, Authenticity, Diplomatics

- “**Life cycle models** are shaping the way we study digital information processes .. represent the life course of a larger system, such as the research process *Chuck Humphreys*
 - ... **Life Cycle Information for E-Literature** (LIFE) and the **Digital Curation Centre** [] have been influenced by this approach. advancing life cycle models **to improve practices** ..”
- ... ‘**life cycle**’ is different from ‘**life span**’ ... implies an environment in which resources are ... **created, curated, made accessible, and preserved** for subsequent research, learning, and policy activity. The challenge is to infuse the data life cycle with the metadata and services ...”
Ann Green <http://www.iassistdata.org/blog/conceptualizing-digital-life-cycle>
- ‘When Documents Deceive: **Trust and Provenance** as New Factors for Information Retrieval in a Tangled Web’, *Clifford A. Lynch*, *JASIST*, 52(1) 2001
- “Archival science, with its principles of uniqueness, **provenance**, arrangement and description, **authenticity**, appraisal, and its tool sets such as **diplomatics** ..., [as] framework for a theoretical foundation for digital libraries.” *Seamus Ross*, Keynote Address ECDL2007, Budapest
- The underlying concept is **Trust**

In whom or in what should be put our trust?

In God We Trust ... for everlasting life

However, for secular purposes we need more than faith

- We still plan for the future, but for this life not the next
 - ‘in perpetuity’ is claim of long life that requires ‘living will’
- What we are interested in the material world: ‘content’
 - The digital is different: we can make ‘copies’, not something ‘unique’
 - * But recall that e-Journal content is intended as ‘record’
- Trust implies risk
 - To have left undone those things which we ought to have done
 - To lose things that we entrusted others to do
- Not (just) trust in people but trust in procedures
 - Importance of fire drill procedure, and of fire drill practice

Trust and the avoidance of disaster & deceit

"We all make mistakes; the trick is being able to recover"

- 'Trust' means more than 'reliance'
 - Failure threatens more than disappointment: significant loss
- When we trust, or are trusted by others, we have interest in:
 - methodology, that is explicit, testable and 'proven'
 - competence to do what is required [who issues certification?]
 - commitment to ;do it, in a timely fashion
 - ***disclosure*** of what is done, and then ***audit*** of what is said and done
- How do you know who, if anyone, is trusted to look after XYZ?

The Keepers Registry & how it came about

- The Keepers Registry is about disclosure of archival action
 - *Who is doing what, and how?*
- Aims to be a global online facility, with dual purpose:
 1. Reporting who is looking after what e-journal (& how)
 2. A showcase for organisations having archival intent
- Idea of e-journals preservation registry mooted in reports,
M.Jones, 2003/4
- 'clarity of public statement by each agency or through a registry'
CLIR Report, 2006
- JISC commissioned scoping study that recommended an e-journals preservation registry should be built, possibly as part of UK Union Catalogue of Serials (SUNCAT), Rightscom & Loughborough U, 2007

The Keepers Registry & how it came about

EDINA & ISSN-IC became partners in the JISC-funded PEPRSproject: Piloting an E-Journal Preservation Registry Services

* Phase 1: August 2008 – July 2010

‘investigate, prototype and build’ [evaluation in Feb. 2010]

* Phase 2: August 2010 – July 2012

‘preparing for service & governance’

Initially UK in scope, but necessarily international, with six of the world’s leading ‘digital preservation agencies:

British Library, CLOCKSS, e-Depot, LOCKSS & ioe Portico

Beta Launch: PEPRS () & The Keepers Registry ()

Now preparing for Phase 3: August 2012 –



Discover who is looking after what e-journals

Search

Input a search term (e.g. free text, or a title or ISSN(s))

[+ Display example searches](#)



[Glossary and definitions](#)

The Keepers Registry allows you to search on title or ISSN

Learn more about the [archiving agencies](#) in The Keepers Registry who have stepped forward to take responsibility for ensuring long-term continuity of access to global scholarship.

- [British Library](#) (Last updated: 30 Dec 2011)
- [CLOCKSS Archive](#) (Last updated: 12 Apr 2011)
- [e-Depot](#) (Last updated: 11 Apr 2011)
- [Global LOCKSS Network](#) (Last updated: 26 Jan 2012)
- [HathiTrust](#) (Last updated: 01 Jan 2012)
- [Portico](#) (Last updated: 06 Jan 2012)

Also acting as showcase for the activity of archiving organisations

Discover who is looking after what e-journals

Search

Input a search term (e.g. free text, or a title or ISSN(s))

[+ Display example searches](#)



[Glossary and definitions](#)

Can also search on
publisher

Learn more about the [archiving agencies](#) in The Keepers Registry who have stepped forward to take responsibility for ensuring long-term continuity of access to global scholarship.

- [British Library](#) (Last updated: 30 Dec 2011)
- [CLOCKSS Archive](#) (Last updated: 12 Apr 2011)
- [e-Depot](#) (Last updated: 11 Apr 2011)
- [Global LOCKSS Network](#) (Last updated: 26 Jan 2012)
- [HathiTrust](#) (Last updated: 01 Jan 2012)
- [Portico](#) (Last updated: 06 Jan 2012)

Soon to add activity of National
Science Library of China
+ 3 others waiting:
from Canada, UK and USA

Royal Society of Chemistry: 74 serials being archived by the Keepers. 3 have ISSN for Online & Print; 2 ISSN for Print only. 3 agencies appear: some content 'Preserved', others 'in progress'.

Royal Society of Chemistry

74 hits found

Current search: "R"

Title	ISSN	Publisher	Current extent of archiving 	Archiving agency
1. Analyst	1364-5528 (Online); 0003-2654 (Print)	Royal Society of Chemistry	Preserved: v. 6-8	CLOCKSS Archive
		Royal Society of Chemistry	Preserved: v. 133-135	CLOCKSS Archive
		Royal Society of Chemistry	In progress: v. 123	Global LOCKSS Network
		Royal Society of Chemistry	Preserved: v. 122, 124-134	Global LOCKSS Network
		Royal Society of Chemistry	Preserved: v. 134	Portico
2. Analytical communications	1364-5536 (Online); 1359-7337 (Print)	Royal Society of Chemistry	In progress: v. 34-35	Global LOCKSS Network
		Royal Society of Chemistry	Preserved: v. 36	Global LOCKSS Network
		Royal Society of Chemistry	In progress: Not yet processed.	Portico
3. Analytical methods	1759-9679 (Online); 1759-9660 (Print)	Royal Society of Chemistry	Preserved: v. 1-2	CLOCKSS Archive
		Royal Society of Chemistry	Preserved: v. 1	Global LOCKSS Network
		Royal Society of Chemistry	Preserved: v. 2	Portico
4. Analytical proceedings	0144-557X (Print)	Royal Society of Chemistry	In progress: Not yet processed.	Portico
5. Annual reports on analytical atomic spectroscopy <i>Variant title: Analytical atomic spectroscopy</i>	0306-1353 (Print)	Royal Society of Chemistry	In progress: Not yet processed.	Portico

'Wiley': 1381 serials being archived by the Keepers. This page highlights significance of change in publisher, co-incident with change in Keeper. But this is only 1 of 139 pages of record ...

1381 hits found
Current search: wil

Title	ISSN	Publisher	Current extent of archiving 	Archiving agency
1241. Risk analysis	1539-6924 (Online); 0272-4332 (Print)	Kluwer Academic Publishers John Wiley & Sons, Inc.	Preserved: v. 18-19, 26-27 In progress: Not yet processed.	e-Depot Portico
1242. Risk management and insurance review <i>Variant title: Risk management & insurance review</i>	1540-6296 (Online); 1098-1616 (Print)	Blackwell Publ. John Wiley & Sons, Inc.	Preserved: v. 9-10 In progress: Not yet processed.	e-Depot Portico
1243. Risknews <i>Variant title: Risk news (Weinheim. Internet)</i>	1616-0045 (Online); 1612-8931 (Print)	John Wiley & Sons, Inc.	Preserved: v. 1-2	Portico
1244. River research and applications	1535-1467 (Online); 1535-1459 (Print)	John Wiley & Sons, Inc.	Preserved: v. 18-26	Portico
1245. Russian review	1467-9434 (Online); 0036-0341 (Print)	Blackwell Publishers John Wiley & Sons, Inc.	Preserved: v. 65-66 In progress: Not yet processed.	e-Depot Portico
1246. Scandinavian journal of caring sciences	1471-6712 (Online); 0283-9318 (Print)	Blackwell John Wiley & Sons, Inc.	Preserved: v. 20-21 In progress: Not yet processed.	e-Depot Portico
1247. The Scandinavian journal of economics	1467-9442 (Online); 0347-0520 (Print)	Blackwell Publishing John Wiley & Sons, Inc.	Preserved: v. 108-109 In progress: Not yet processed.	e-Depot Portico
1248. Scandinavian journal of immunology	1365-3083 (Online); 0300-9475 (Print)	Blackwell Science John Wiley & Sons, Inc.	Preserved: v. 63-66 In progress: Not yet processed.	e-Depot Portico
1249. Scandinavian journal of medicine & science in sports <i>Variant title: Scandinavian journal of medicine and science in sports</i>	1600-0838 (Online); 0905-7188 (Print)	Blackwell John Wiley & Sons, Inc.	Preserved: v. 16-17 In progress: Not yet processed.	e-Depot Portico
1250. Scandinavian journal of psychology	1467-9450 (Online); 0036-5564 (Print)	Blackwell Publishers John Wiley & Sons, Inc.	Preserved: v. 47-48 In progress: Not yet processed.	e-Depot Portico

Progress, but far from 'job done'

The Keepers Registry <thekeepers.org> reports archival action for:

- 20,024 titles (unique ISSN-L), of which 14,829 have an e-ISSN
Print only = 5,193 Online only = 1,252
- 16,558 have status 'Preserved' 😊
['In progress' reported for a further 9,886]

However ...

- that includes 'annual reports' and journal content having an ISSN assigned for the print but (not yet) the 'e'. [digitised journals]
- as shown, the coverage of volumes for Titles is sparse

AND

ISSN Register has 1.6m entries, with over 98,000 ISSNs for e-serials 😊

Ulrichweb* reports 227,604 Active Titles, of which:

- 75,621 Academic/Scholarly Titles
- 29,907 Refereed Titles [the "30,000" that everyone quotes]

[Yvette Diven of SerialsSolution At Mid-summer ALA 2011 (New Orleans)]

We have been busy building international support...

2008: JISC Journals WG, London; ISSN National Directors Meeting

2009: NASIG Annual Conference, Ashville NC, USA;
Lib. of Academy of Science, Beijing; ISSN Directors, Beijing;
PARSE.Insight, Germany; Knowledge Exchange, Edinburgh

2010: E-journals are Forever Workshop, JISC/DPC, London;
IFLA 2010 Gothenburg; RLUK Conference, Edinburgh;
Columbia Univ., NYC

2011: UKSG; ISSN Governing Body; ARL, Montreal; ALA; UNESCO;
JISC Archiving Implementation Group (JARVIG) ...

2012: Digital Preservation Coalition;

P.Burnhill, F.Pelle, P.Godefroy, F.Guy, M.Macgregor, A.Rusbridge & C.Rees
Piloting an e-journals preservation registry service.
Serials 22(1) March 2009. [UK Serials Group]



How to assist DPC members with other works?

“key barriers to preservation being adopted more widely”

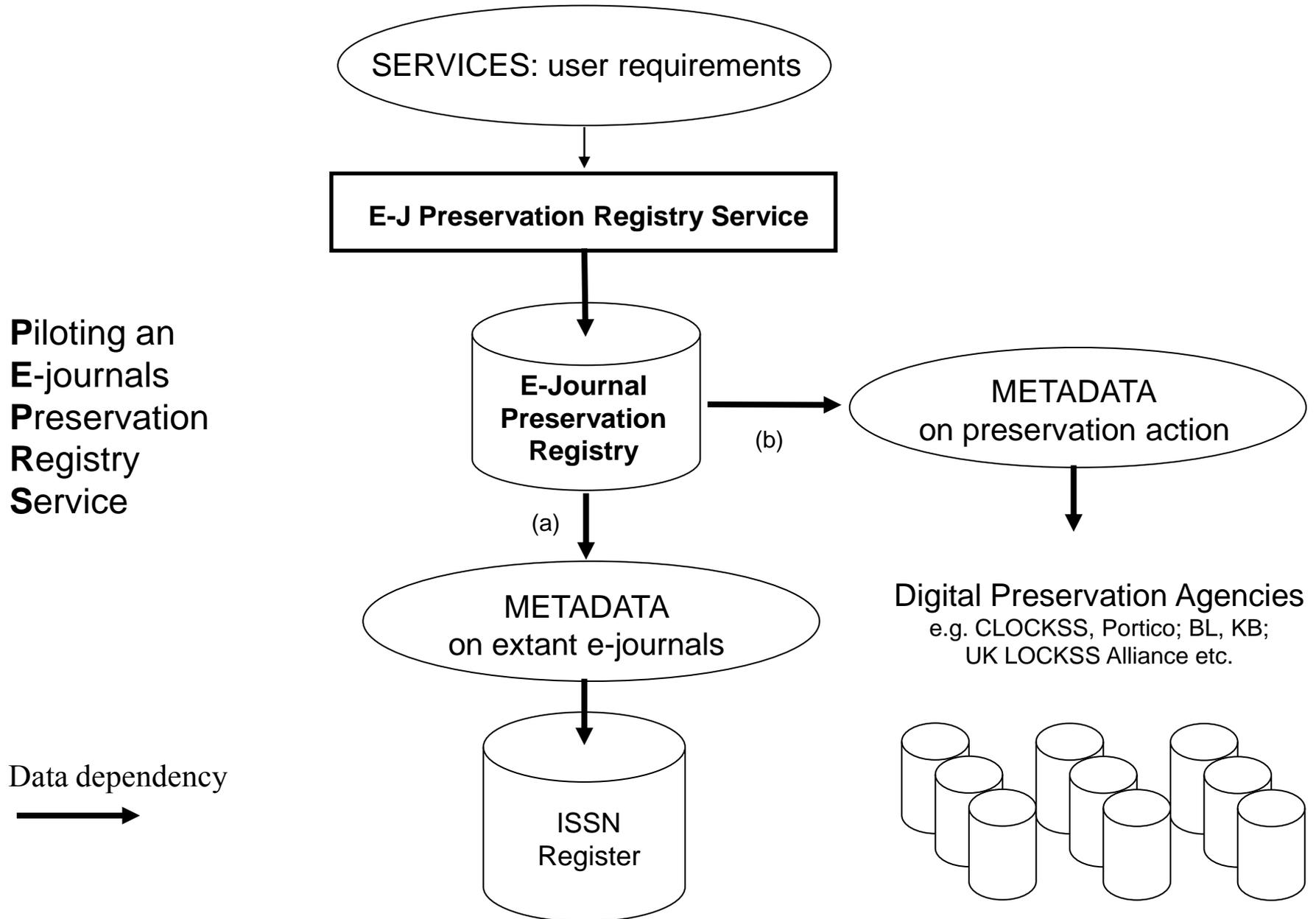
What lessons from the Keepers Registry, which has its focus on e-journal preservation, for preservation of other digital works?

- It enables those with archival intent to advertise what they are doing, with purposeful replication [not ‘unnecessary’ duplication]
 - A means for <stewards> to demonstrate that their content is being preserved, and potential users to know that content they care about is safe

The Abstract Data Model may generalise, with explicit dependence upon:

1. good metadata on the digital object (e-journals) and
2. metadata about the archival activity of the Keepers

Abstract Data Model: Figure 1 in reference paper in Serials, March 2009



1. good metadata on the digital object

Keepers Registry benefits greatly for e-journals because:

- prior existence of scheme for assignment of identifiers (ISSN)
- availability of a content registry (ISSN Register)
- ISSN-L, a link field between ISSN for print & that for digital
- The idea of there being many copies [print had led the way!]
- Could there be other aspects: Consensus that this is 'record'?

BUT still non-trivial, with challenges because serials are 'living' objects having successive parts:

- Articles are the objects of desire, contained within issues
- Serial-level: ISSN assignment for growth of digitised journals; some errors in extant metadata for archived content, variation in data on publisher
- Volume-level: universal holdings statements: records & display
- The Keepers Registry needed to a global online facility
-

2. (good?) metadata on the archival action

Working with the (leading) Keepers

- Self-statement (**disclosure**) on 'areas of archival action':
 - need to define/implement consensus on fields & vocabulary
- a) *Brief summary*
- b) *Ingest/preservation workflow*: what each does (methodology)
- c) *Library access to content*: statement of terms & conditions
- d) *Auditing of content, policies & procedures (inc. external audit)*: to ensure ongoing authenticity & accessibility of content
- e) *Latest data*: report of serials/journals being archived, with archival status and volume detail.
- The Keepers Registry needed to a global online facility

How to assist DPC members?

“key barriers to preservation being adopted more widely”

What lessons from the Keepers Registry, which has its focus on e-journal preservation, for preservation of other digital works?

What's difference between journals & 'special collections'?

Copies!

What's different about the digital?

- both the born digital and the digitised

Everything can readily be copied!

- and improved, value-added, combined, curated ..

And that gives opportunity for safe-keeping action

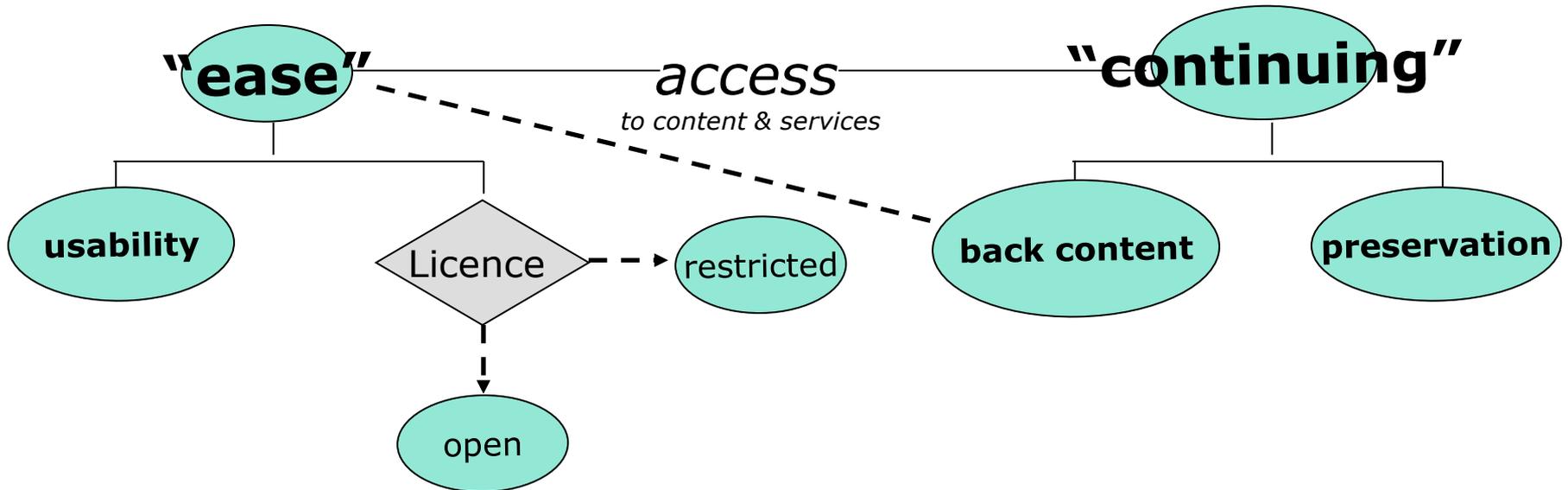
Re-thinking Trust: assurance of continuing access

- *Stewardship* (for digital content) as action on trust
 - the trust put in libraries and archives by present and future children, citizens and scholars
 - * not just budgets and provision of cost-effective access
- *Digital Accident*: ease of online access *also* means that libraries no longer hold back copy on their shelves
 - A challenge of stewardship for publishers as well as libraries
 - * *“What to do with that responsibility: outsource to libraries?”*
- *The Central Task* (for research and university libraries et al):

Ensure that researchers, students & their teachers have continuity of access to the (online) scholarly resources they need.

Ensuring

researchers, students and their teachers have
ease and continuing access
to online scholarly resources

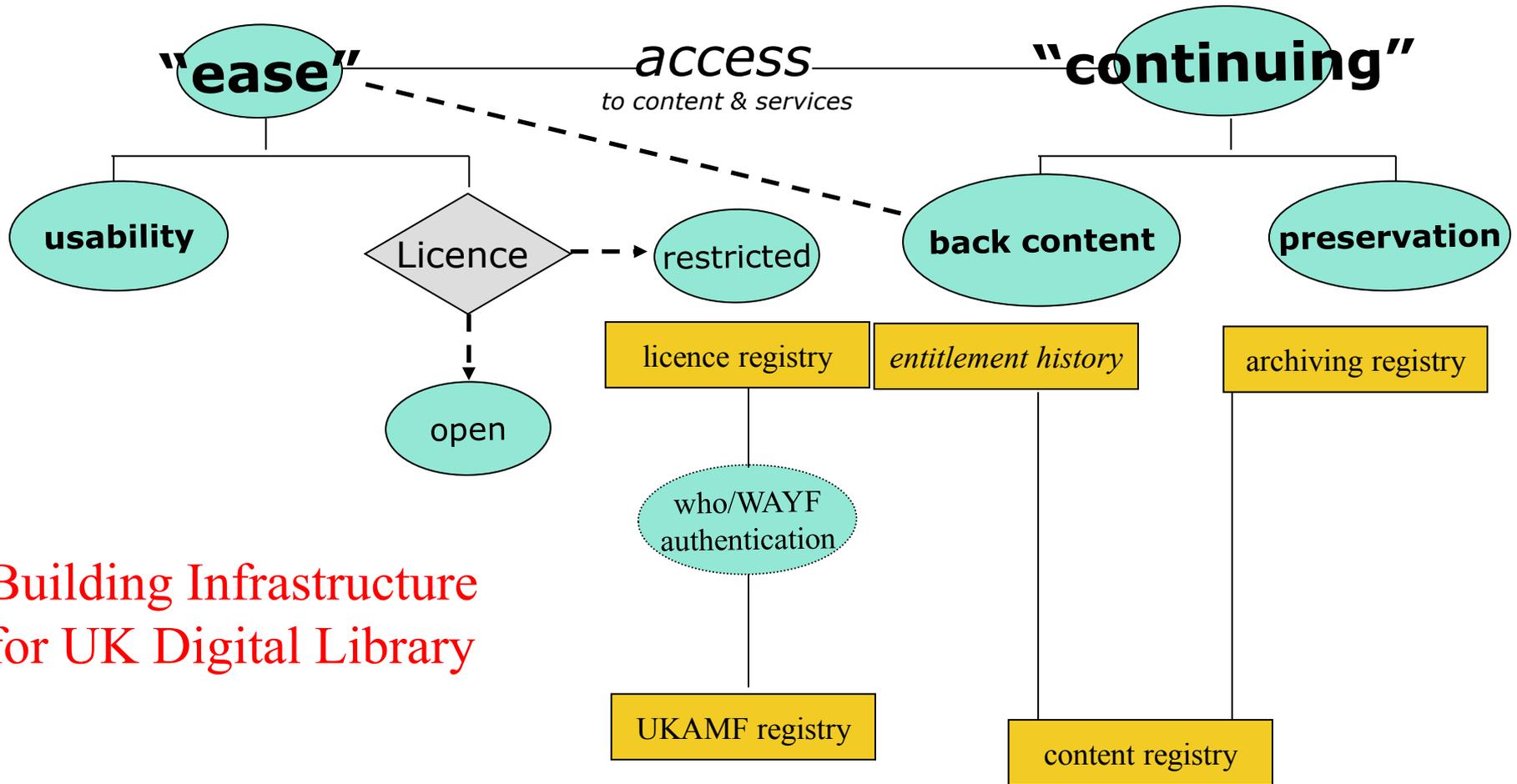


**Building Infrastructure
for UK Digital Library**

Classic use case: article-length work as journal content

P.Burnhill, Edinburgh 2009

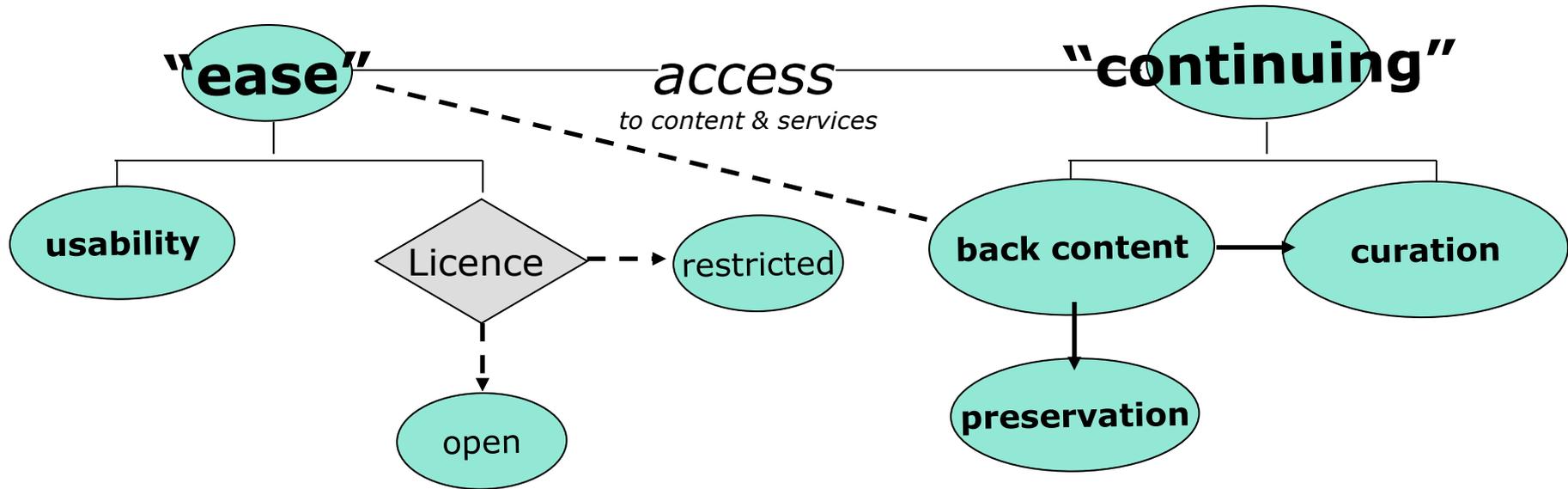
Ensuring ease and continuing access



Building Infrastructure
for UK Digital Library

Ensuring

researchers, students and their teachers have
ease and continuing access
to online scholarly resources

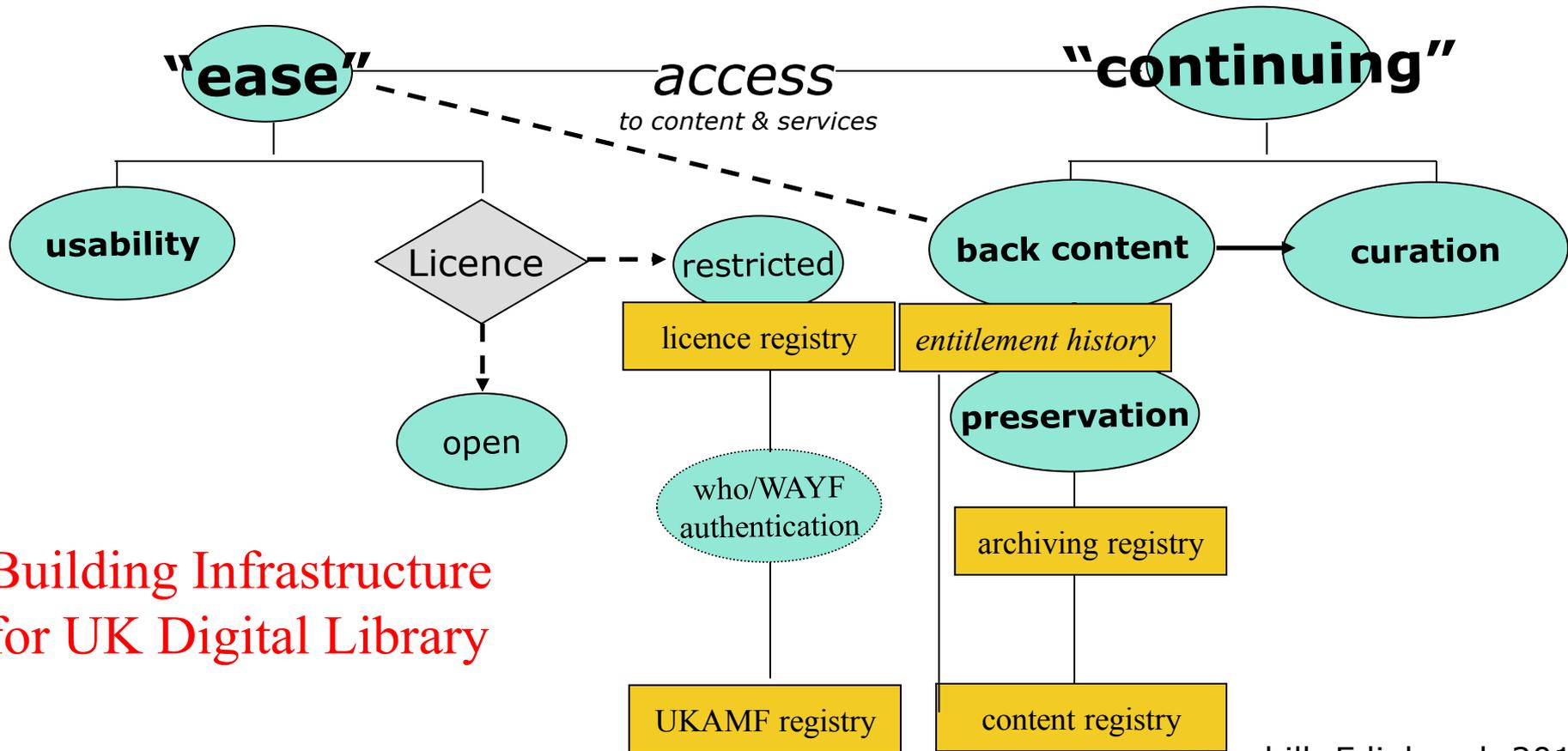


A broader view of Digital Library
f (digital preservation, data curation)

f (“document tradition & computation tradition”), M.Buckland, 1998

Ensuring

researchers, students and their teachers have
ease and continuing access
to online scholarly resources



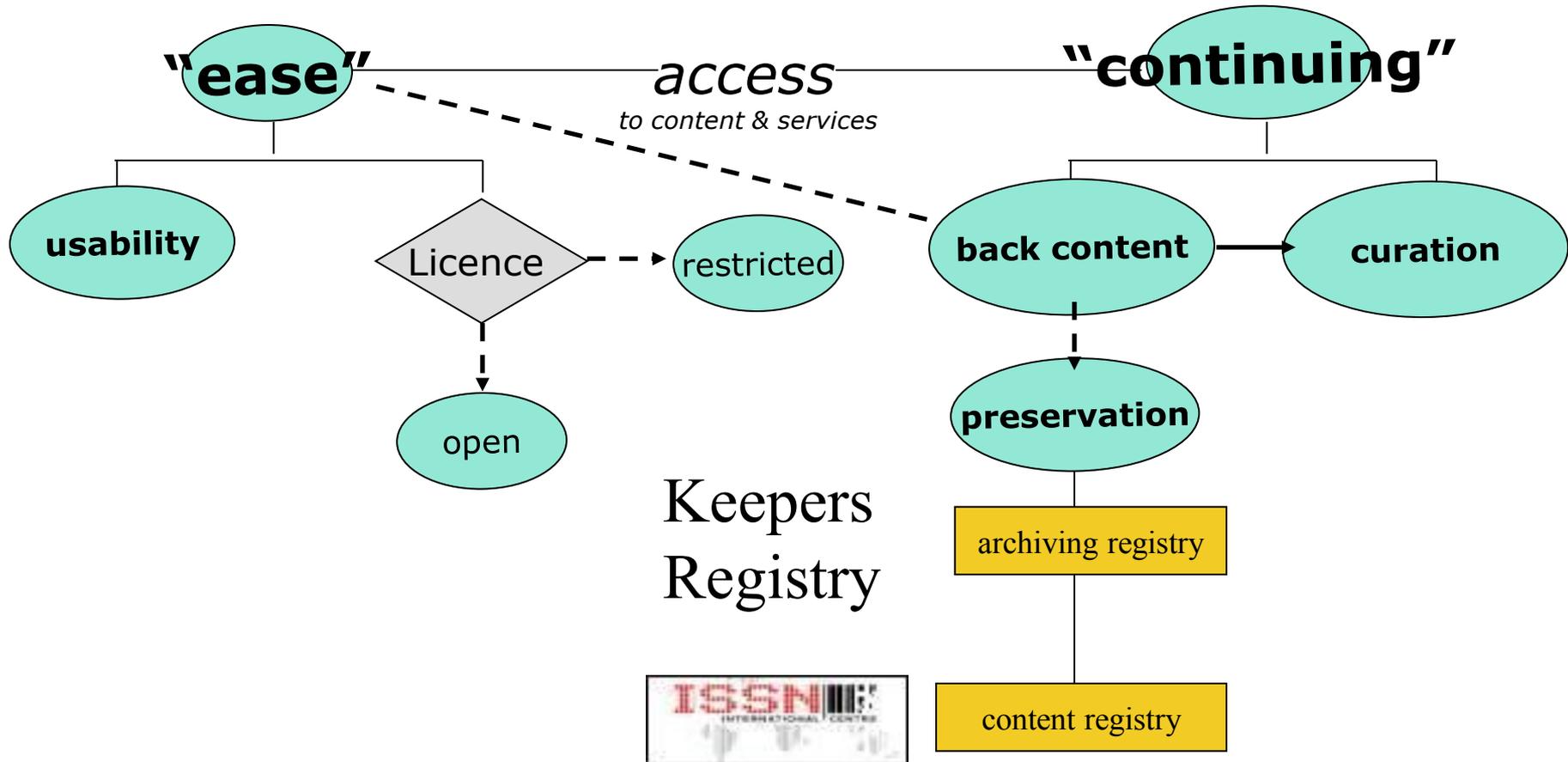
Building Infrastructure
for UK Digital Library

Ensuring

researchers, students and their teachers have

ease and continuing access

to online scholarly resources



Trust: avoiding disaster by sharing the task

- When we trust, or are trusted by others, the steps are:
 - 1. methodology**, that is explicit, testable & 'proven'
 - * Locally that should include '**copies**'
 - * Globally level that should include **different approaches**
 - 2. competence** to do what is required
 - * Mix of education, skill, knowledge and wisdom
 - 3. commitment** to do it, in a timely fashion
 - * That means **funds**, to secure and **sustain effort**
 - 4. disclosure** of what is being done
 - * Can the 'Registry of Keepers' model be generalised?
 5. and then **audit and report** what was said & done

"We all make mistakes; the trick is being able to recover" 😊

How to assist DPC members?

What's different about the digital?

- both the born digital and the digitised

Everything can readily be copied!

- and improved, value-added, combined, curated ..

And that gives opportunity for safe-keeping action

Enlightenment = Many Copies for Content

$$E = MC^2$$