



Metes and Bounds: E-Journal Archiving Landscape

Anne R. Kenney

Joint workshop on e-journal archiving and
preservation

British Library Conference Centre

April 2007



Cornell University
Library



Why Worry?

- “Grand challenge facing higher education”
- Shift to e-only in publishing, purchasing, and use
- Inadequate protection of digital content
- Uncertainty

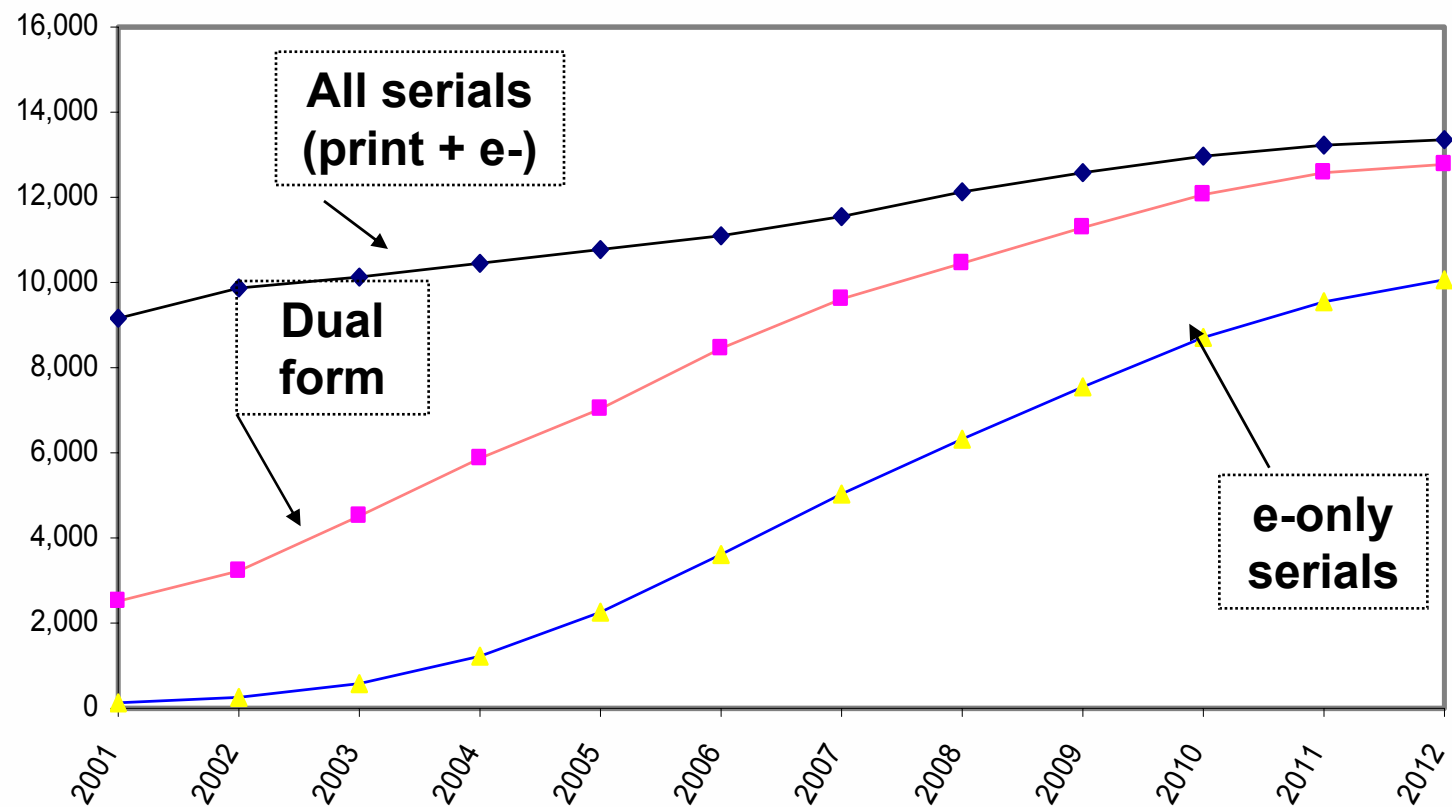


Shift to Electronic Publishing

- Dramatic increase in online publications in past decade
- Elimination of print versions
- Estimate by 2016 half of serials will be e-only
- STM titles and smaller publishers most likely to go e-only first



Predicted Growth of Serials Publications (EPS study for BL)





User Prefer Online Journals

- Use of online journals dwarfs print use
- Over 75% of scholars view e-journals as invaluable research tools
- Users worry about preservation, but willing to accept cancellation of print if choice has to be made



Libraries Move to E-only

- 1,400% increase in funds spent on e-journals in a decade; currently represent 42% of serials expenditures. (ARL)
- STM e-journal subscriptions to exceed print by 2008 (EBSCO)
- Cancellation of print subscriptions when e-journals available
- Perpetual access vs archiving



Scope of Study

- Define information needs of library directors
- Identify and review most promising programs
- Assess data
- Issue report and recommendations

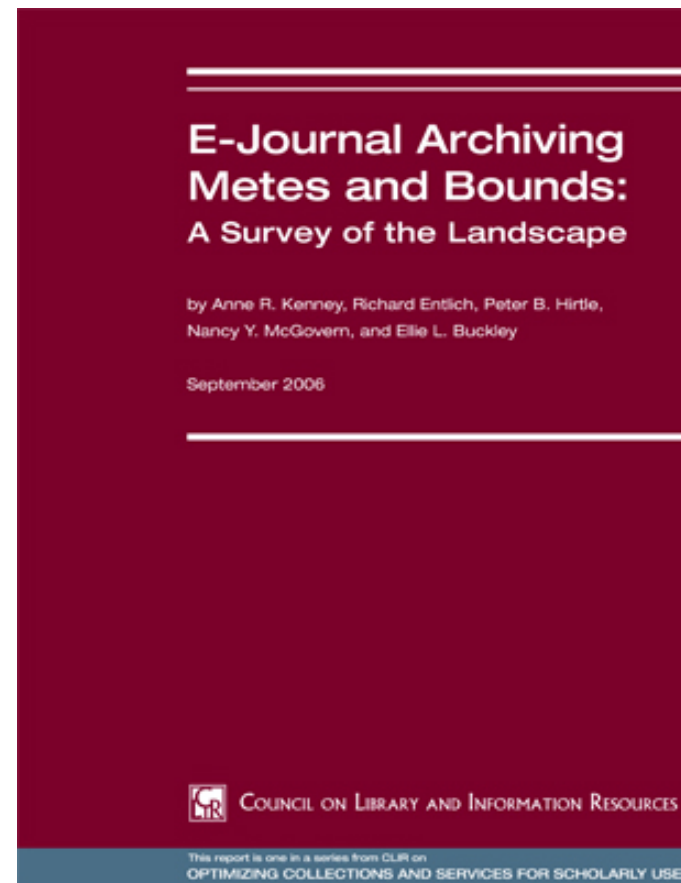
CLIR pub 138:

E-Journal Archiving Metes and Bounds: A Survey of the Landscape

by Anne R. Kenney, Richard Entlich, Peter B. Hirtle,
Nancy Y. McGovern, and Ellie L. Buckley

September, 2006. 120 pp. \$30
ISBN 1-932326-26-X
ISBN 978-1-932326-26-0

[www.clir.org/pubs/reports/pub138/pub138.
pdf](http://www.clir.org/pubs/reports/pub138/pub138.pdf)





Library Director Concerns

- Library motivation (why invest in this?)
- Content coverage (is content I'm interested in included?)
- Access (what will I gain access to? when? under what conditions?)
- Program viability (will these efforts last?)
- Library responsibilities (what will this cost in terms of time, expertise, funding?)
- Technical approach (will this really preserve the material?)



3 Themes Emerged

- Sense of urgency
- Resource commitment and competing priorities
- Need for collective response



Identifying Programs

- Not-for-profit program independent of publisher
- Commitment to archiving scholarly peer-reviewed e-journals
- Formal arrangements with publishers
- Program in place
- Beneficial to academic libraries



The Group of 12

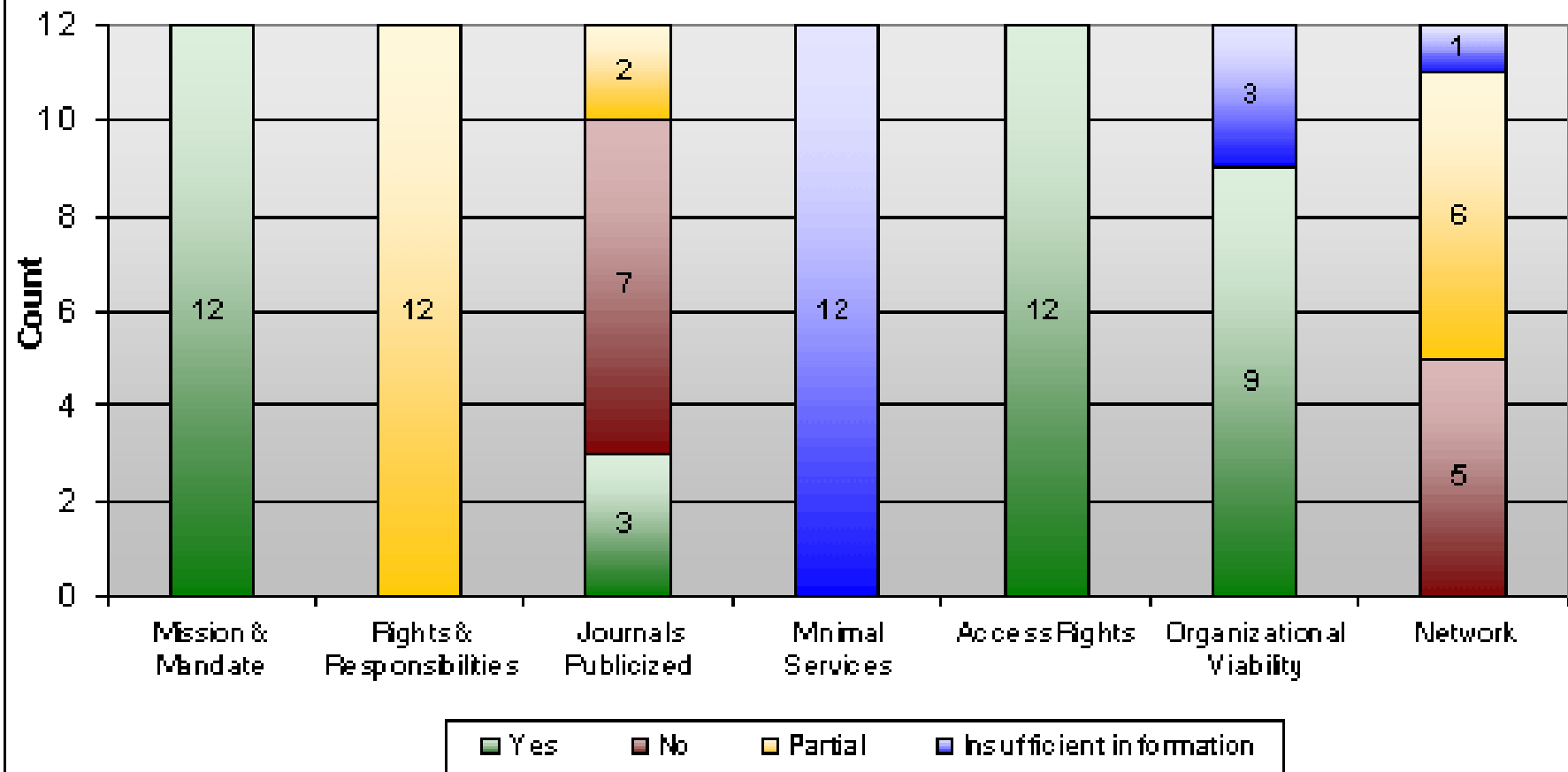
- CISTI (Canada)
- CLOCKSS
- OCLC ECO
- OhioLINK EJC
- KB e-Depot (Holland)
- kopal/DDDB(Germany)
- LANL-RL
- LOCKSS Alliance
- Ontario Scholars Portal (Canada)
- PANDORA (Australia)
- Portico
- PubMed Central



Seven Program Indicators

- Explicit mission and necessary mandate
- Necessary rights and responsibilities
- Content coverage
- Minimal set of services
- Access and triggers
- Organizational viability
- Network

Is Evidence of the Following Indicators in Place?

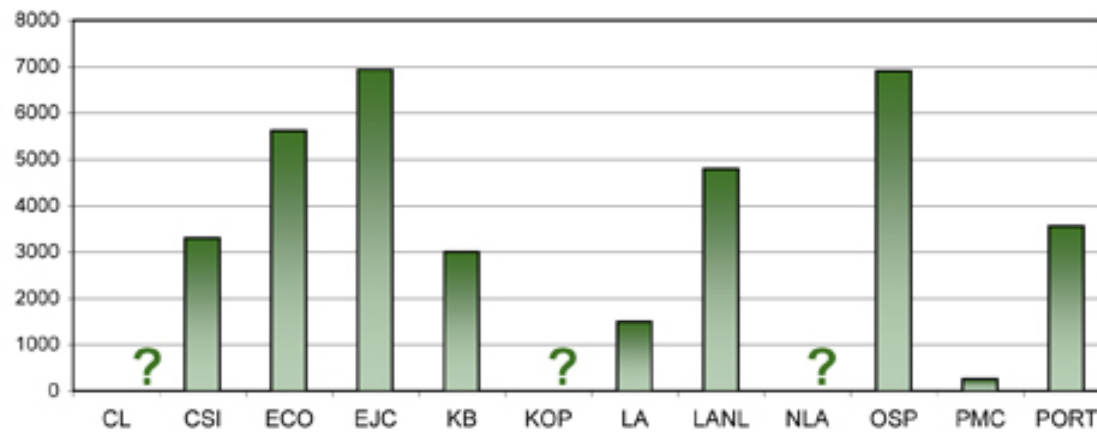




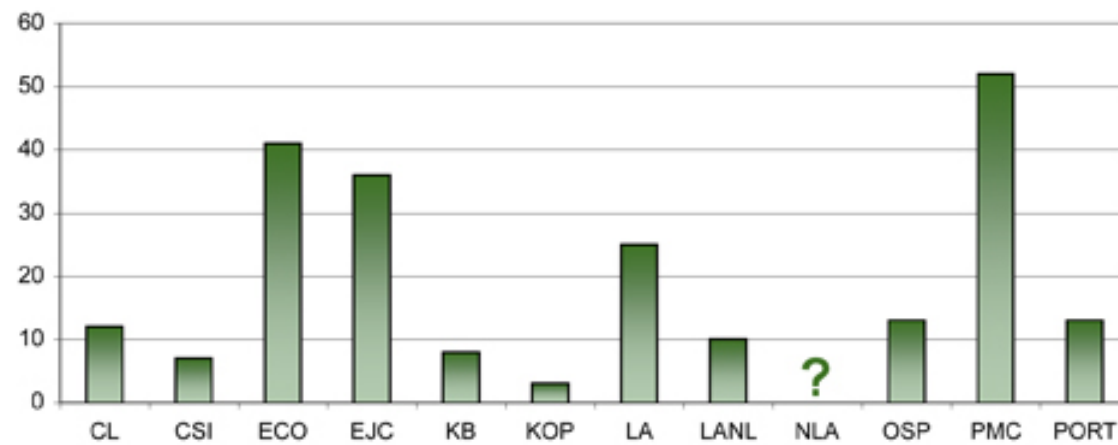
Content

- 128 publishers involved
- Over 34,000 titles included -- with significant duplication across services
- Difficult to create definitive list
- Major publishers well represented
- Redundancy vs greater content coverage

Approximate Number of Titles Reported

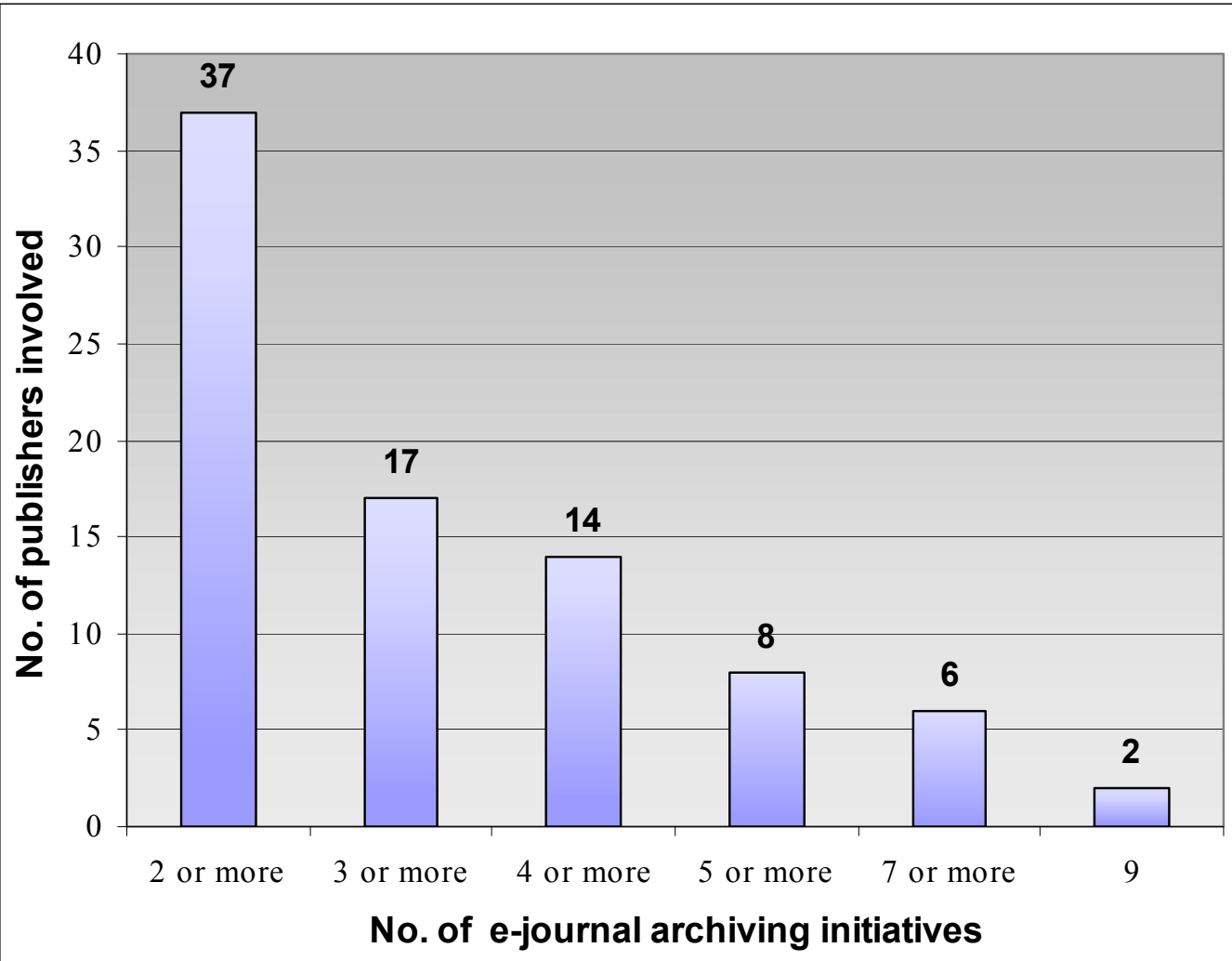


Number of Publishers





Publisher Overlap





Minimal Services

- Receive files in standard form
- Store in non-proprietary formats
- Integrity testing
- Files processing
- Security
- Ensure transparency



Commitment to Standards

Standard or Practice	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
OAIS	•			•	•	•					•
Trusted Digital Repositories				P		•	P				•
PREMIS	•			P		P	P				•
OAI-PMH	•			P		•	•			•	•
JAI DTD (NLM)	•									•	•
Audit Checklist				P		P					•
Open source software	•				•	•	•	•			•
Open file formats					•	•	•		•	•	•
Nonproprietary storage media						•	•				•
Other				P	•						

Table 5. Responses to question: "Do you follow any of the following standards and best community practices for archiving?" (• = yes; P = plan to within six months)



Validation/Testing

	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Completeness upon ingest	•	•	•	•	•	•	•	•	•	•	•
Integrity upon ingest	•	•		•	•	•	•	•	•	•	•
Ongoing integrity		•		•	•	•	P	•	N/S	•	•

Table 10. Responses to question: "Do you conduct validation/testing?" (• = yes; N/S= not sure; P= plan to within six months)



Archiving Strategies

Archiving Strategy	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Migration	•	•	•	•	•	•	•	•		•	•
Emulation				•	•			•			
Normalization	•						•	•	•	•	•
Reliance on standards	•			•	•		•	•		•	•
Refreshing	•		•	•	•			•		•	•
Use of durable media				•	•				•	•	•

Table 11. Responses to question: "What type of archiving strategies do you use or plan to use?"



Redundancy Procedures

Redundancy Procedure	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Local backups	•	•	•	•	P		•	•	•	•	•
Mirror sites	•	•		P						P	P
Off-site storage	•	•	•	P	•		•	•	•	•	•
Not sure											
Other						•					
Minimum number of copies (including master or production system)	2	4	3	2	4	See text	5	3	2	6	6

Table 12. Responses to questions: "Do you use any of the following redundancy procedures?" and "How many copies of your content do you maintain?" (• = yes; P = plan to within six months)



Documentation

Documentation Type	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Mission statement			•	•	•					•	•
Publishers agreements	•		•	•	•	•	•	•		•	•
Membership agreements							•				•
Selection/Acquisition policies	•		•	P	•	•	•	•	•	•	•
Transfer requirements and deposit guidelines	•		•	•	•					•	
Ingest			•	•	•		•			•	•
Archival storage							•			•	•
Quality control							•			•	•
Auditing							•				•
Data management	•						•			•	•
Disaster planning/Recovery	•		•							•	P
Preservation planning					P						P
Metadata				•			•			•	•
Access and use policies							•	•			•
Financial reports											•
Annual reports				•							

Table 14. Responses to question: "Do you have the following written documentation that explicitly refers to e-journal archiving?" (• = yes; P = plan to within six months)



Current Access Conditions

- Online-limited (5)
- Online-open & moving wall (2)
- On-site (2)
- Trigger/audit only (3)



Trigger Events

Trigger Event	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Publisher ceases operation				•	•	•					•
Publisher no longer offers back issues				•	•	•					•
Copyright expires					•	•		•			
Journal ceases publication				•		•					•
Catastrophic failure				•		•	•				•
Temporary failure				•		•					
Other								•			•

Table 15. Trigger events that spark changes in access for the authorized community



Organizational Viability

- Very recent efforts; limited track record
- Limited auditing/reporting
- Sources of funding
- Stakeholder buy-in



Network of Repositories

- Exchanging information, strategies, software, documentation
- Little selection coordination
- No formal succession plans

Network Indicators

[illegible]



Conclusions

- Trigger events will happen
- Libraries cannot do this alone
- Current license terms for libraries are inadequate (perpetual access does not equal preservation)
- Viable options are emerging
- No single archiving program will meet all needs
- Coverage is very uneven
- Much content is at risk
- Libraries can and should influence developments
- Legislation needed -- legal deposit
- All programs need greater support, transparency, etc.



Recommendations for libraries

- Press publishers to enter archiving relationships
- Share information about what they are doing and how they are making decisions
- Join at least one initiative
- Press existing programs to meet their needs
- Develop a registry of archived publications
- Lobby programs to participate in networks for information sharing, best practices, etc.



Recommendations for publishers

- Enter into relationships with one or more e-journal archiving programs
- Provide adequate information and data to archivers
- Extend liberal archiving rights in their licensing agreements.



Recommendations for e-journal archiving programs

- Present evidence of minimal level of services for long-term, well managed collections (open to audit, certified)
- Be overt and explicit about what is archived
- Assure appropriate property rights
- Negotiate with regard to eventual placement in the public domain
- Form a network of mutual support and interdependence



Some Caveats

- The report does not endorse any of the programs
- Libraries/consortia/publishers should determine whether programs meet their needs
- Involvement of all parties essential
- It's a moving target



Post report action

- ARL endorses report and urges members to
 - support and participate in trustworthy and effective e-journal archives
 - press e-journal publishers to participate in trustworthy archiving of their titles.
- CRL/RLG Programs, Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist (March 2007)
- JISC Review and analysis of report for UK education sector



What Can UK Community Do?

- Review the scholarly literature most critical to higher education in UK; consider subject approach
 - ✓ NESLI2 and JISC/CURL LOCKSS Pilot
- Endorse set of principles *and* actions
 - ✓ E-journal archiving workshop
- Acknowledge power of collective action
- Consider role for funding bodies in research, development and requirements
- Foment international cooperation

Digital preservation represents one of the grand challenges facing higher education. Yet... the responsibility for preservation is diffuse and the responsible parties have been slow to identify and invest in the necessary infrastructure. The shift from print to electronic publication of scholarly journals is occurring at a particularly rapid pace; the digital portion of the scholarly record is increasingly at risk and solutions may require unique arrangements within the academy for sharing preservation responsibility.

Adapted from "Urgent Action Needed to Preserve Scholarly Electronic Journals," Don Waters et al, 10/2005