

Policy, Preconditions and Costs

Opportunities and Pitfalls in Long-term Digital Preservation

Neil Beagrie

and

Marcel Ras

Curating Research, KB, April 2009

Agenda

- Introductions
- Format of Workshop
- One point of view: David Rosenthal
- Long-term preservation of scientific publications
- Costs – Keeping Research Data Safe
- Policy – Digital Preservation Policies Study
- Final Discussion and Questions

One Point of View

David Rosenthal – Spring CNI Plenary
Presentation [edited/abridged]

The Content

- **Ancient History: before 1995**
 - Jeff Rothenberg's 50year look forward from 1995
 - What he predicted & why
- **Modern History: from 1995 to 2009**
 - Impacts of Jeff's article
 - What else happened
 - How Jeff rates as a prophet & why
- **The Future: following Jeff's example**
 - Looking forward to identify the real problems

The Summing Up



- **Jeff being wrong is Good News!**
 - Collections that survive aren't as hard as we thought
- **Just collect and keep the bits**
 - Not collecting is the major reason for stuff being lost
- **If you keep the bits, all will be well**
 - Current tools will let you access them for a long time

The Real Problems Were..

➤ **Scale**

- Not individual documents but vast collections of them

➤ **Cost**

- Preservation not by individuals but large organizations

➤ **Intellectual Property**

- If content worth saving someone is making money from it

Economics

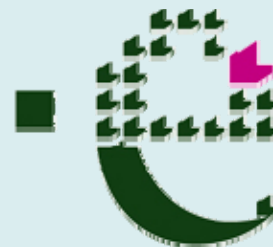
- 2008 Preservation Buzzword: Sustainability
- Future stuff will be much more expensive. There'll be a lot more bytes of it
- Bytes vulnerable to money supply glitches
 - Data needs to be endowed if it is to survive hard times. Endowing up front means preserving less.
- Collection development: what must be kept? But it has really bad scaling problems

Discussion and Questions?

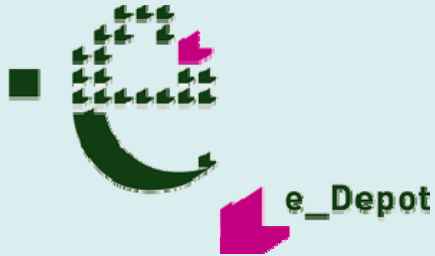
Long-term preservation of scientific publications in practice

The KB e-Depot

Koninklijke Bibliotheek
National Library of the Netherlands



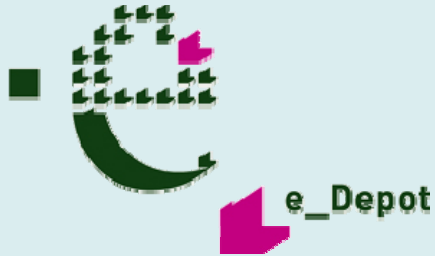
 e_Depot



The KB e-Depot (1)

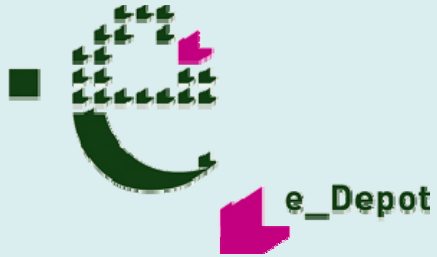
- KB is national library of The Netherlands
- Task of a national library is to collect, describe and preserve national imprint
- Paper, but also digital

The mission of the KB e-Depot is to collect published information, preserve it and provide permanent access to the information for use in research, education or for any other purpose in society

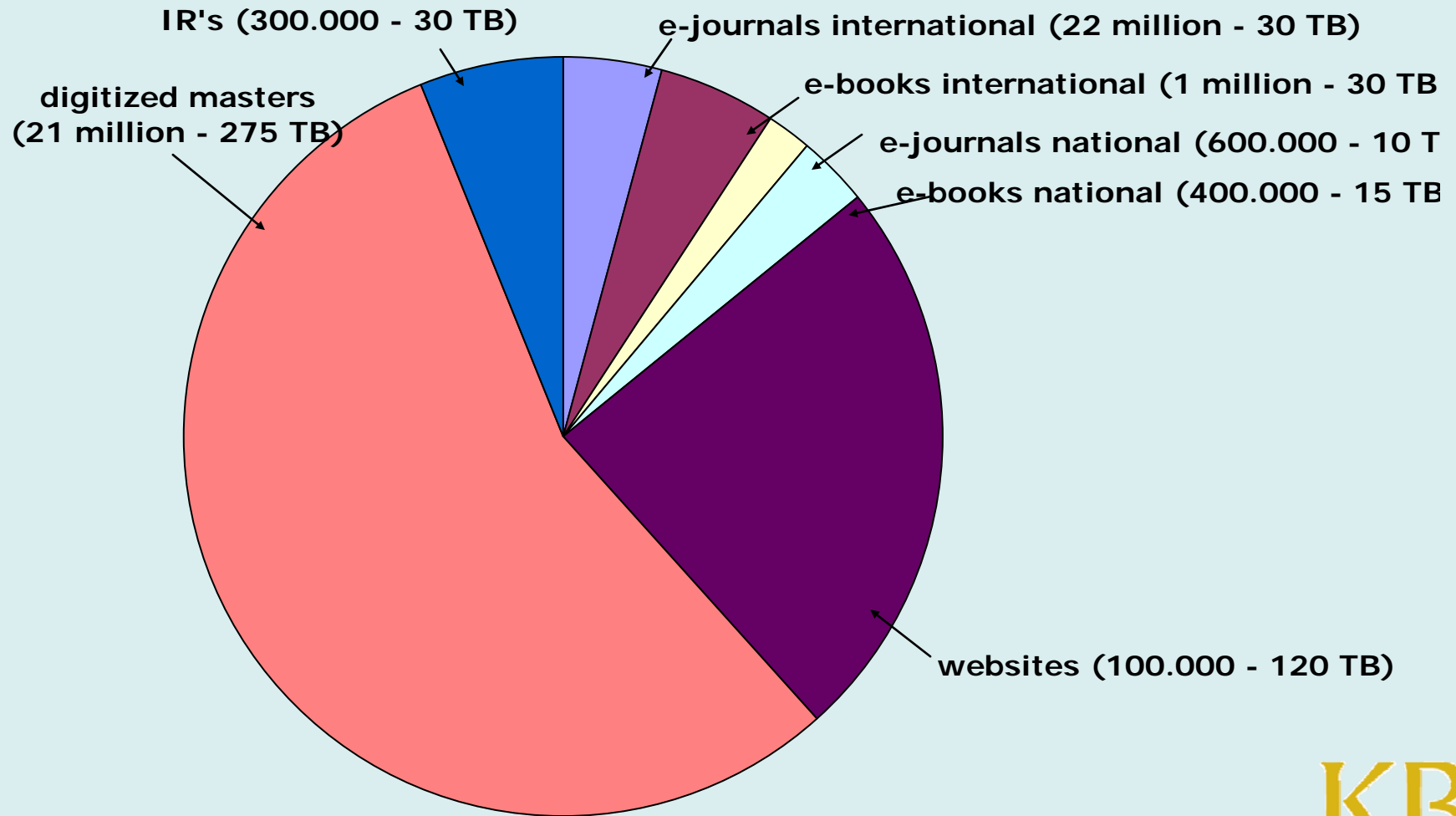


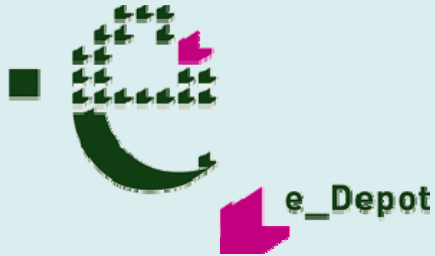
The KB e-Depot (2)

- Digital version of traditional depot
- Operational since 2003
- No legal deposit legislation
- Based on agreements with publishers
- International focus
- International scope of scientific output
- Safe Places Network



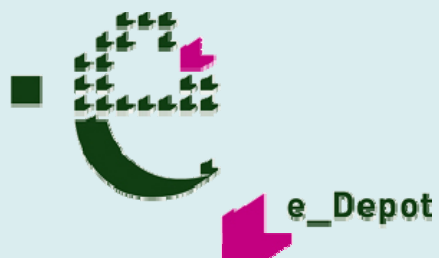
What is in it (in 2012)?





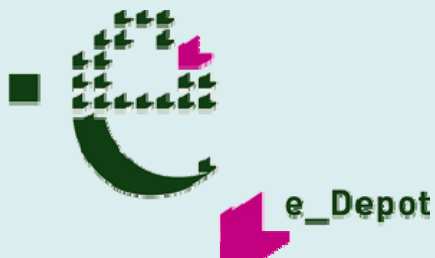
What does it cost?

- Initial costs for development (2000 – 2003)
- Annual costs, of which
 - Operational Staffing: 30 %
 - Project staffing & development: 25 %
 - Maintenance + hard- software licenses: 25 %
 - Storage: 20 %
- How to calculate
 - Preservation management
 - Preservation actions

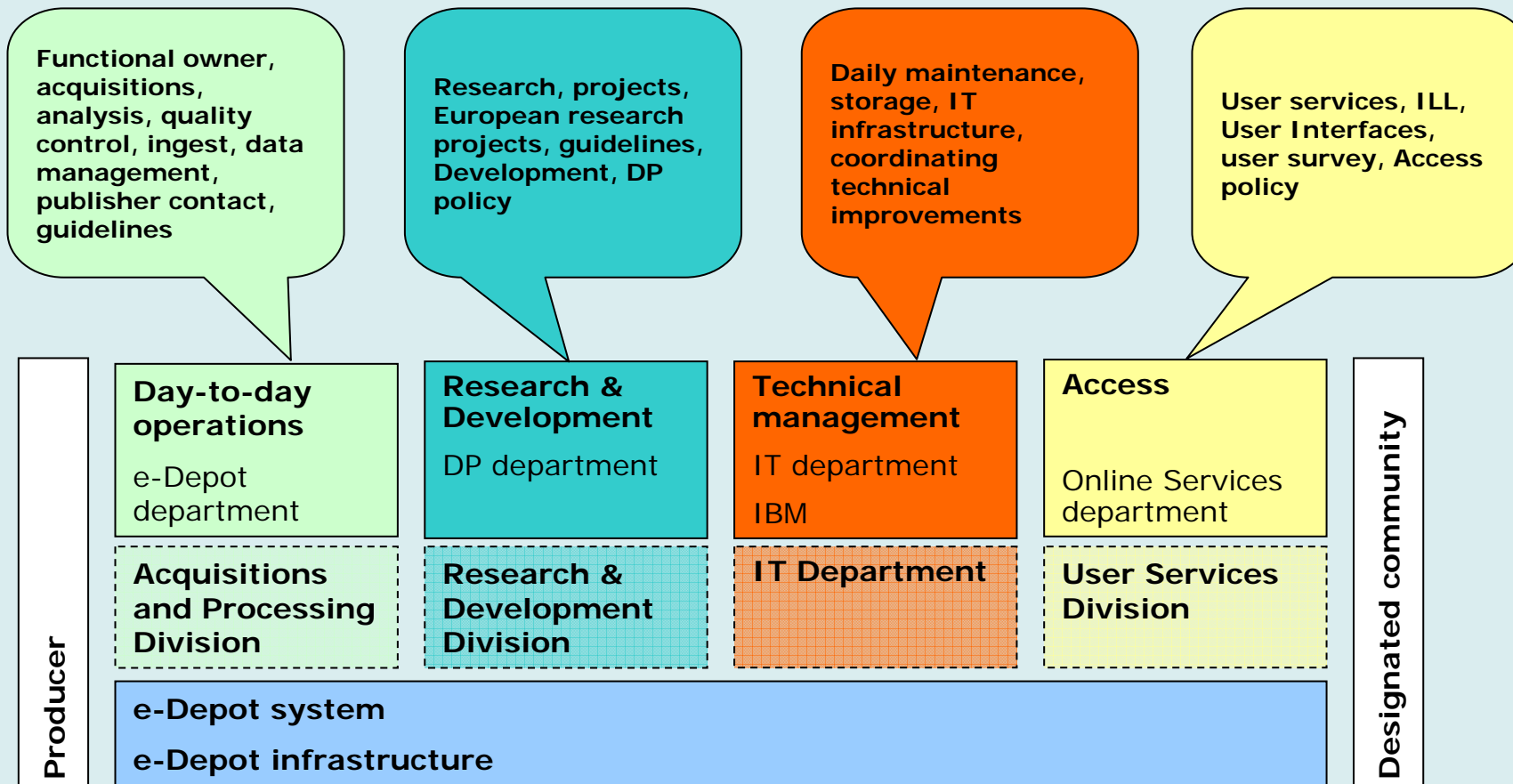


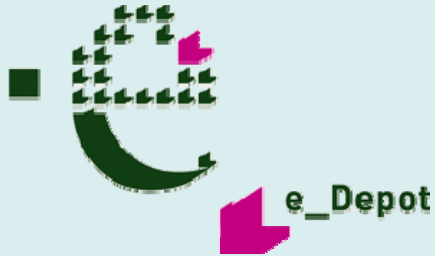
What does it cost?

costs		2010	2011	2012	2013
staffing	operations (8 fte)	450	500	500	500
	research (6 fte)	1500	1500	1500	1500
	maintenance (2 fte)	200	250	250	250
	k€	2.150	2.250	2.250	2.250
materiel	Hosting & backup	1500	1500	1500	1500
	LTP-system licence	1000	1000	1000	1000
	Application development	1000	1000	1000	1000
	Preservation actions	250	250	250	250
	k€	3.750	3.750	3.750	3.750
Total amount	k€	5.900	6.000	6.000	6.000



DP is not only a technical issue





Conclusions

- Preservation is not just storage and technique
- It demands for long-term organizational commitment
- It requires continuous research
- It requires substantial investments in infrastructure and up-to-date expertise and skills
- It brings organizational changes
- It asks for constant rethinking: next generation LTP solution
- Running a DP service has great impact on the organization
- But it has (organizational) benefits

Discussion and Questions?

Keeping Research Data Safe

JISC Research Data Digital Preservation Costs Study

Overview

- Aim – investigate costs, develop model and recommendations
- Method – detailed analysis of 2 cost models (LIFE & NASA CET) in combination with OAIS and UK Research TRAC;
- Plus literature review; 12 interviews; 4 case studies.

What have we Produced?

- A cost framework consisting of:
 - activity model in 3 parts: pre-archive, archive, support services
 - Key cost variables divided into economic adjustments and service adjustments
 - Resources template for Transparent Costing (TRAC)
 - Used in combination to generate cost/charging models
- 4 detailed case studies (ADS, Cambridge, KCL, Southampton)
- Data from other services.

Findings

Institutional Repository (e-publications):	Staff	Equipment (capital depreciated over 3 years)
Annual recurrent costs	1 FTE	£1,300 pa

Federated Institutional Repository (data): Annual recurrent costs	Staff	Equipment (capital depreciated over 3 years)
Cambridge	4 FTE	£58,764 pa
KCL	2.5 FTE	£27,546 pa

Findings

- **Timing.** costs c. 333 euros for the creation of a batch of 1000 records. Once 10 years have passed since creation it may cost 10,000 euros to ‘repair’ a batch of 1000 records with badly created metadata (Digitale Bewaring Project)
- **Efficiency Curve effects** – start-up to operational
- **Economy of scale effects** – Accession rates of 10 or 60 collections - 600% increase in accessions will only increase costs by 325% (ULCC)
- **“First mover innovation”** – costs of being first to solve a problem and how to finance this.

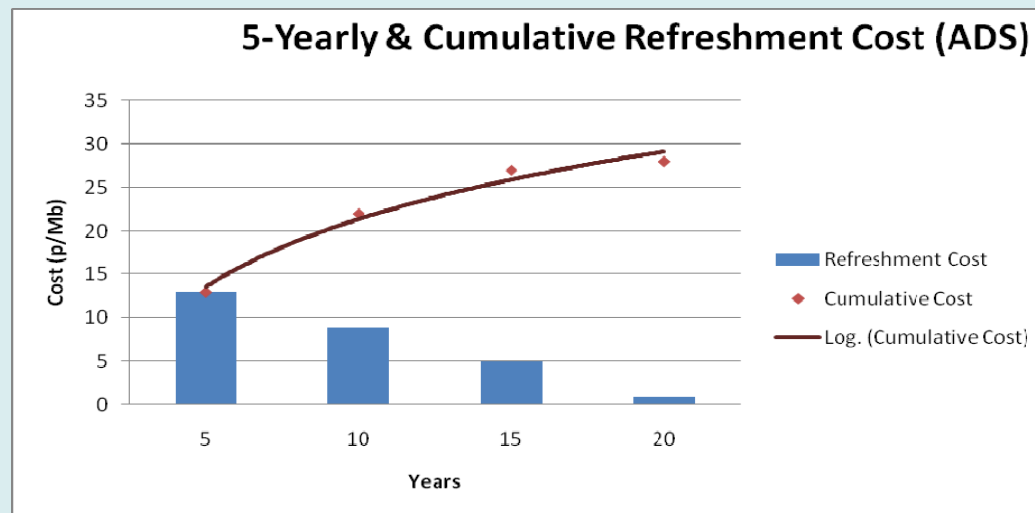
Findings

- National subject repositories costs (UKDA)

Acquisition and Ingest	Archival Storage and Preservation	Access
c. 42%	c. 23%	c. 35%

Findings

- ADS projection of long-term preservation costs
- Implications for sustainability via project charges/endowment



- Preservation interventions (file format migrations)
- Long-term storage costs
- Assumptions of archive growth (economies of scale)
- Assumptions on “first mover innovation”

Conclusions

- Data preservation costs not just formula of function costs
- Can illustrate effect of preservation choices on costs
- “first mover innovation” costs v operational costs
- Endowment archive funding model?
- Not last word on costs....recommendations for future work

Discussion and Questions?

The JISC Digital Preservation Policies Study

Overview

- The Challenge – too few digital preservation policies in institutions
- Study Aim – to support institutions wishing to create digital preservation policies and enhance their impact
- For UK HE/FE but of much wider relevance and interest

The Model Policy

- Amalgam of a range of policies
- Few clear technical strategies
- Mapped principle themes
- Eight generic clauses
- Exemplars, useful references, quotes
- Separate section for Guidance and Implementation
- Annotated bibliography

Conclusions

- A major business driver in all institutions has been harnessing digital content and electronic services for access
- **Long-term access and future benefit** will be heavily dependent on relevant digital preservation policies being put in place...
-and underpinned by **implementation procedures.**

Final Discussion and Questions

Further Information

David Rosenthal CNI Plenary see

<http://blog.dshr.org/2009/04/spring-cni-plenary-remix.html>

“Keeping Research Data Safe” Final report and Executive Summary at

<http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx>

“Digital Preservation Policies Study” Final Report and Appendices at

<http://www.jisc.ac.uk/publications/publications/jiscpolicyfinalreport.aspx>