



Koninklijke Bibliotheek
National library of the Netherlands



→ The International e-Depot

e-Journal Archiving at the national library of the Netherlands

Erik Oltmans

Head, Acquisitions & Cataloguing Division

National Library of the Netherlands

iPRES 2006, Cornell, October 10, 2006

Overview

1) Setting and Scope

- Koninklijke Bibliotheek
- History of the e-Depot
- System characteristics
- Scale

2) Archiving e-journals: The KB Policy

3) Digital Preservation Research



Setting and Scope

Governance, funding, organizational structure

- Koninklijke Bibliotheek
 - Medium-sized national library, founded in 1798
 - Financed by Ministry of Education, Culture and Science
 - Annual budget € 50 million, 270 fte
- Digital archiving and R&D Permanent Preservation:
 - € 1,1 million structurally allotted to staff, system maintenance
 - € 1,2 million permanently dedicated to research
 - Research funding will increase
- Digital archiving and preservation embedded in organization
- Departments: e-Depot, Digital Preservation, IT

Director General

Corporate Management

Communication

Planning & Control

Human Resources

Finance

Building & Facilities

Information
Technology

Collections & Services

Acquisitions & Cataloguing

Acquisitions

Cataloguing &
Metadata

e-Depot

Information & Collections

Collection
Management

Collections Care

Expert Services

User Services

Front Desk

Storage & Control

Innovation

Product & Service
Development

National
Programmes

Digital
Preservation

History of the KB e-Depot

- 1993: Decision to start e-Depot
- 1995: Experiments with Elsevier, Dutch Publishers Association
- 2002: Landmark archiving agreement with Elsevier
- More archiving agreements:
 - Kluwer Academic Publishers, BioMed Central (2003)
 - Blackwell, Oxford University Press, Taylor & Francis (2004)
 - Sage, Springer, and Brill Academic Publishers (2005)
 - General agreement Dutch Publishers Association (updated 2005)

The e-Depot system

- Experiments with AT&T, IBM from 1995-2002
- DIAS: OAIS-compliant, developed with IBM, implemented in 2002
- Integrated with other library modules
- Functionalities:
 - Ingest of e-journals, e-books, digitized publications, and CD-roms
 - Authentic publications are archived, standard formats (PDF, XML)
 - Automatic validation (checksum, JHOVE), error handling
 - Metadata conversion
 - Batch delivery
- KB aims at continuous auditing, resulting in certification

Scale

- Volume
 - 8 million e-publications currently
 - 10 million e-publications eventually (based on *current* agreements)
 - Expected to be fully implemented this year
- Size
 - 1 e-publication equals 1 Mb on average
 - 1 Terabyte for every 1 million publications
- Capacity
 - 5,000 – 50,000 e-publications ingested per day



Archiving e-Journals: the KB policy

The KB Policy: identifying risks

No obvious guardian for international publications

- Who ensures long-term accessibility of international e-journals?
- Most current e-Journals lack a natural a fatherland
- Geographic model not sufficient for systematic approach
- Risk of information getting lost forever

Disrupted Access

- Calamities damaging publisher's services
- Withdrawal of publications, journals (commercial motivation)
- No successor

Technological Obsolescence

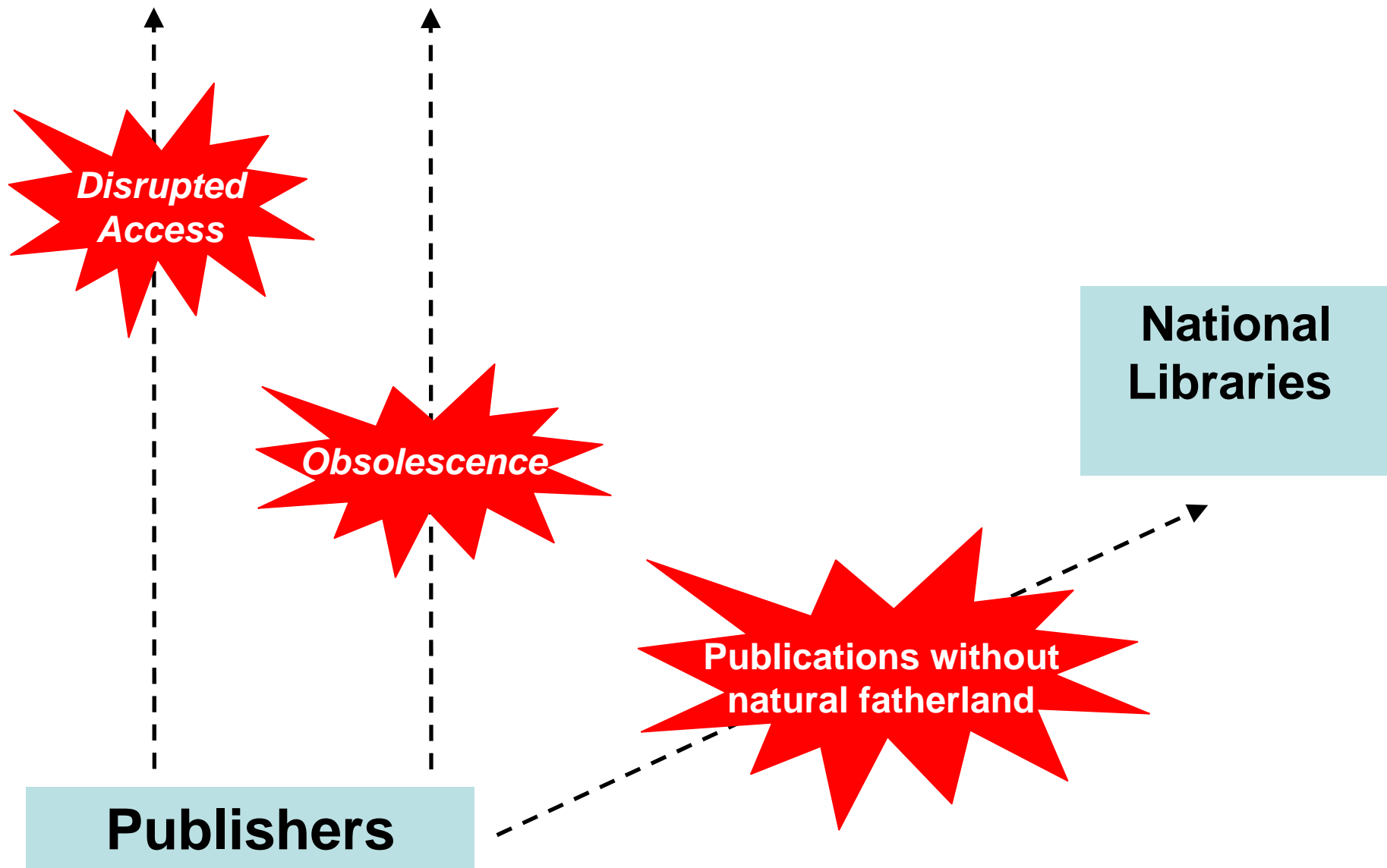
- Inaccessible file formats

The KB Policy: addressing risks

- Publishers are unlikely to deposit at large number of places
- Spread the (geopolitical) risk and identify trustworthy partners
- Collaboration and coordination
- Centers of expertise may emerge: the Safe Places Network

- Safe Places Network secures systematic, coordinated preservation
- In case of loss, libraries know where to go
- Permanent archiving calls for
 - Substantial investment
 - Permanent commitment
 - Continuous efforts and permanent R&D

Research Libraries, End Users



*Disrupted
Access*

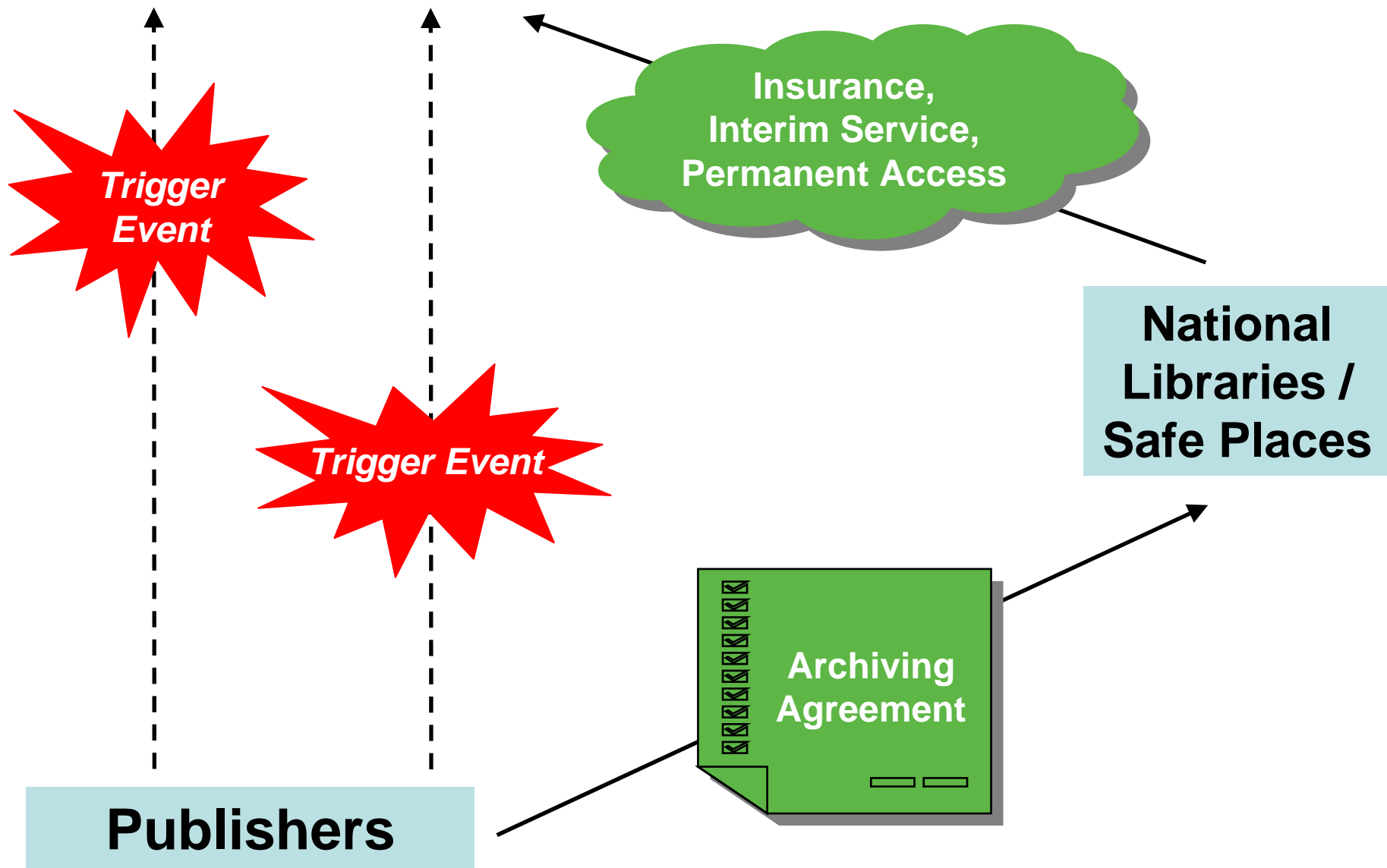
Obsolescence

*Publications without
natural fatherland*

Publishers

**National
Libraries**

Research Libraries, End Users



Publishers

Trigger Event

Trigger Event

Insurance,
Interim Service,
Permanent Access

**National
Libraries /
Safe Places**

Archiving
Agreement

Who are we serving? Designated Community

Publishers

- The e-Depot is a safe place for the scientific community
- Additional service for publisher's customers
- e-Depot access does not compete with publisher-provided access
- Retrieval, access, printing, downloading *for private use only*
- Systematic reproduction is not allowed
- COUNTER compliant

Libraries, end users

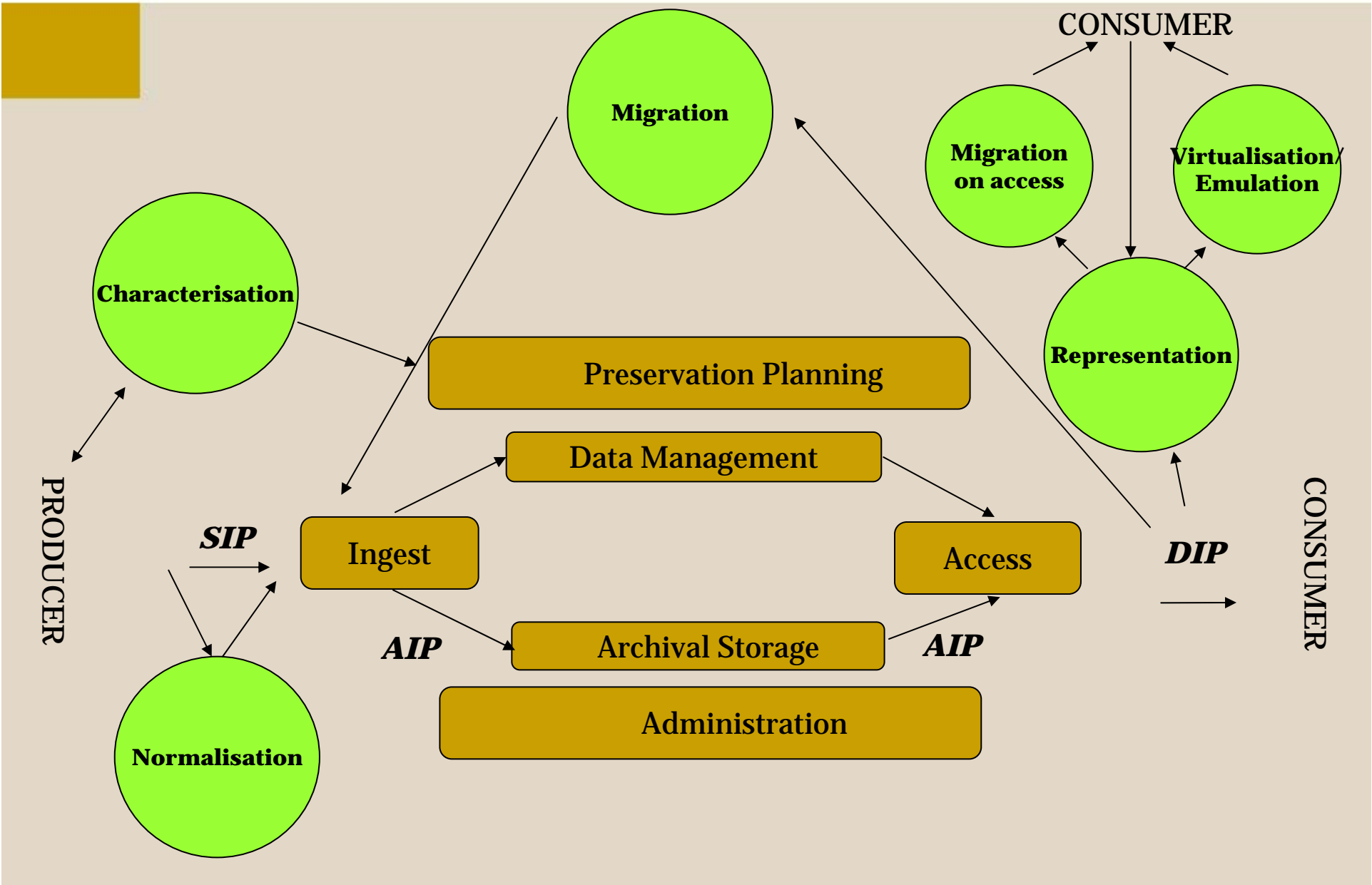
- The e-Depot is a safe place for the scientific community
- On site access for KB visitors, ILL supply within the Netherlands
- Exceptional use: Trigger events cause broader access
- If a publisher stops making journals available: Open access to all
- In case publishers cannot meet obligations: Part of interim service



Digital Preservation Research

e-Journal archiving

- Preservation planning is not a separate module, but requires action at all stages of the archiving process
- The file format concept is not enough: we need to distinguish between collections and need to know about document settings
- Research on PDFs: how can we influence publishers to create durable PDFs without restricting them to PDF-A?
- Font sets: should be embedded
- Version Control: how do we deal with updates?
- Full range of digital preservation: both migration and emulation



Next steps towards operational solutions

- Characterization
 - Identification is the first step towards permanent access
 - Implement, add-ons to JHOVE, implement other characterization tools
 - New requirements for the Preservation Manager
 - Interoperability with external file format registries
 - More granularity: collections, significant properties, complex objects
- Migration, operationalization in three steps:
 - Migration on ingest: normalization (operational module early 2007)
 - Batch-migration, migration on access
- Emulation: a modular approach
 - New design of a modular emulator, with Dutch National Archive (2007)
- Virtualization
 - Existing solutions have proven very useful
 - Integrate virtualization techniques in representation process

Cooperation in Digital Preservation

- Planets: European Project to develop Preservation Planning tools
- KB and Portico are investigating further ways to collaborate
- DIAS: enhance the system, international user group
- IBM: new research agenda, implementing preservation functionality
- The Netherlands: set up National Digital Preservation Coalition

Summary

- E-Journals: range of possible disruptions
- E-Journals must be preserved by institutions who take responsibilities;
- Systematic and coordinated by means of Safe Place Network
- KB is committed to be part of the Safe Places Network: substantial R&D, permanent commitment
- KB seeks more international collaboration, e-Depot to be certified



The KB e-Depot

Contact:

erik.oltmans@kb.nl

www.kb.nl/e-Depot