

→ The International e-Depot

Digital Archiving at the national library of the Netherlands

Erik Oltmans

Head, Acquisitions & Cataloguing Division

National Library of the Netherlands

Deutsche Forschungsgemeinschaft

Berlin, June 11th, 2007

Overview

1) Short introduction of the e-Depot

- Koninklijke Bibliotheek
- History of the e-Depot
- The KB Digital Archiving Policy
- Designated Community

2) The e-Depot system: system characteristics and technical basis

- Functionalities
- System Characteristics
- Technical Basis
- Authenticity, Integrity, Readability

Setting & Scope: The KB



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 increases every year
- 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 Care

Research Collection

Special Collections

User Services

Online Services

Front Desk

Storage & Control

Innovation

Product & Service
Development

(Inter) 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)
 - IOS Press (2007)

The KB Policy: identifying & addressing risks

1. No obvious guardian for international publications
 2. Disrupted Access
 3. Technological Obsolescence
- Publishers are unlikely to deposit at large number of places
 - Spread the (geopolitical) risk and identify trustworthy partners
 - Centers of expertise may emerge: the Safe Places Network
 - Safe Places Network secures systematic, coordinated preservation
 - Permanent archiving calls for
 - Substantial investments and permanent commitment
 - Continuous efforts and permanent R&D

Who are we serving? **Designated Community**

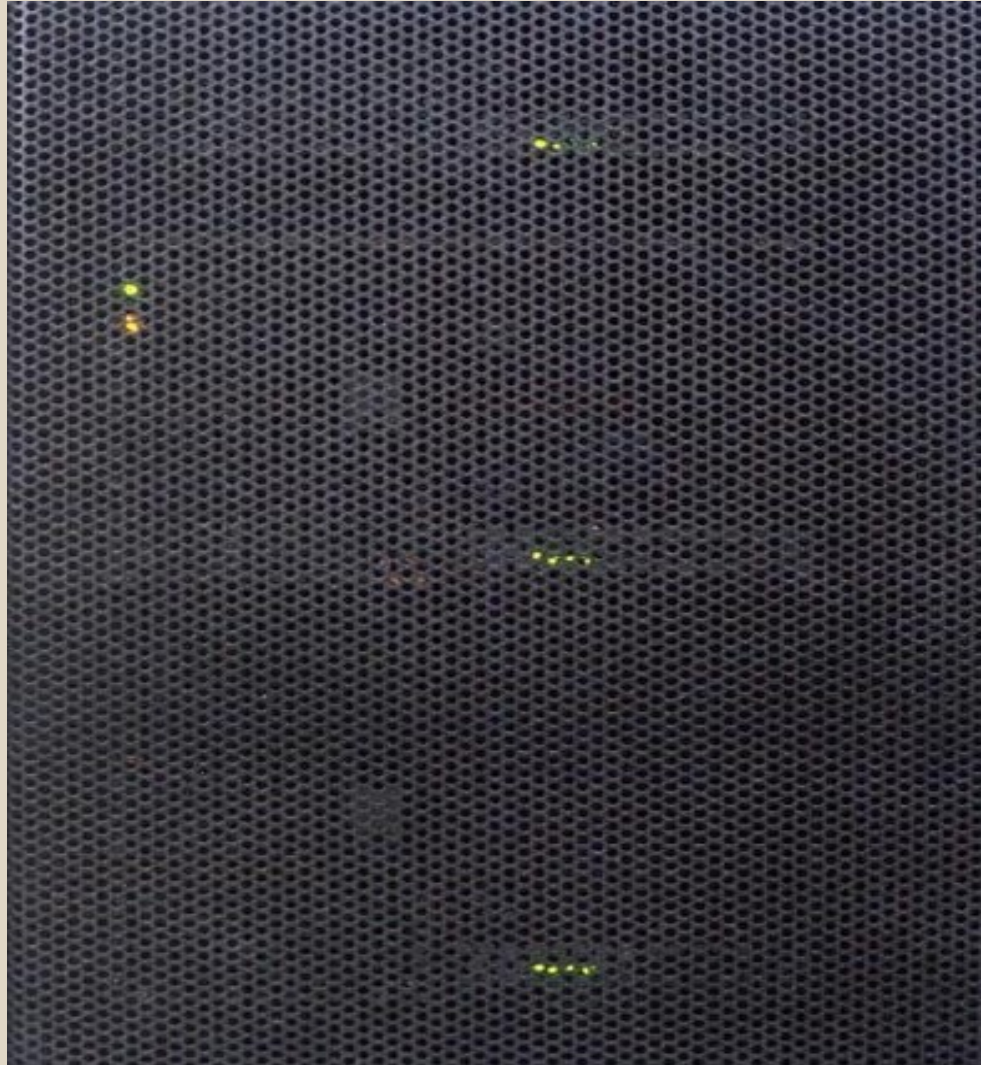
Libraries, end users

- 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

Publishers

- 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

The e-Depot System



Scale: what is in it?

Volume

- 11 million e-publications (PDFs, based on current agreements)
- 10 million digitized publications to be expected (TIFFs)
- Thousands of web sites

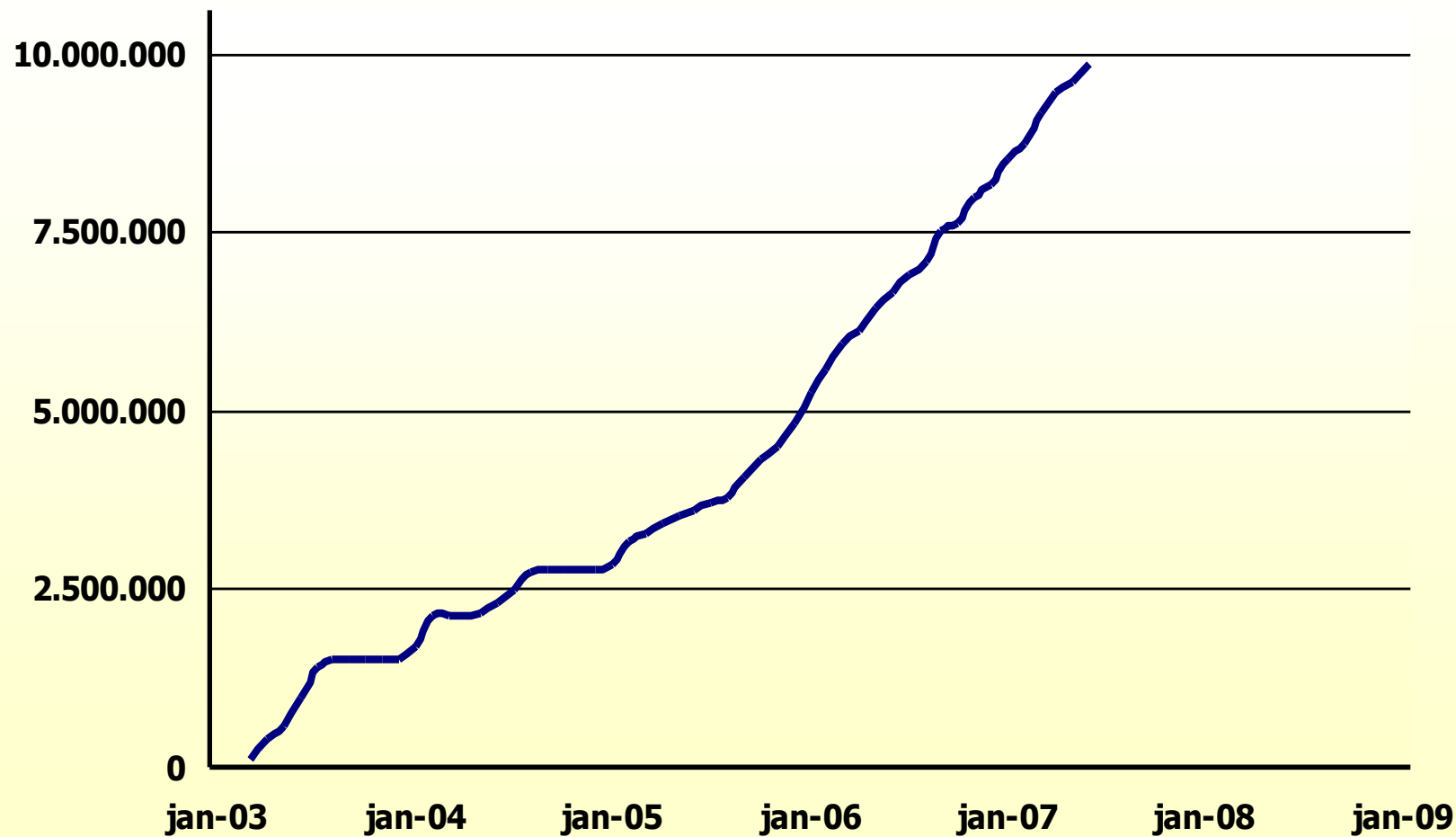
Size

- 1 PDF equals 1 Mb on average: 11 terabyte e-publications
- 1 TIFF equals 90 Mb on average: 900 terabytes TIFFs
- 1 Web site equals 40 Mb on average: 120 terabytes web sites

Capacity

- e-publications: 50,000 objects ingested per day
- TIFFs: 10.000 per day (to be expected)

Total volume publications ingested since 17-3-2007



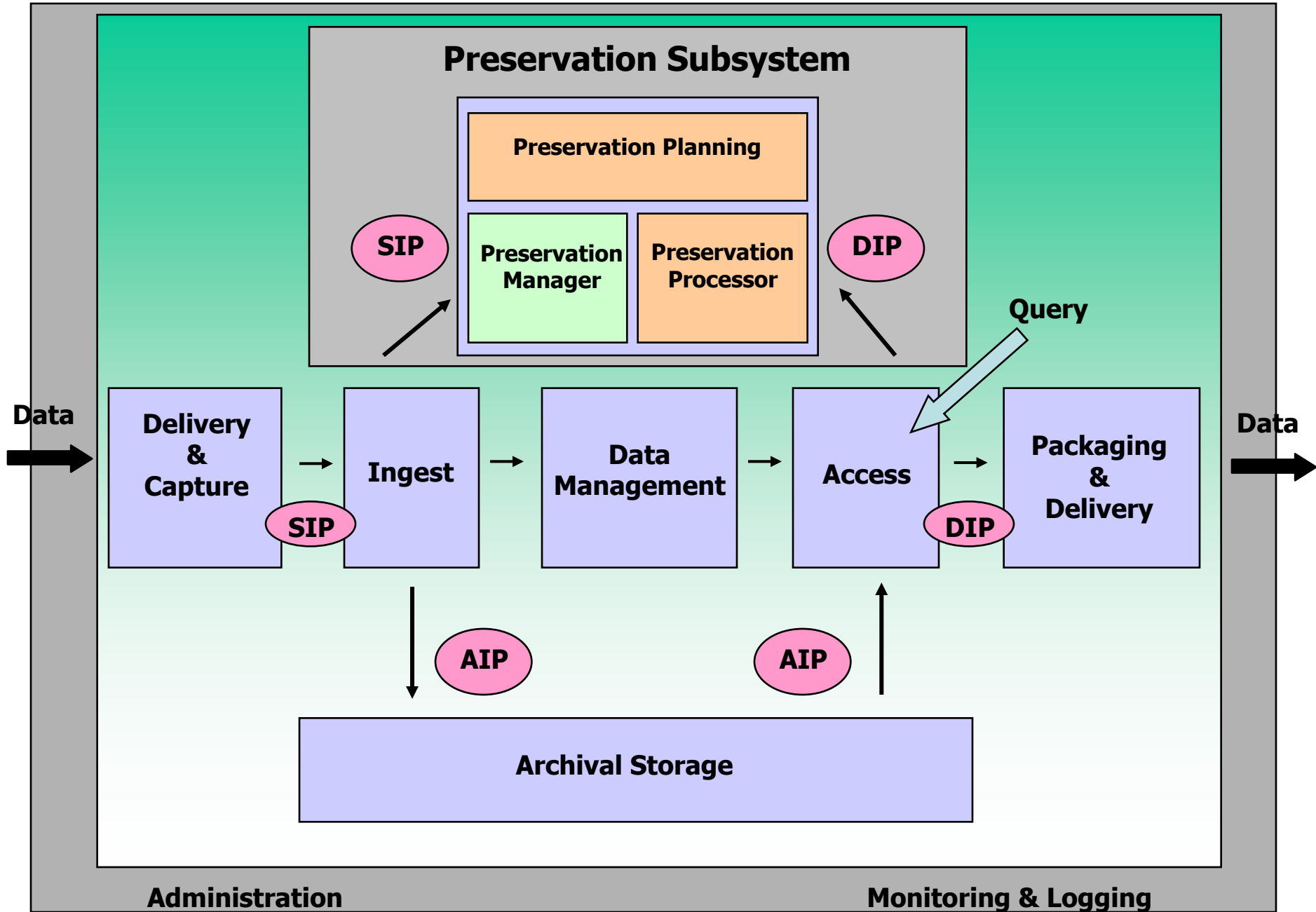
The e-Depot system

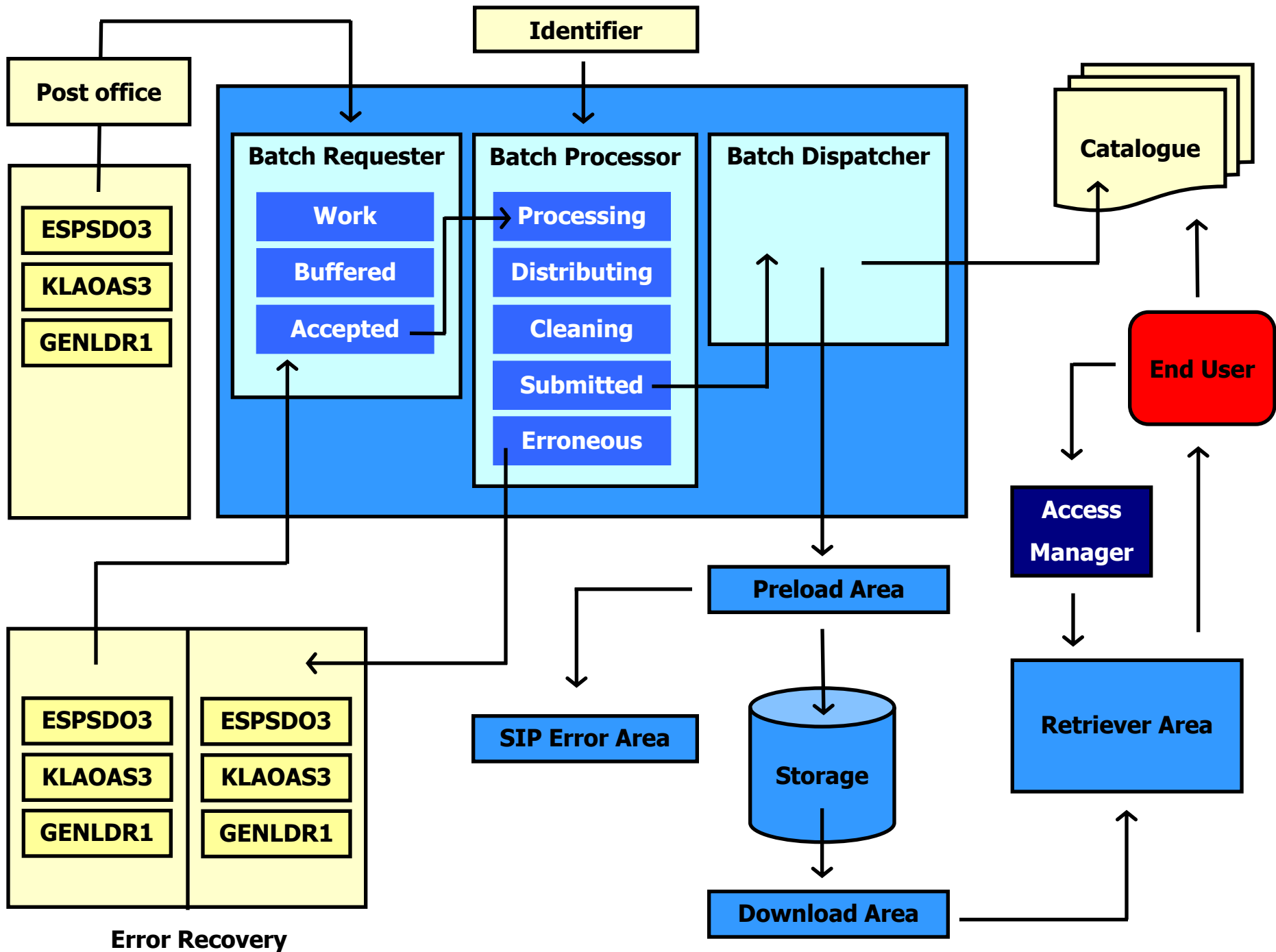
- Experiments with AT&T, IBM from 1995-2002
- IBM implemented system in 2002
- Integrated with other library modules
- Functionalities:
 - Ingest of e-journals, e-books, digitized publications, and CD-roms
 - Authentic publications archived, standard formats (PDF, XML)
 - Automatic validation (checksum, JHOVE), error handling
 - Metadata conversion
 - Batch delivery
 - Web archiving

Technical characteristics

- Digital Information Archiving System, developed by IBM
- (Also used by kopal ; International UG meets twice a year)
- **System properties:**
 - Storage capacity 25 TB, to be extended to 1,5 PB
 - Operating system AIX 5.3
 - Software: Tivoli Storage Manager, Access Manager, DB2, Content manager, Web Sphere, Java
 - Hardware: IBM P570, P520, 4 Gb fibre channel switches

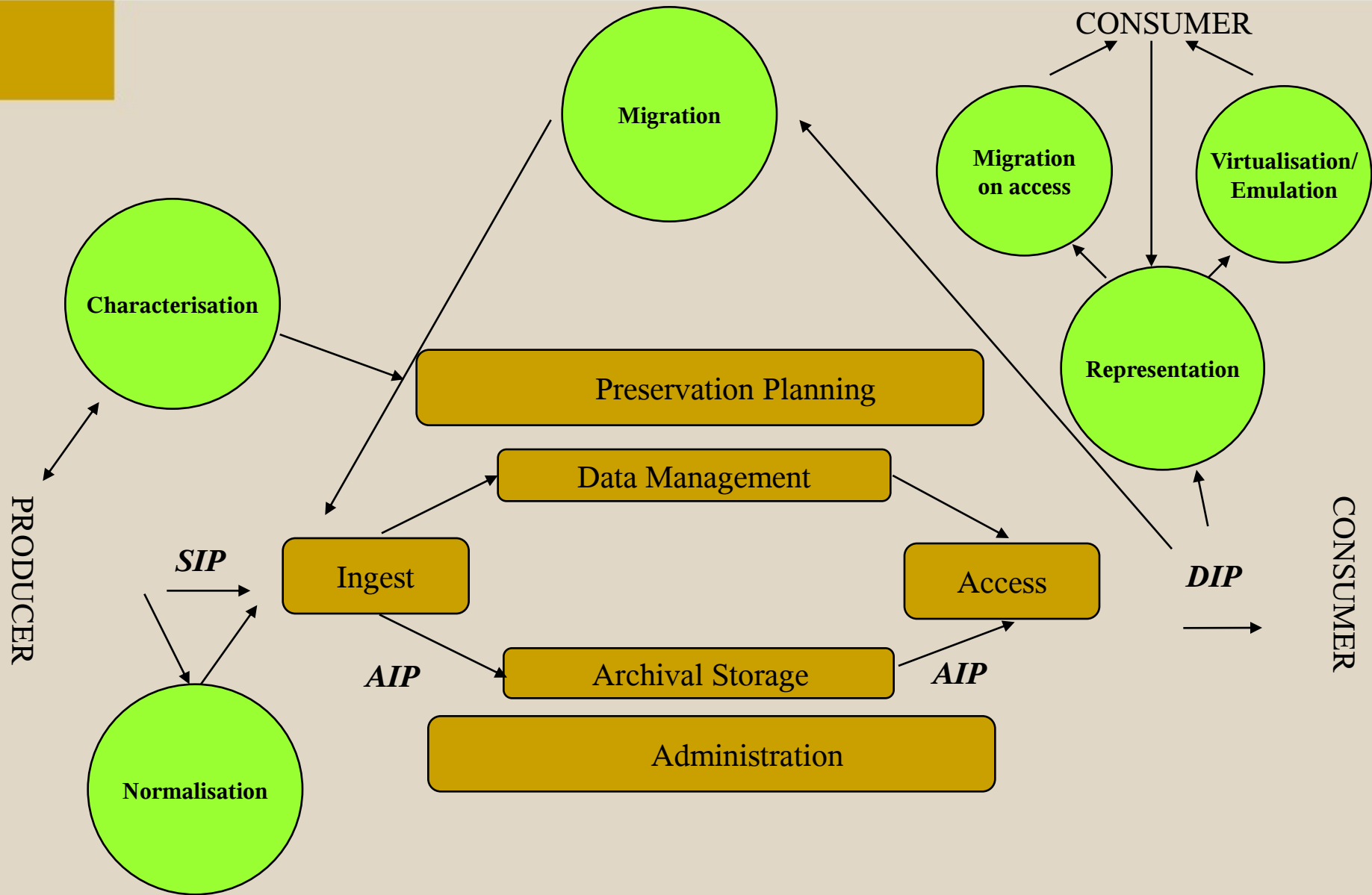
D I A S





Digital Preservation Strategies

- Emulation – so as to preserve look & feel and functionalities
 - Universal Virtual Computer
 - Modular Emulation
 - Virtualization
- Migration – so as to preserve data and functionalities
 - Batch Migration
 - Migration on request
- Normalization – so as to preserve the data
 - No implementation yet, planned though



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 2008)
 - Batch-migration, migration on access
- Emulation: a modular approach
 - New design of a modular emulator, with Dutch National Archive (2007)

More technical aspects

- Existence of objects over time
 - Media Refreshment: moving files to newer media carriers
 - Administration, regular checks
- Readability: Migration to actual formats, emulation
- Safety: intrusion tests; risk analysis by KPMG
- Testing of new software: Reference set, *The daily trip*
- Integrity checks: Checksums, Validation (JHOVE, Droid)
- Authenticity: Checksums, unique IDs, comparison with publishers
- Organizational aspects: permanent commitment!

Certification

- The e-Depot is a solid digital repository with well-planned processes and procedures underpinning its operations and technical infrastructure. As a core piece of a national library's infrastructure and service delivery, the KB's e-Depot is well-funded by a government ministry and the commitment to continue this funding is evident. [...]
- The KB appears to provide very good stewardship of the valuable research resources in its care. [...]
- Stakeholders in the KB electronic journal archives and the national image archives should feel confident about the state of the organization, as well as the processes, procedures, technologies, and technical infrastructure employed by the organization.
- *KB Final Report, March 21, 2007, produced as part of a test of the RLG/NARA Draft Audit Checklist for the Certification of Trustworthy Digital Repositories.*

Summary

- KB e-Depot focuses on e-journals, e-books, TIFFs, web sites
- KB aims for collaboration in context of Safe Places Network
- Technical basis is OAIS-compliant IBM product: DIAS
- Digital Preservation strategy: both migration and emulation
- Technical soundness, organizational commitment, certification



Contact:

erik.oltmans@kb.nl

www.kb.nl/e-Depot

