Things Mr. Gates will not tell you about the Digital World

(because he doesn't understand them)

James Currall
University of Glasgow

ICA
Vienna
DELOS Workshop on Preservation in Digital Libraries
Tuesday, 24th August 2004 Room A353

Thought for the day

"Stewardship is easy and inexpensive to claim but expensive and difficult to honor."

Clifford Lynch, 2002

Some Questions

- Is it safe for Digital Assets to be left on desktop machines for management by individuals?
- Do we need special systems to preserve Digital Assets or are normal file systems satisfactory?
- Should important Digital Assets be stored on-line or off-line?
- How much management do Digital Assets require?

Some More Questions

- At what stage do actions need to be take to preserve Digital Assets?
- How can we afford Digital Repositories?
- To what extent are Digital Repositories a technical problem?
- What are the major issues in setting up and maintaining a Digital Repository?

The Plan

- Introduction
- Types of Digital Asset store
- Software Approaches
- Some Important Issues
- DAM a Strategic Matter?
- Some Examples
- Some Challenges

... and the Questions Revisited

Fundamentals

- What is one?
- Why have one?
- What does it do?

What is a Repository?

- Somewhere safe to put stuff
- An efficient asset store
- An infrastructure to manage assets
- A mechanism for finding and distributing digital assets
- A set of services for the management and dissemination of digital assets

Why have a Repository?

- Risk of loss from:
  - degradation of media
  - technical obsolescence
  - accidental deletion
  - malicious damage
  - poor management
- Value of digital assets to you and your organisation
- Business continuity
- Legislative requirements
- The present situation is unsustainable

What purpose might a Repository serve?

- Management of Current Records
- Management of Archival Material
- A Publication Medium
  - pre-prints
  - e-prints
  - reports
  - etc.

Types of Digital Asset store

- Unmanaged collection of digital objects
- Well managed hierarchical file store
- Specific systems
  - off-line
    - live computer system
    - static non-active media (CDRs, etc.)
  - on-line
    - primary source
    - secondary source
  - multi-mode
- Models
  - Open Archive Information System (OAIS)
    - SIP >> AIP >> DIP
    - management processes

Software Approaches

- Just the file system
- Just a metadata database
- File system + metadata database
- Public Domain Tools
- Multi-level, multi-machine processing systems (see examples later)

Public Domain Tools

DSPACE

Developed by MIT and HP, emphasises different 'communities' and appropriate workflow, uses DC metadata and permits OAI metadata harvesting.

FEDORA

Developed by Virginia (implementation) and Cornell (architecture), based on a strong model separating management of the bitstream from rendering for the user, permits OAI metadata harvesting.

LOCKSS

Developed by Stanford and SUN to 'simulate' the preservation of printed materials in libraries, works well for syndicated materials such as e-journals.

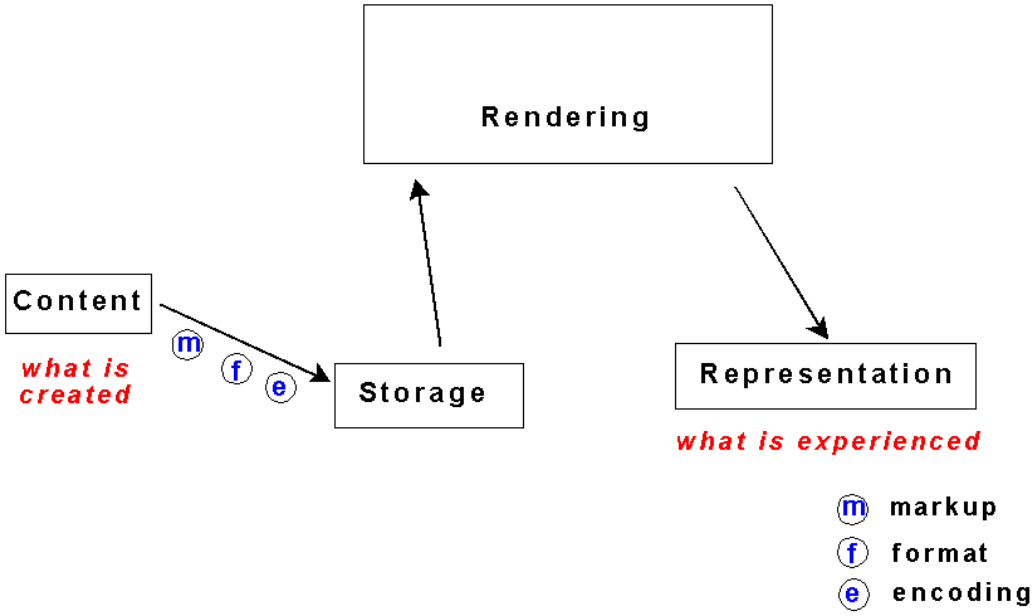
What Bill doesn't understand!

- Representation
- Metadata
- Trust
- Real Security

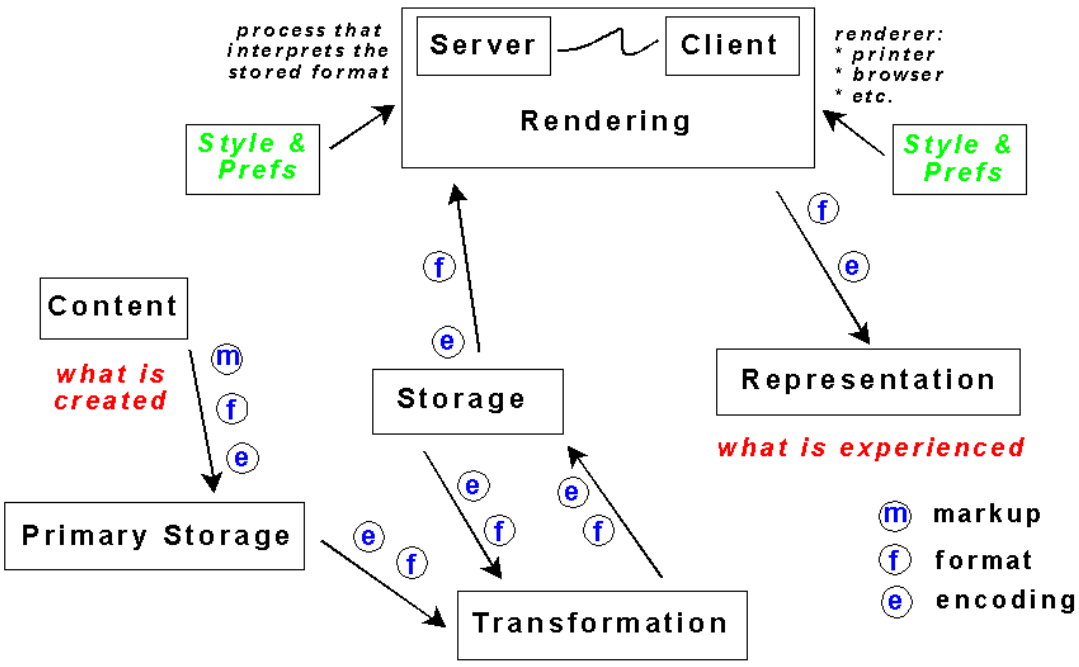
Representation

- What do we experience of digital objects?
- File Formats in which digital objects are stored
- Will the deposited bitstream be sufficient?

Mediated Bitstreams



Mediated Bitstreams



File Formats

Choices

- Standards based or proprietary?
- Binary or 'character'?
- Unicompressed or compressed?

File Format Registries

- Technical details
- Compatibilities
- Tools to handle

Will the deposited bitstream be sufficient?

- Deposited form - probably proprietary
- Canonical form - standards-based
- Transformation - from canonical form to user representation
- Migration - to deal with format obsolescence
- Emulation - to deal with format obsolescence
- Behaviours - to deal with different media and audiences, and as technology develops

Metadata

- Types
- Specialised Standards
- Metadata Harvesting

Types

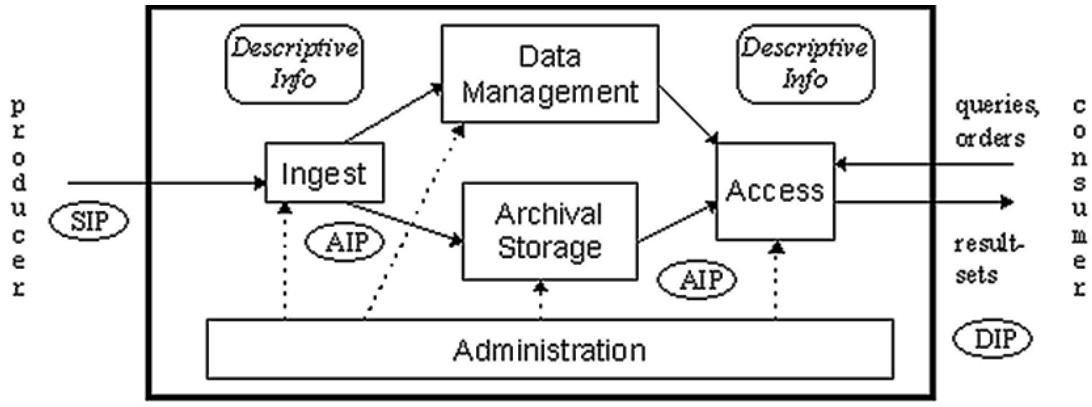
- Administrative
- Preservation - OAIS
- Structural
  - General packaging - METS
  - Learning Object packaging - IMS - LOM
- Descriptive - Dublin Core
- Behaviours
- Persistent Identifiers

Administrative

- Technical characteristics
- Source object (what was digitised)
- Provenance (history of repository operations)
- Rights management

OAIS

Presents views of the archive (and archival process) at different levels



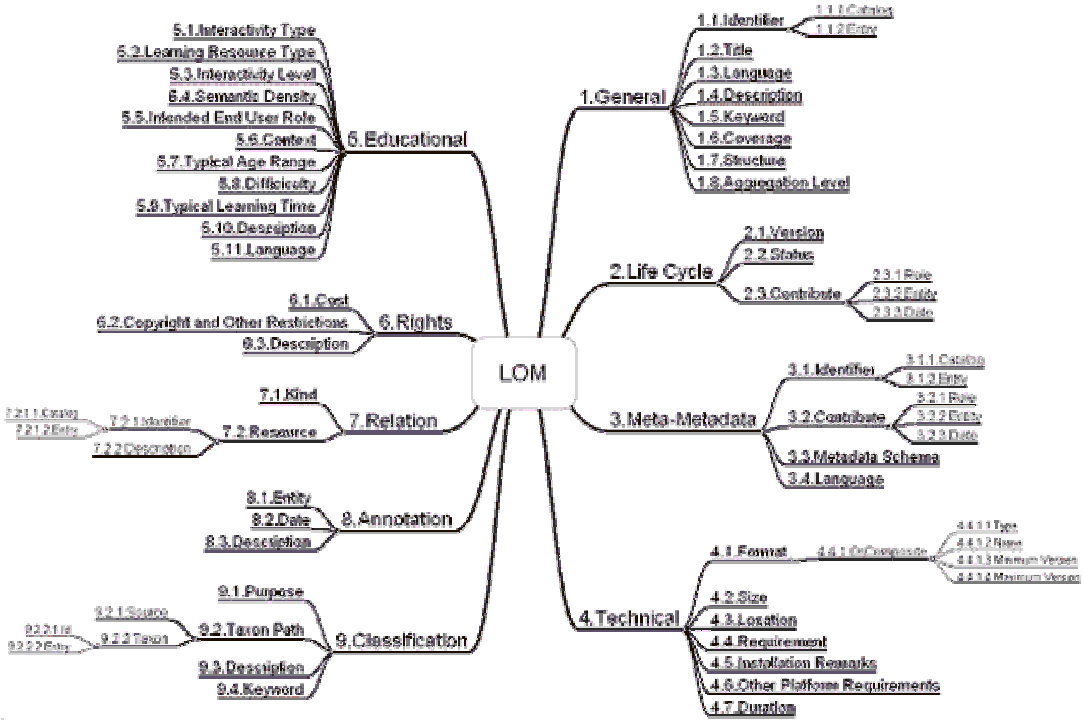
Metadata Encoding and Transmission Standard

A METS 'document' consists of 7 elements:

- METS Header
- Descriptive Metadata
- Administrative Metadata
- File Section
- Structural Map
- Structural Links
- Behavior

IMS - Learning Object Metadata

Learning Resource Meta-data Specification



Dublin Core

A standard for metadata - 15 key metadata elements.

- identifier
- format
- date.publication
- relation.isFormatOf
- title
- creator.author
- contributor.editor
- subject
- description
- publisher
- rights
- date.creation
- date.modification
- type
- language
- source

Specialised Standards

Archive

- ISAD(G), ISAAR(CPF), EAD

Library

- MARC21, MARCXML, etc.

Museum

- SPECTRUM, CDWA

Open Archives Initiative

- Qualified Dublin Core metadata
- Simple Harvesting protocol
- Harvesting service providers (e.g. OAIster)
- Register repository with service providers
- Allows cross-searching of many repositories

Trust

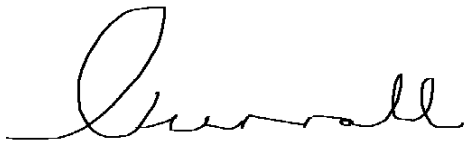
- Is the item what the depositor says it is?
- Is the item what was deposited?
- What operations have been performed on it since?
- Capturing such assertions at the time of deposit and at key points later
- Allowing the user to validate or test the assertions
- The role of Digital Signatures
- The role of Trusted Repositories

Digital Signatures

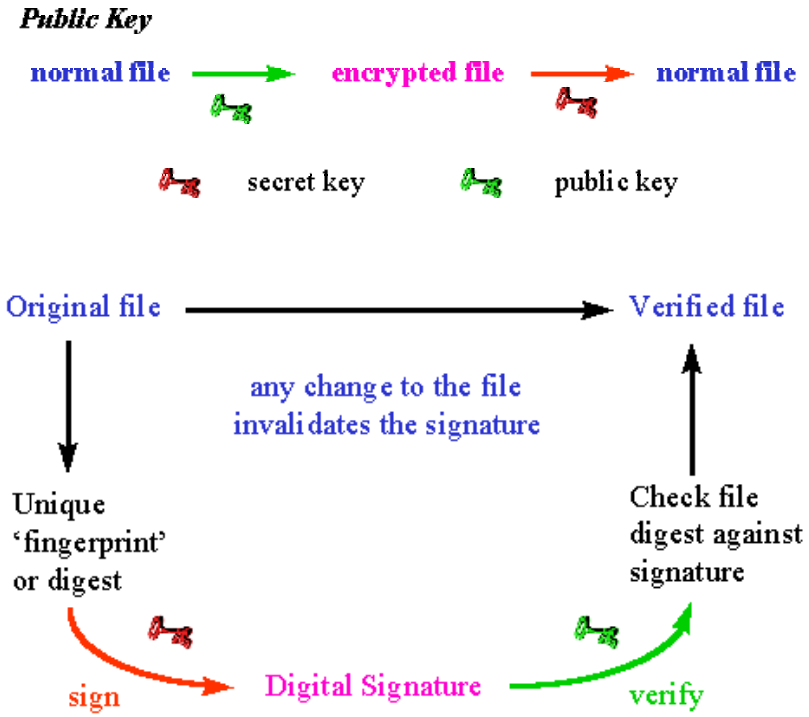
```
-----BEGIN PGP SIGNATURE-----
Version: PGPfreeware 6.5.8 for non-commercial use <http://www.pgp.com>

iQA/AwUAPr5HPSpYUiSnnmBwEQKu4gCgxITOZp7OqhB+MniiYGlfzeBW7QAAnO6r
rtGenpb4i8YmEsfzkCvkyPkL
=biy8
-----END PGP SIGNATURE-----
```

not



Digital Signatures



Trusted Repositories

- What attributes does one have?
- What responsibilities rest with the TR and what remain with the content owning organisation?
- Are organisations ready and willing to 'out-source' this activity?
- Certification schemes (e.g. DINI)
- Are there 'certified' service providers?
- The role of National Libraries and Archives?

Security

Avoiding:

- accidental damage (through bitrot)
- deliberate damage (through malicious act)
- incidental damage (through anonymous hacking)

By:

- physical and virtual separation
- monitoring of objects
- backup
- environmental control

Providing:

- physical security
- environmental security
- virtual security

DAM a Strategic Matter?

- Organisational Strategies
- A DAM Strategy
- Processes

Organisational Strategies

- Business models that value digital assets
- Cultural change in relation to ownership and responsibility
- Buy-in from
  - management and decision makers
  - creators and support staff
- Acceptance that repositories are 'institutional' and not a game for enthusiastic amateurs
- Understanding that digital assets need to last longer than a PC and longer than one person's working life

A DAM Strategy

- Policies
- Workflows
- Metadata definition
- Metadata management
- Asset 'capture' processes
- File management
- Rights management
- Access mechanisms

Processes

- Establish standards for file formats
- Establish standards for storage media
- Identify content and accompanying metadata
- Set up secure storage
- Data migration/transformation
- Media refresh
- Policies and procedures

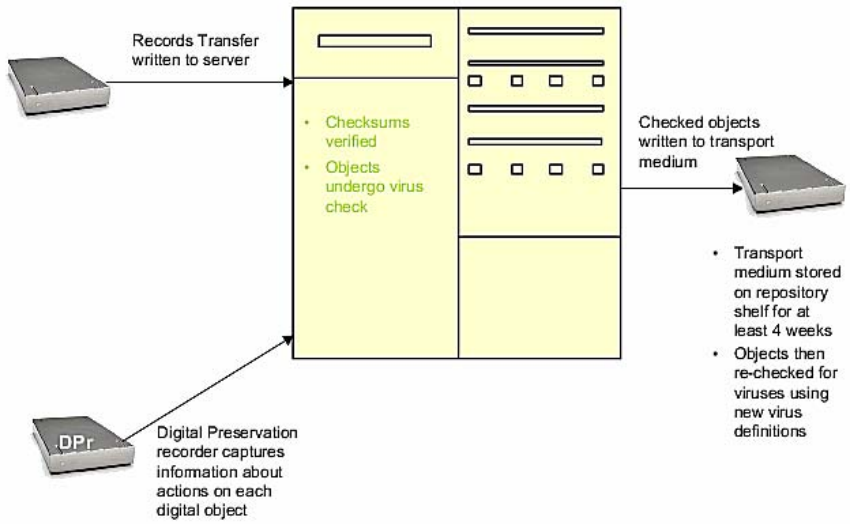
Some Examples

- National Archives of Australia
- Uppsala University Library
- An Abstract Model

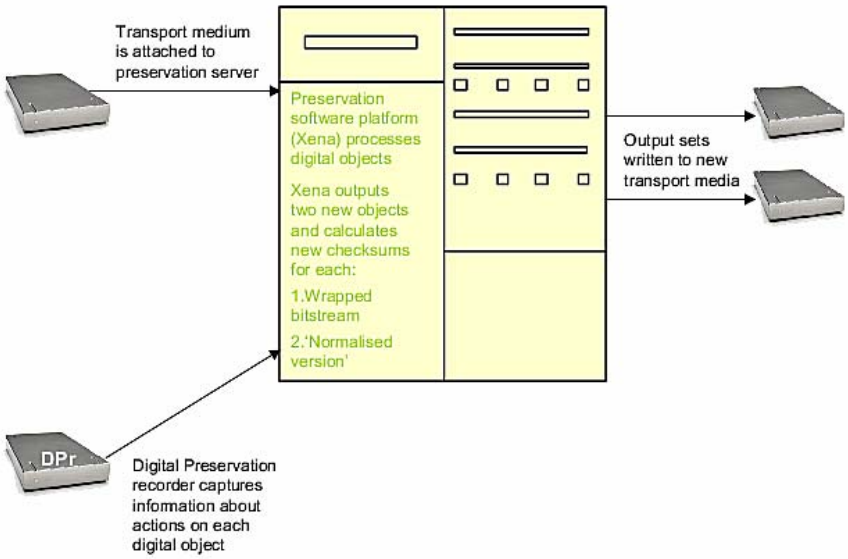
National Archives of Australia

- 3 separate components
  a) Quarantine Server
  b) Preservation Server
  c) Digital Repository
- Components physically separated from each other and all other NAA networks
- Access to hardware restricted to digital preservation staff

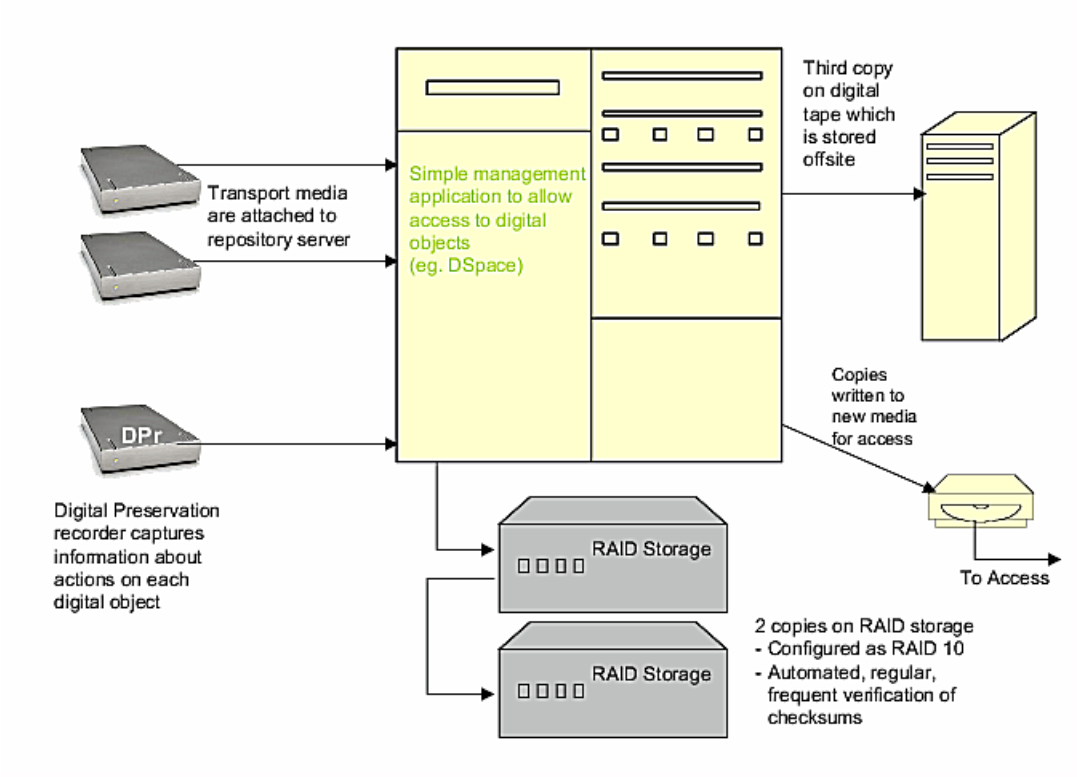
Quarantine Server



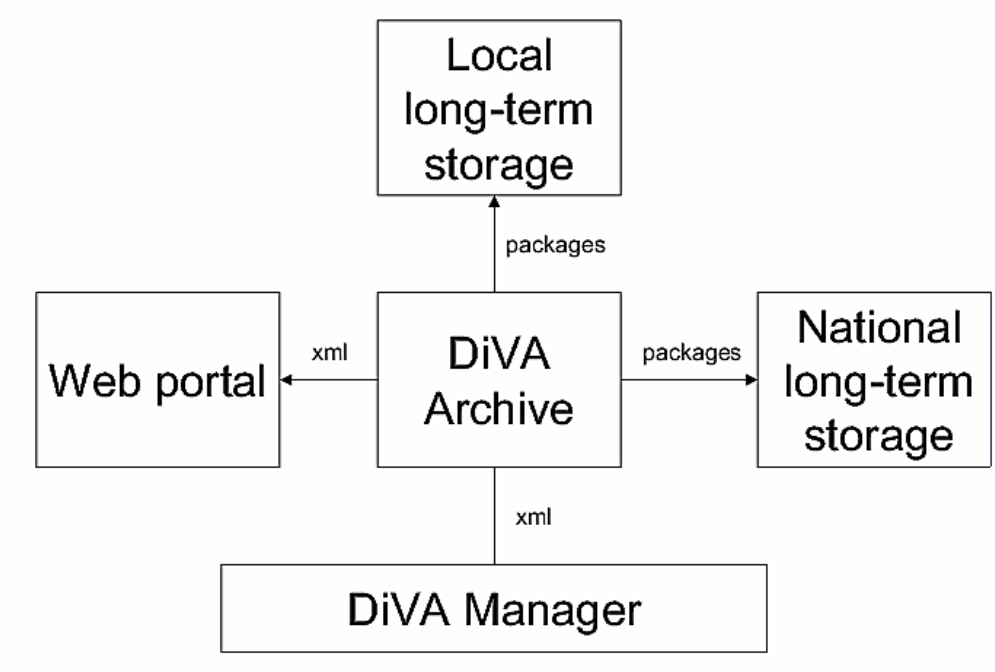
Preservation Server



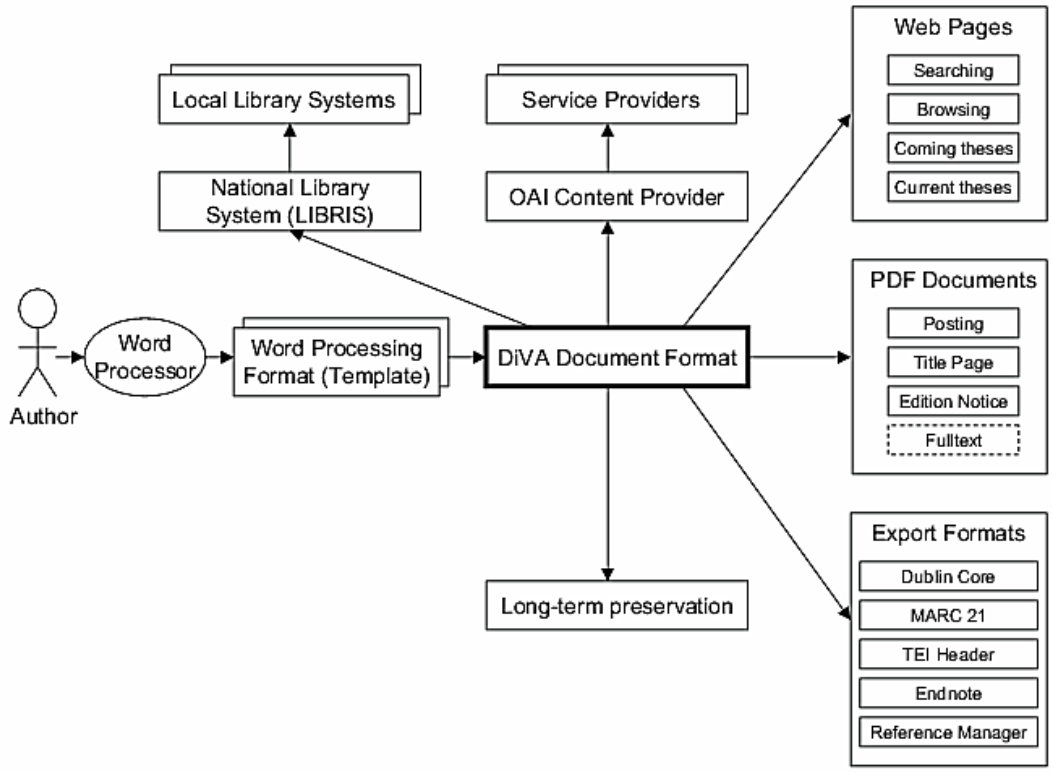
Digital Repository



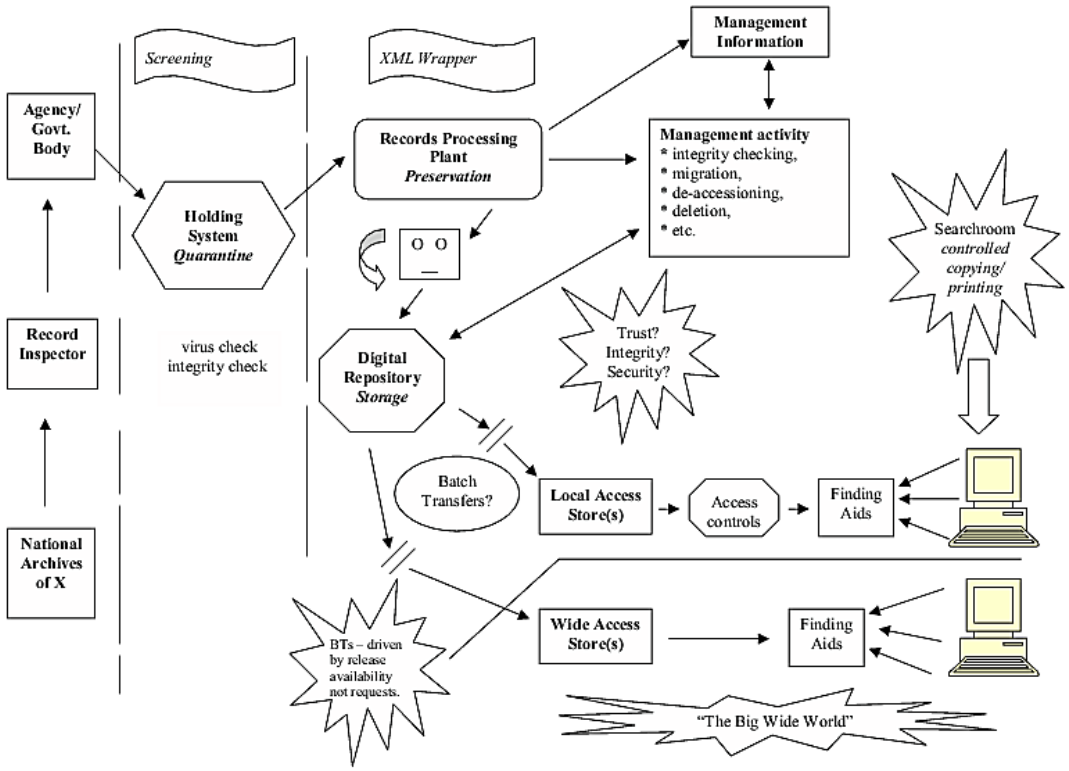
UUL -DiVA Project



UUL -DiVA Project



A Model of a Digital Repository



Some Challenges

- To make curation of digital assets strategic
- To establish necessary processes and procedures
- To bring about cultural change
- Setting up the technical systems required
- To take Preservation seriously

Preservation

- bit availability
- bit interpretation
- preservation description
  - provenance
  - context
  - fixity/authenticity
  - technical definitions and requirements
- descriptive information
- preservable formats
- persistent identifiers
- rights management