wePreserve

Historisch
Kulturwissenschaftliche
Informationsverarbeitung

An Introduction to Digital Preservation

Manfred Thaller

Universität zu* Köln

March 23rd, 2009

*University ***at*** not ***of*** Cologne

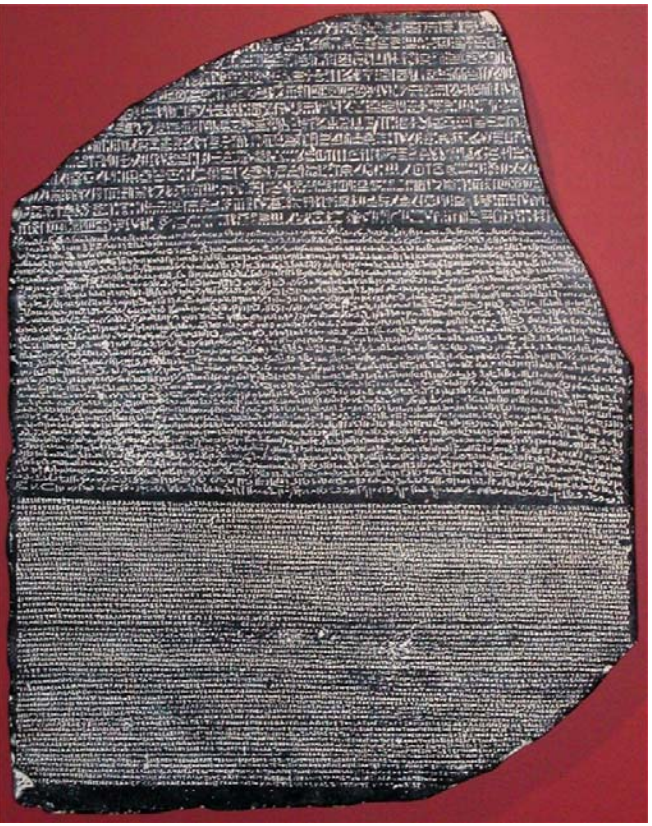
Information Society

digital preservation europe

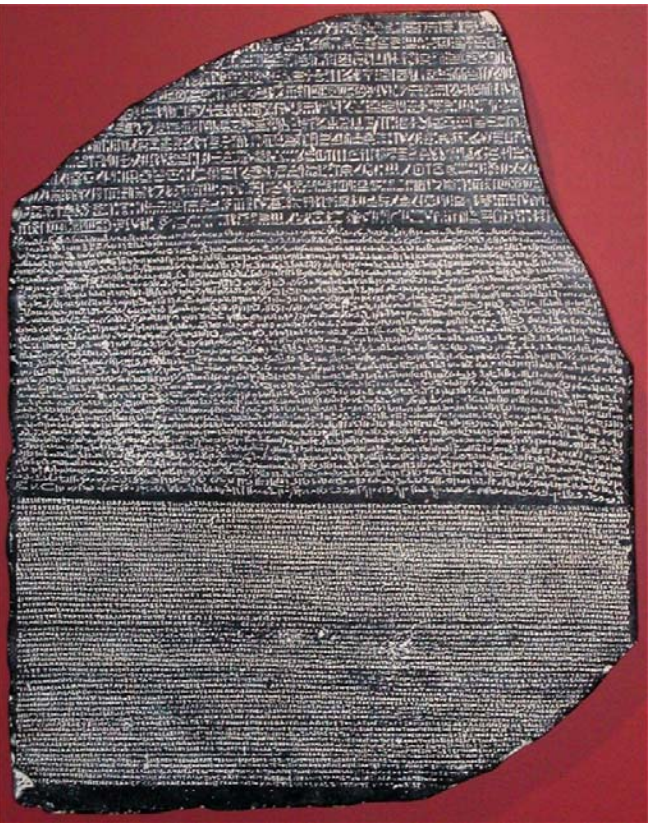
planets

CASPAR

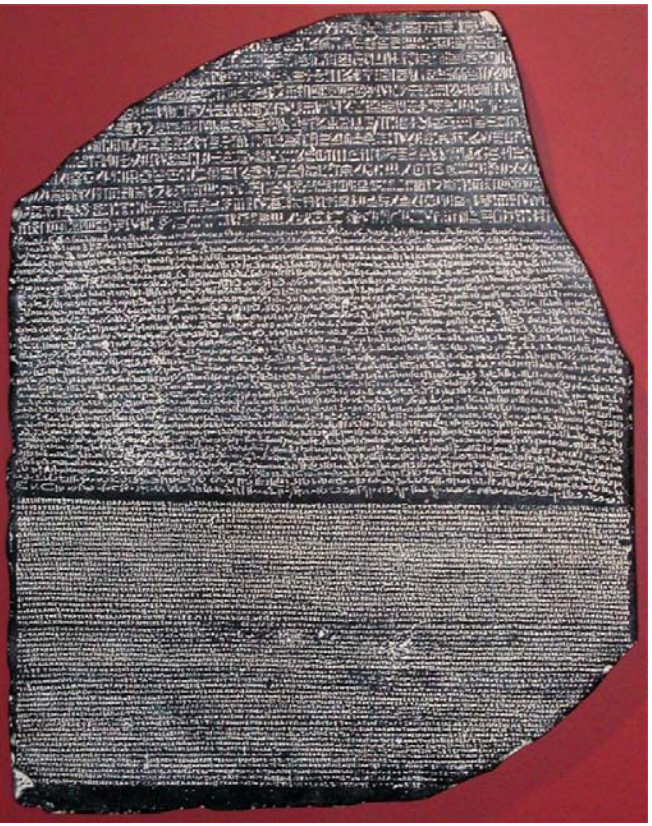
nestor



Modern information technology allows all memory institutions to make substantial amounts of their holdings accessible without delay and with great comfort.



Modern information technology allows all memory institutions to make substantial amounts of their holdings accessible without delay and with great comfort.



… for how long?

I – Challenges and responses

Challenge 1

Media fragility

Media curation

Challenge 2

Hardware obsolescence

Media migration

Challenge 3

Software obsolescence

Technology watch

Challenge 4

Format obsolescence

Migration

Emulation

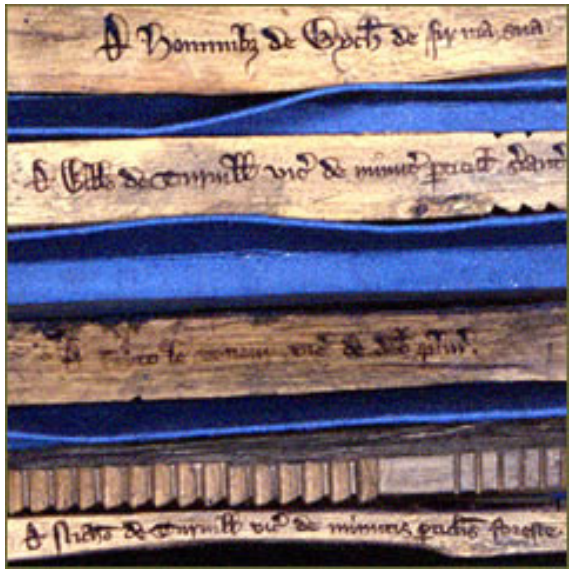
Challenge 5

Loss of technical context

Keep references alive

Challenge 6

Loss of semantic context



Challenge 6

Loss of semantic context

Make assumptions explicit

Summary I

Traditional media can be stored by passive protection (defend against disaster).

Digital media need permanent active curation.

“Active Curation”

Content Creation

Content Integrity

Content Maintenance

Association for Library Collections and Technical Services

(American Library Association)

“Active Curation 1”

Content creation includes:

- Clear and complete technical specifications
- Production of reliable master files
- Sufficient descriptive, administrative and structural metadata to ensure future access
- Detailed quality control of processes

“Active Curation 2”

Content integrity includes:

- Documentation of all policies, strategies and procedures
- Use of persistent identifiers
- Recorded provenance and change history for all objects
- Verification mechanisms
- Attention to security requirements
- Routine audits

“Active Curation 3”

Content maintenance includes:

- A robust computing and networking infrastructure
- Storage and synchronization of files at multiple sites
- Continuous monitoring and management of files
- Programs for refreshing, migration and emulation
- Creation and testing of disaster prevention and recovery plans
- Periodic review and updating of policies and procedures

Summary I

Traditional media can be stored by passive protection (defend against disaster).

Digital media need permanent active curation.

[Or new computer science paradigms …]

Summary II

"Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.“

American Library Association

II – Why should we care?

Business view

Because we are legally required.

To protect us from law suits.

To save investments.

Scientific view

To avoid duplication of costly experiments.

To document trends.

To protect integrity.

Cultural view

How can we not?!

How can we not?!

How can we not?!

However …

Learning how to decide what to throw away is as important, as learning how to keep things alive.

III – Are we serious?

CASPAR provides a comprehensive solution to digital preservation for all types of digitally encoded information, based on the OAIS reference model (ISO:14721:2002). It enhances techniques for capturing the many types of "metadata" which are needed to ensure that the preserved information will be usable and understandable by future users, despite changes in technologies. **CASPAR also integrates cutting edge** storage technologies, digital rights management, authentication, and accreditation as standard features, and will provide sophisticated access to and use of preserved digital resources including intuitive query and browsing mechanisms.

http://www.casparpreserves.eu/

planets

Planets builds practical services and tools to help ensure long-term access to digital cultural and scientific assets. With a special focus on the needs of libraries and archives, **Planets will deliver a downloadable "click-andinstall"** software package that supports administration, configuration, and deployment of preservation services and workflows.
Planets empowers organisations to define and execute preservation plans; analyse digital objects; identify their significant properties; take action to provide access to old data and software using sophisticated emulation techniques; and a Testbed for comparing digital preservation tools and assessing their effectiveness.

http://www.planets-project.eu/

digital preservation Europe

DigitalPreservationEurope (DPE) fosters collaboration and synergies between many existing national initiatives across the European Research Area. **DPE addresses** the need to improve coordination, cooperation and consistency in current activities to secure effective preservation of digital materials. **DPE leads work** to raise the profile of digital preservation, promote the ability of Member States acting together to add value to digital preservation activities across Europe, use cross-sectoral cooperation to avoid redundancy and duplication of effort, ensure auditable and certificated standards for digital preservation processes are selected and introduced.

http://www.digitalpreservationeurope.eu/

nestor – Network of Expertise in Long-Term Storage of Digital Resources has brought together existing knowledge and expertise from over 30 institutions as a starting point for future alliances for the preservation of Germany's digital memory. **nestor has** created an information and communication platform as a centre of contact for questions on long-term storage **nestor forms a visible focus in Germany for** cooperative efforts and for the creation of international alliances.

http://www.langzeitarchivierung.de/

Since 2008: Shaman

http://shaman-ip.eu

Since 2009: PrestoSpace

http://www.prestospace.org/

http://wiki.prestospace.org/pmwiki.php?n=Main.HomePage

This means …

Since 2006 the European Commission has initiated projects with a total sum of funding of close to 50 million Euros.

Yes, digital preservation is serious.

However …

"Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.“

Dangers

Too much planning, too little action.

Too many chiefs, too few little Indians.

The giant awakes …

The software suppliers for libraries / archives become preservation aware:

IBM: DIAS

ExLibris: Rosetta

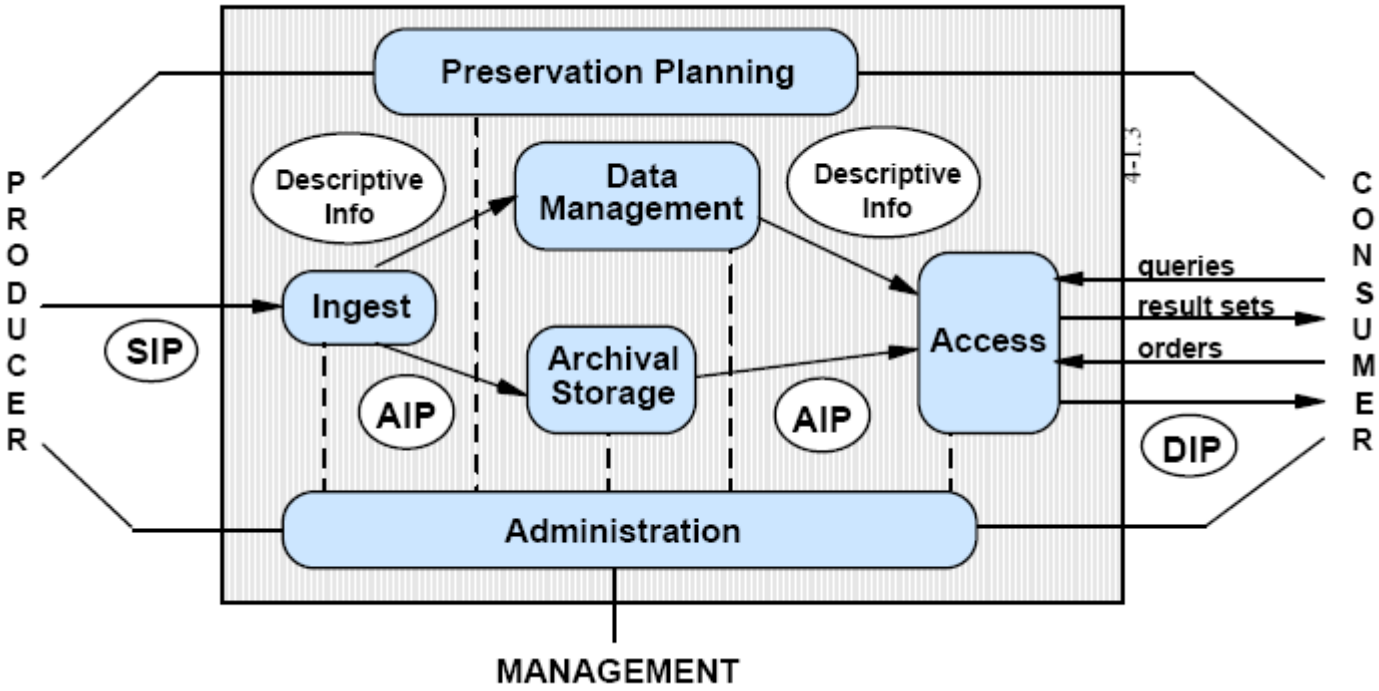
The giant awakes …

SNIA (Storage Networking Industry Association): XAM (eXtensible Access Method).

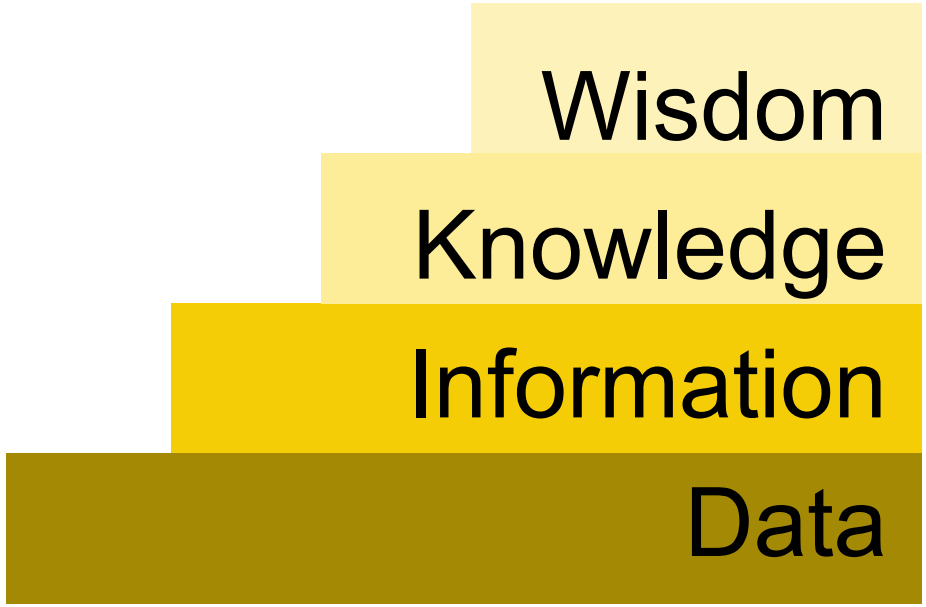
OAIS aware.

IV – Aspects of the problem

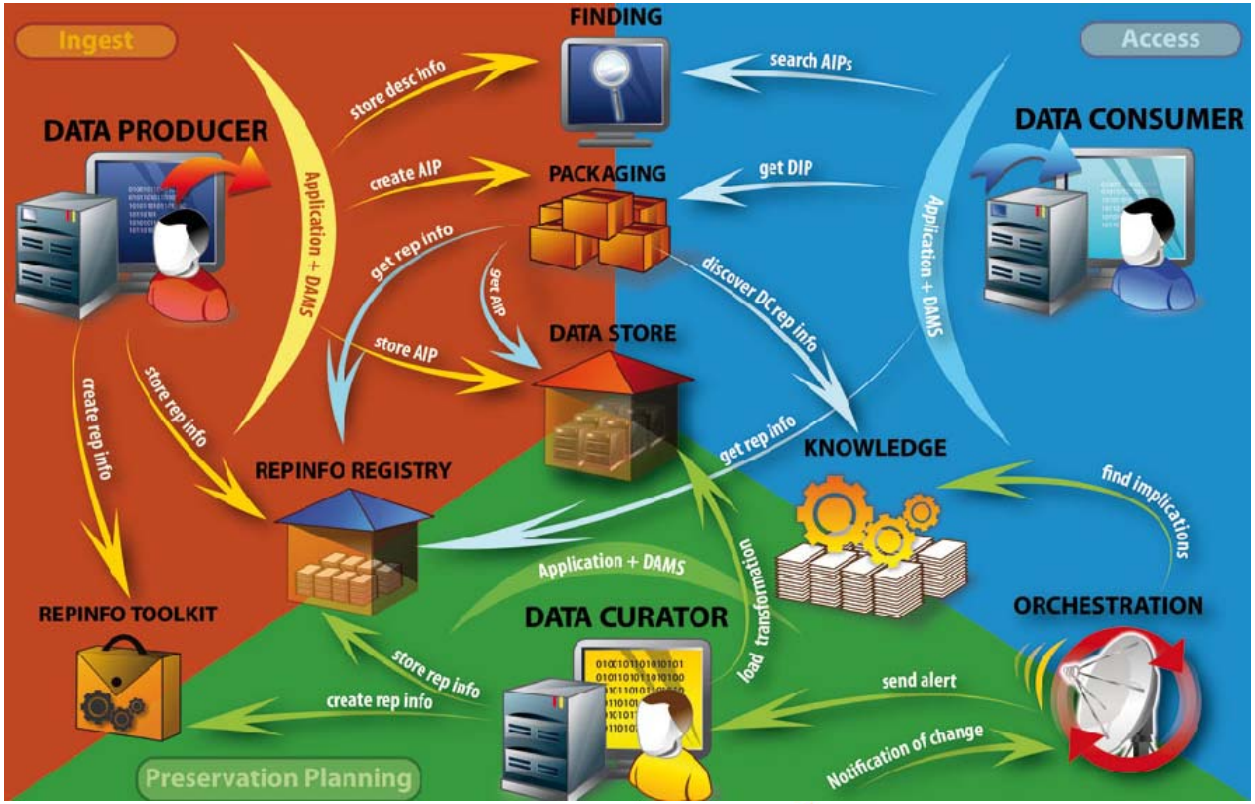
OAIS


Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
Ingest
SIP
Archival Storage
AIP
AIP
Access
queries
result sets
orders
DIP
CONSUMER
Administration
MANAGEMENT
4-1.3

Knowledge for Preservation



Infrastructure for Preservation



Intellectual Property Rights

How to protect, what you may not access?

How to preserve, what you may not observe?

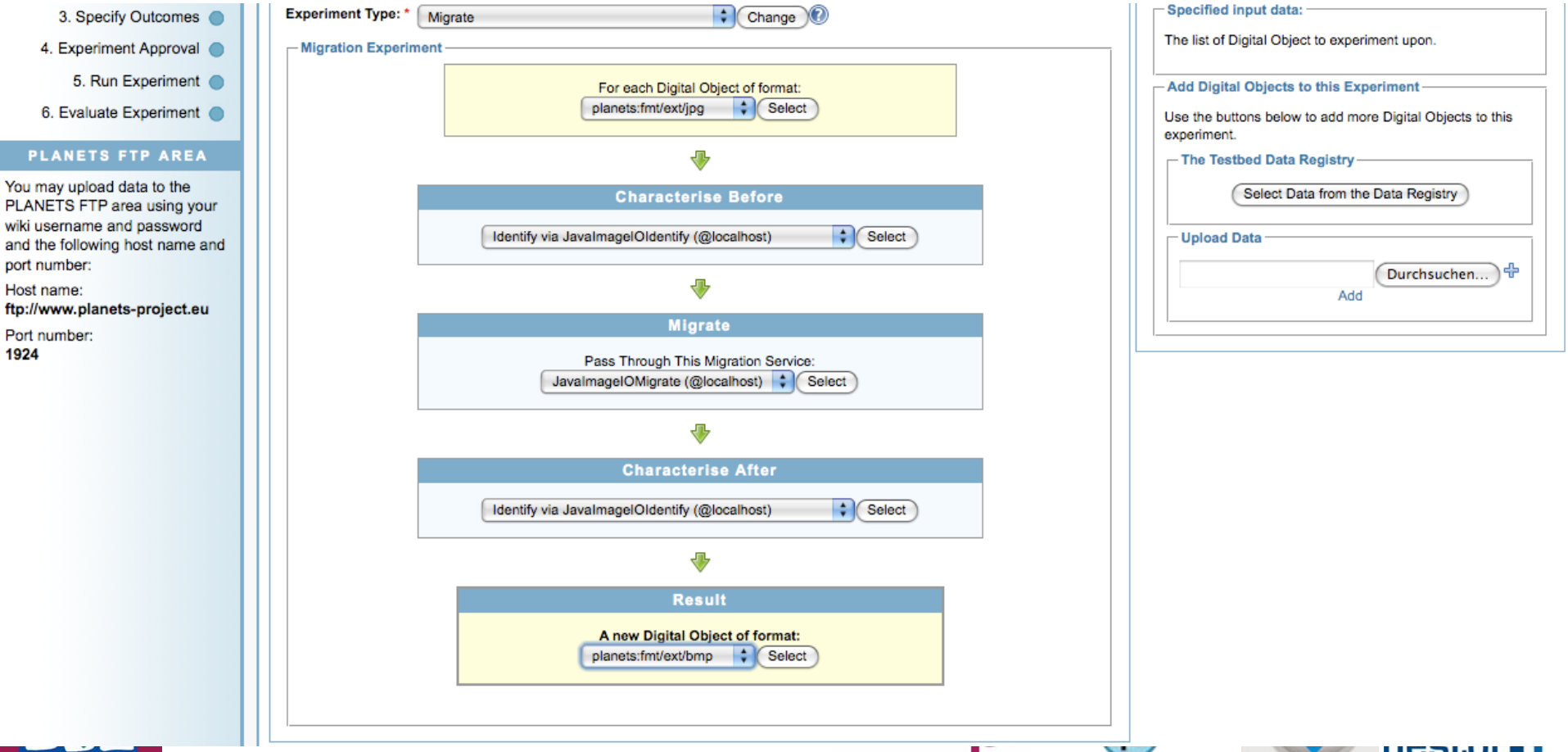
How to keep, what you do not have?

File formats and significant properties

► Photoshop ►

► Photoshop ►

Testbed


3. Specify Outcomes
4. Experiment Approval
5. Run Experiment
6. Evaluate Experiment
PLANETS FTP AREA
You may upload data to the PLANETS FTP area using your wiki username and password and the following host name and port number:
Host name:
ftp://www.planets-project.eu
Port number:
1924
Experiment Type: * Migrate Change
Migration Experiment
For each Digital Object of format:
planets:fmt/ext/jpg Select
Characterise Before
Identify via JavaImageIOIdentify (@localhost) Select
Migrate
Pass Through This Migration Service:
JavaImageIOMigrate (@localhost) Select
Characterise After
Identify via JavaImageIOIdentify (@localhost) Select
Result
A new Digital Object of format:
planets:fmt/ext/bmp Select
Specified input data:
The list of Digital Object to experiment upon.
Add Digital Objects to this Experiment
Use the buttons below to add more Digital Objects to this experiment.
The Testbed Data Registry
Select Data from the Data Registry
Upload Data
Durchsuchen...
Add

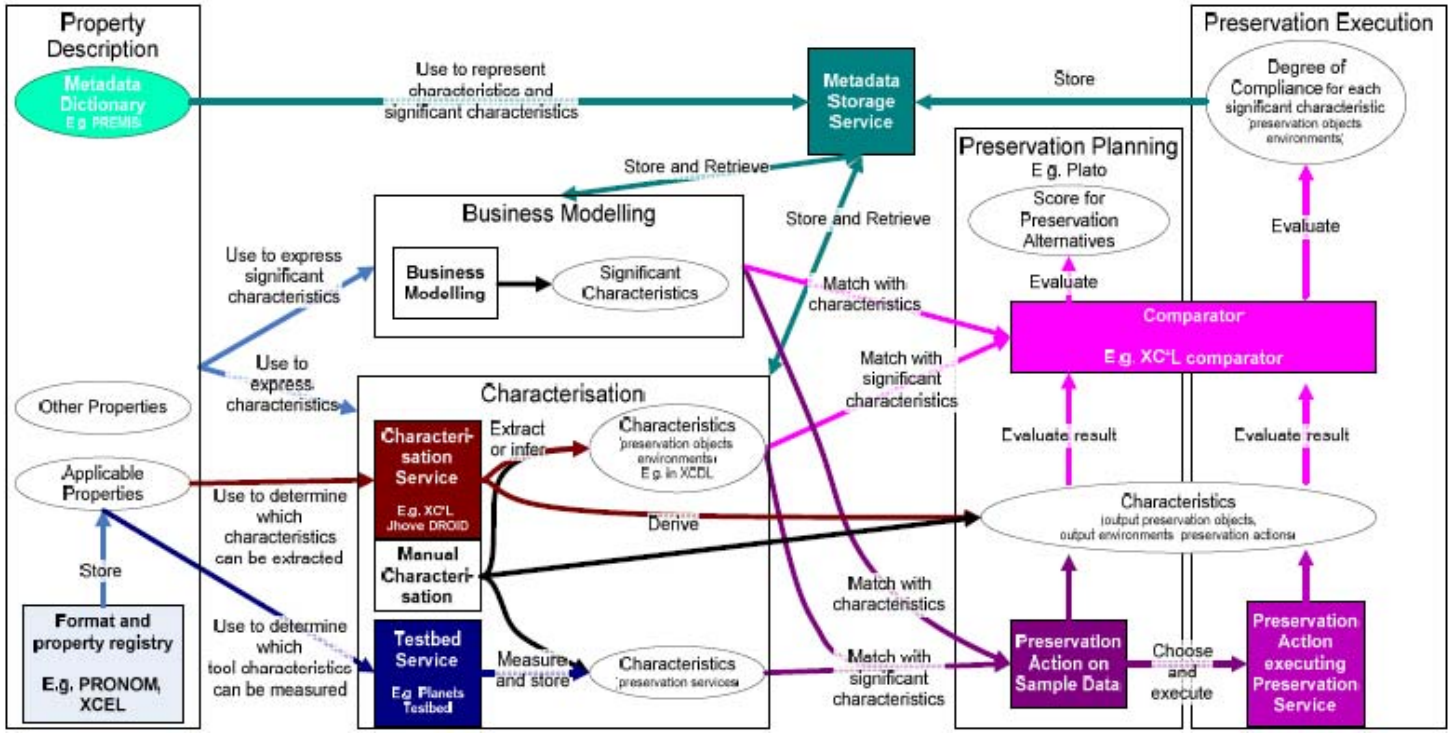
Trusted repositories

Will my byte streams survive?

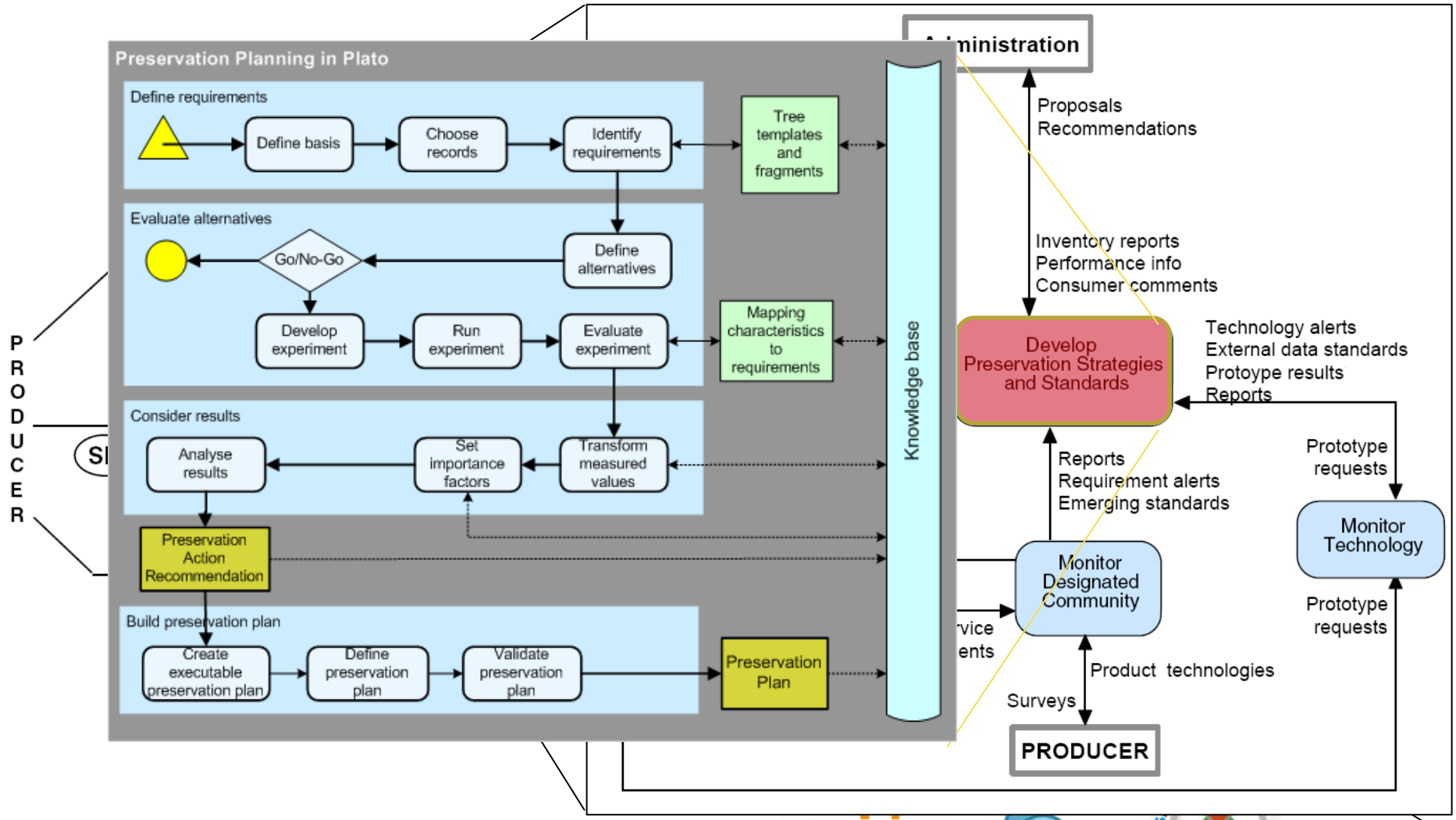
Will my information objects remain useful?

Will the institution continue to exist?

Preservation Metadata



digital preservation europe
planets
CASPAR
nestor
Information Society


Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results
Set importance factors
Transform measured values
Preservation Action Recommendation
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan
Knowledge base
PRODUCER
Administration
Proposals
Recommendations
Inventory reports
Performance info
Consumer comments
Develop Preservation Strategies and Standards
Technology alerts
External data standards
Protoype results
Reports
Reports
Requirement alerts
Emerging standards
Prototype requests
Monitor Technology
Monitor Designated Community
Prototype requests
Product technologies
Surveys
PRODUCER

Preservation Process: Preparation and Requirements

Identify your needs

Identify your resources

Document your decisions

Self Audit and Self-Certification: DRAMBORA

Γνώτι σ‘άυτων

Be aware of your strengths and weaknesses

Check, before flying

Thank you!

Manfred.Thaller@uni-koeln.de

Dieses Werk ist unter einem Creative Commons Namensnennung 3.0 Deutschland Lizenzvertrag lizenziert. Um die Lizenz anzusehen, gehen Sie bitte zu http://creativecommons.org/licenses/by/3.0/de/

oder schicken Sie einen Brief an

Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.