Digital Preservation Introduction

Andreas Rauber

Department of Software Technology and Interactive Systems

Vienna University of Technology

http://www.ifs.tuwien.ac.at/~andi

Overview

Part 1: Introduction

- What is Digital Preservation?
- What is the OAIS Reference model?
- How do we build a preservation plan?
- From Data to Processes
- Other issues in DP?

Why do we need Digital Preservation?

Questions / discussion:

- What is *Digital Preservation?*

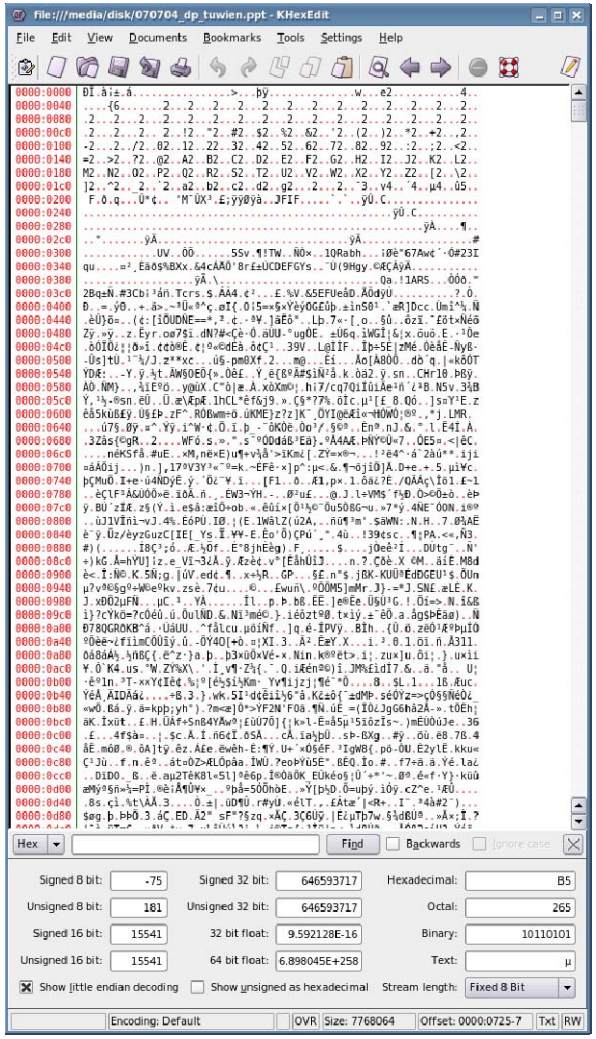
Why do we need Digital Preservation?

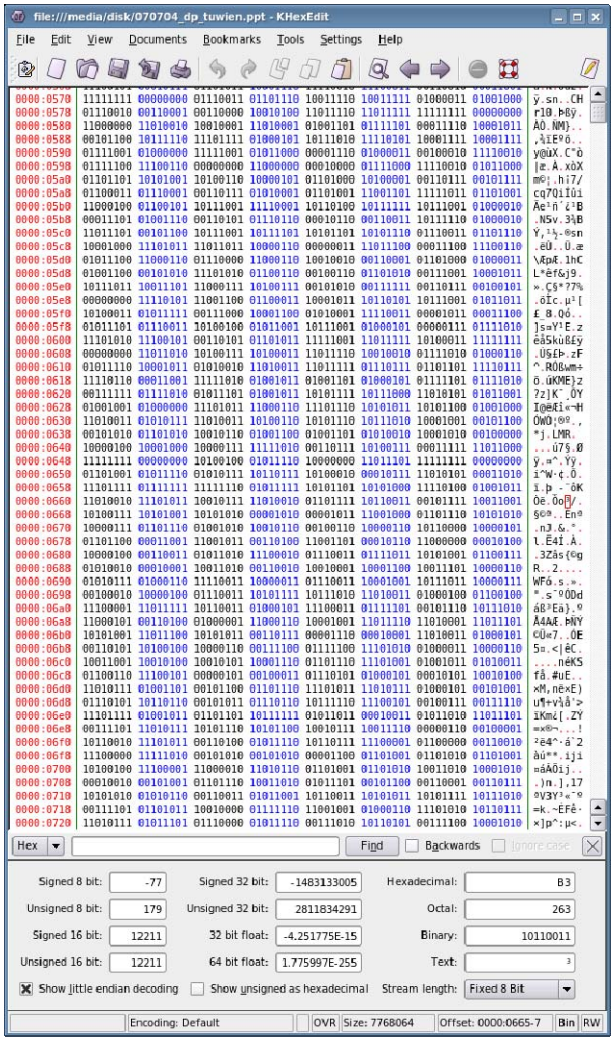
Why do we need Digital Preservation?

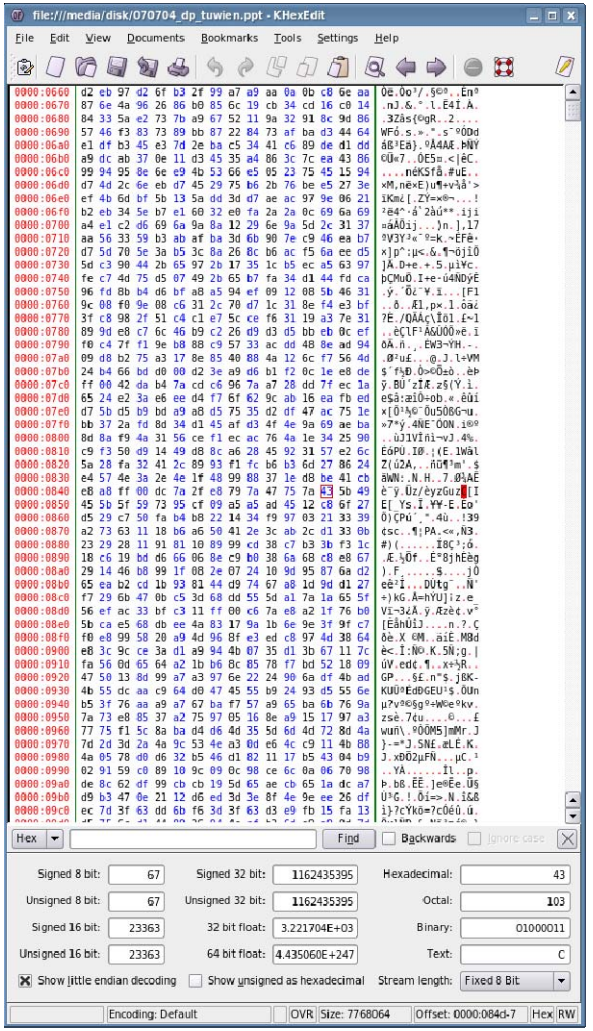
1. Physical Preservation (Bit-stream preservation)

- Transfering to current storage systems
  - note: transfer may not be trivial
    (file systems, encodings, relative references, copy protection,...)
- Ensure redundancy
  - technologically
  - geographic spread
- Access, security
- Error detection, recovery, disaster planning

Why do we need Digital Preservation?





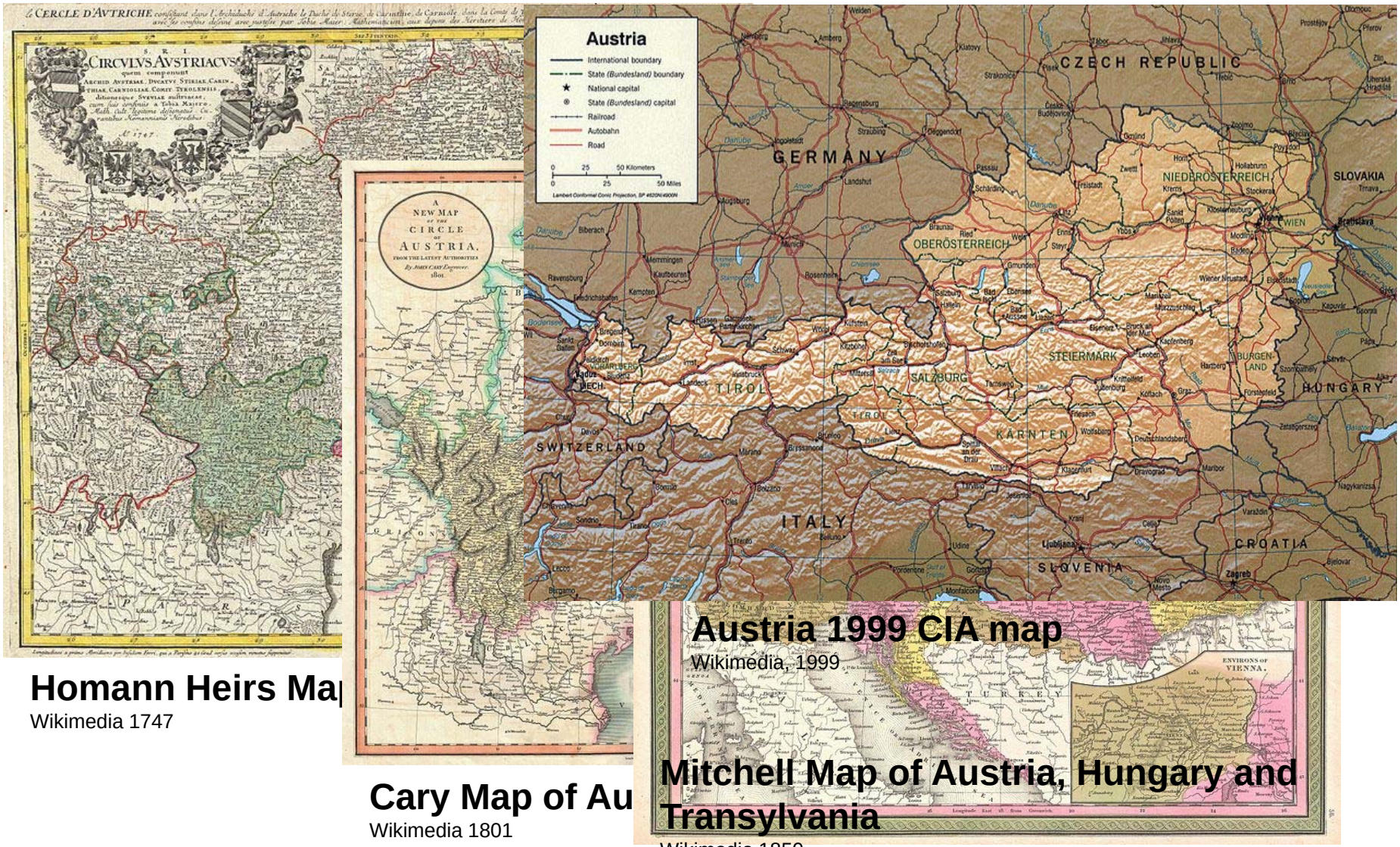


Why do we need Digital Preservation?

2. Logical Preservation

- Digital Objects require specific environment to be accessible :
  - Files need specific programs
  - Programs need specific operating systems (-versions)
  - Operating systems need specific hardware components
- SW/HW environment is not stable:
  - Files cannot be opened anymore
  - Embedded objects are no longer accessible/linked
  - Programs won‘t run
  - Information in digital form is lost
    (usually total loss, no degradation)
- Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.

Why do we need Digital Preservation?



Homann Heirs Ma
Wikimedia 1747

Cary Map of Au
Wikimedia 1801

Austria 1999 CIA map
Wikimedia, 1999

Mitchell Map of Austria, Hungary and Transylvania
Wikimedia 1850

Why do we need Digital Preservation?

3. Semantic Layer: information object

- How to interpret the data (information?) in the objects?
  - terminology changes:
    changes in country names, borders, connotation of words,...
  - concept changes:
    drunk driving: before 1998: 0.8‰ , afterwards 0.5‰
  - transformations: currencies/exchange rates, sensor resolutions,
  - provenance: actions applied to objects
    sources: who? / which sensor?, transformations, post-processing
  - context of objects:
    understanding the context of decisions, side-effects, quotations, calibration timestamps
- For preserving digital information, all 3 layers need to be addressed

Why do we need Digital Preservation

- The goal of Digital Preservation is to **maintain digital objects accessible and usable in an authentic manner for a long term** into the future.

Why do we need Digital Preservation?

- Essential for all digital objects
  - Office documents, accounting, emails, …
  - Scientific datasets, sensor data, metadata, …
  - Applications, simulations, business processes, …

- All application domains
  - Cultural heritage data
  - eGovernment, public administration
  - Science / Research
  - Industry
  - Health, pharmaceutical industry
  - Aviation, control systems, construction, …
  - Private data
  - …

Why do we need Digital Preservation?

Questions / discussion:

- What is *digital data*?
- What is *digital storage*?
- What do we mean by
  - *accessible*?
  - *authentic*?
  - *long-term*?

Why do we need Digital Preservation

3 levels of threat / preservation

1. Bit rot – physical preservation / bit preservation
   Physical Layer: how to keep the 0's and 1's
2. Object formats – logical preservation
   Logical Layer: how to remain able to open a file, run a program
3. Authenticity, interpretability – semantic preservation
   Semantic Layer: how to ensure we can understand/interpret data correctly

- What can we do?

Bit-level preservation

- Maintain bit-sequence
- Redundant storage:
  - Lockss: lots of copies keeps stuff safe
  - Cloud
- Distributed storage – physically separated
- Different technologies / platforms / production batches
- Controlled storage conditions
- Regular maintenance: tape rewinding, disc spinning, …
- Maintain devices for accessing storage!
- Trade-off capacity, energy, effort

Bit-level preservation

Questions / discussion:

- How long do tapes / CDs / DVDs / HDDs / SSD last?
- What are the costs of bit-level preservation?
- What are the logistic challenges?
- Is a DVD that lasts for 200 years a solution?
- What would be the most durable storage technologies?
- What is "digital storage"?
- Distribution and Trust?
- Are we allowed to store redundantly? in the cloud?
  - Copyright
  - Copy protection

Why do we need Digital Preservation

3 levels of threat / preservation

1. Bit rot – physical preservation / bit preservation
   Physical Layer: how to keep the 0's and 1's
2. Object formats – logical preservation
   Logical Layer: how to remain able to open a file, run a program
3. Authenticity, interpretability – semantic preservation
   Semantic Layer: how to ensure we can understand/interpret data correctly

- How? What can we do?

Logical Preservation

Deja vue:

- Digital Objects require specific environment to be accessible :
  - Files need specific programs
  - Programs need specific operating systems (-versions)
  - Operating systems need specific hardware components
- SW/HW environment is not stable:
  - Files cannot be opened anymore
  - Embedded objects are no longer accessible/linked
  - Programs won‘t run
  - Information in digital form is lost
    (usually total loss, no degradation)
- Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.

Strategies for Logical Preservation

Technology Museum

- Keep the hardware (drives, computer,…)
- \+ Maintains full functionality
- \+ Creates time buffer to develop more permanent strategies
- \+ Requires detailed documentation of HW and SW, but this also helps
- \+ Only strategy for some types of objects? (which?)
- \- Economically and technically infeasible to maintain spare parts forever
- \- Requires huge "museum"
- \- Requires highly specialized know-how for all platforms and software

Strategies for Logical Preservation

Migration

- Transform into different format
- Continually or on demand (Viewer)

\+ Widely used

\+ Possibility to compare at time of migration

\+ Resulting objects are always accessible

\- Possibly undesired changes during migration

\- Needs to be repeated again and again

Strategies for Logical Preservation

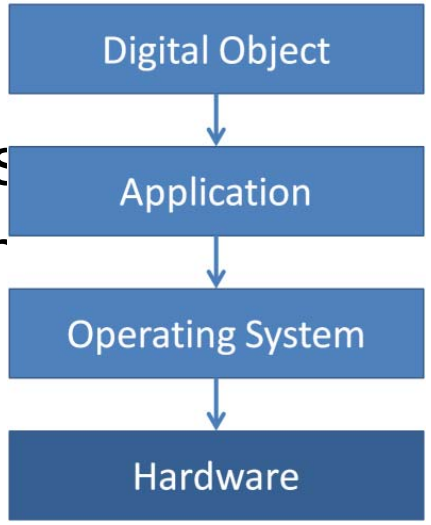
Emulation

- Emulation of Hardware or Software (OS, application)
+ Widely used principle
+ Many emulators available
+ Potentially preserving complete functionality
+ *Document is unchanged*
- *Document is unchanged*
- Complex technology, lot of research required
- Requires detailed documentation of the system
- Requires experience how to interact with emulated historic system in the future
- Emulators must be migrated as well
- Emulators potentially erroneous (Complexity)

Strategies for Logical Preservation

Excursion: Emulation vs. Migration

- Different on the pragmatic level, but conceptually identical
- Change occurs somewhere in the viewpath
- Have basically the same advantages/disadvantages and characteristics
- None of them guarantees identical rendering/performance of digital objects
- Many variants (e.g. viewer, virtualization)
- Need to be evaluated the same way



Strategies for Logical Preservation

Standardization

- Using open or de-facto standards
- \+ Simplifies DP process
- \+ Many tools available
- \+ Tools for standards are easier to build also in the future
- \- Significant effort required for standardization
- \- Loss at converting into standard
  (who is responsible?)
- \- Some object types cannot be standardized

Strategies for Logical Preservation

Standardization - Excursion into file formats
Proprietary vs. Open

- Proprietary
  - Documentation mostly not available
  - License and patent rules
  - License agreements subject to change
  - Restrictions for use and modifications may apply
- Open
  - Documentation available!
  - Unlimited use
  - No license fee
  - Open for modifications
  - No patent owners
- But: sometimes proprietary may better than open - **why**?
- Is the concept of "file formats" still useful?

Strategies for Logical Preservation

Limiting Accepted Formats

- ▪ Similar to standardization
- \+ Reduces challenge to smaller number of formats
- \- Does not solve the problem
- \- Limits the type of objects that can be accepted
- \- Potential loss at conversion
- \- Requires strict control of formats (and what's in them!)

Strategies for Logical Preservation

Data/Information Extraction

- Create abstract representation of information (e.g. databases or documents -> XML)
- \+ Independent of specific infrastructure
- \+ Many tools available
- \+ Easier to develop tools in the future
- \- High effort to develop tools for specific abstraction scenario
- \- Limited functionality of tools designed to interpret information, many aspects not preservable
- \- Cannot be applied to all types of objects

Strategies for Logical Preservation

Encapsulation

- Add metadata, software,... (representation information) to object („onion“)
- \+ Simplifies search for preservation solution on demand, offering several potential layers
- \+ Always allows for the application of several other strategies at different levels
- \- Does not solve the problem
- \- Even with all information encapsulated we may not be able to find a solution

Strategies for Logical Preservation

Universal Computing Platform

- Example: UVC: Universal Virtual Computer (IBM)
- Abstract virtual machine, intermediate platform that can be implemented on many other platforms

\+ works for documents and software

\+ A kind of standardization for platforms, reduces development effort

\+ Can test solution at time when being developed

\- Pretty complex (cf. Java, but that's still simple)

\- High effort at time of preservation

\- Requires cooperation of the producers of information

\- High risk of loosing aspects of information

Strategies for Logical Preservation

Backwards Compatibility and Version Migration

- current SW reads old versions and performs migration
- \+ Usually available
- \+ Creates time buffer for more permanent solutions
- \+ sometimes equal or better functionality
- \- Doubtful whether this will work for a long time (why?)
- \- Each change might lead to unwanted changes
- \- No guarantee from part of the producer of the SW

Strategies for Logical Preservation

Viewer

- Migration on demand, interpretation by Viewer software
- \+ Original datastream unchanged, interpreted directly
- \+ No continuous migration
- \+ No cummulative errors
- \- Viewer sometimes cannot process all (parts of) objects
- \- Time delay when developing viewers, increasing
- \- Viewer SW must be carried along with technology changes
- \- Hard to evaluate whether viewer is correct

Strategies for Logical Preservation

Non-digital Strategies

- Printing to paper, microfilm, ...
+ Reqires transformation to readable form -> stable
+ Coding of digital data is possible
+ Lots of experience in handling analog data carriers
+ High stability -> Bit-stream Preservation
- Loosing functionality, loosing advantage of digital technology
- Not applicable for all objects
- High costs for preserving some of the analog data carrier material, low storage density, ...

Strategies for Logical Preservation

Data Recovery, Data Archeology

- Analysis of bit-stream to interpret data, digital forensics
+ Probably only approach to recover "lost" information
- No guarantee that it works
- Without sufficient documentation close to "guessing"
- Extremely high costs per object
- Hard to estimate on whether it may be successful for a given object

Logical Preservation

Summary

- Changing object, environment
- Loss upon migration / emulation
- Decision of what to preserve → **Significant Properties**!
- How to detect/document what you lost?
- Range of strategies available, none is perfect
- Combination of strategies
- No solution forever -> DP is a process!

Logical Preservation

- Preservation Planning
- Identify objects at risk
- Standardization reduces risk (why?)
- Apply preservation actions such as migration / emulation / HW-museum
- Identify what you need to preserve (significant properties)
- Identify suitability of tools
- Find out what you can preserve / what you loose
- Do it, document it, verify it, monitor it

Logical Preservation

Questions / Discussion:

- What are the problems of logical preservation?
- What is the optimal strategy?
- What is the optimal strategy for a specific object?
- What is a good format / platform (e.g. to migrate to)?
- What are characteristics of good formats/platforms/… ?
- How can we identify objects at risk?
- When is a format "more/less risky"?
- What is a file format?
- How can we find out what we loose with a strategy?

Logical Preservation

Questions / Discussion (2):

- What is the difference between emulation and migration? Are they different? Are they not different?
- What are the significant properties of an object / process?
- “I want to preserve everything” – (how) can we do this?
- What is the “original object”?
- Is XML the solution to DP?
- What is the complexity of each strategy? Costs? Effort?
- What know-how do we need to decide on a strategy?
- What would be potential risks/difficulties e.g. for construction plans? Medical imaging (DICOM)?

Logical preservation

Questions / discussion (3):

- Which objects are most at risk?
- Which objects are most difficult to preserve?
- How do we preserve entire business processes?
- If we loose significant properties with a strategy, what is the impact on authenticity? Can we use a “changed” object?
- What is the difference to systems engineering?

Why do we need Digital Preservation

3 levels of threat / preservation

1. Bit rot – physical preservation / bit preservation
   Physical Layer: how to keep the 0's and 1's
2. Object formats – logical preservation
   Logical Layer: how to remain able to open a file, run a program
3. Authenticity, interpretability – semantic preservation
   Semantic Layer: how to ensure we can understand/interpret data correctly

- How? What can we do?

Semantic preservation

- Threats at semantic level
  - meaning of terms change: city names, …
  - measurement scales, sensor sensitivity, …change
  - interpretation of facts change: alcohol levels, …
- Rather long-term, but subtle to notice
- Consider context of objects
  - purpose, setting, limitations, cultural context, related objects, …

Semantic preservation

- Approaches / solutions:
  - Semantic enrichment
  - Metadata
  - Migration at semantic level
  - Documentation of context
  - Tracing of metadata
  - Document intended meaning / interpretation

Semantic preservation

Questions / discussion:

- How do we identify need for action?
- What is the risk of missing timely action?
- How do we solidly identify and document context?
- How can we implement semantic enrichment / semantic migration, …?
- What about security issues?
- Is PDF save? PDF/A?
- Who is allowed to have access to which documents? Who had access to them?
- Are differences in the communication protocol at an API level a problem of logical or semantic preservation?

From Data to Processes

- Assume we know how to preserve data - **Is this sufficient?**
- Preserving data: Data Management Plans
  - describing data and context: provenance, authenticity, representation information,...
  - range of (ambiguous) definitions of context
  - But: mostly not actionable, not enforcable,...
  - BUT: data are (just) results of processes!
- Processes may be needed to
  - verify data
  - understand provenance
  - re-use process on new data
  - integrate data over time
- **Process curation instead of data curation!**

Digital Preservation - Summary

- Is a complex task
- Requires a concise understanding of the objects, their intellectual characteristics, the way they were created and used and how they will most likely be used in the future
- Requires a continuous commitment to preserve objects to avoid the „digital dark hole“
- Requires a solid, trusted infrastructure and workflows to ensure digital objects are not lost
- Is essential to maintain electronic publications & data accessible
- Will become more complex as digital objects become more complex
- Needs to be defined in a preservation plan

Digital Preservation

- Reference Models
  - Records Management, ISO 15489:2000
  - OAIS: Open Archival Information System, ISO 14721:2003
- Audit & Certification Initiatives
  - RLG- National Archives and Records Administration Digital Repository Certification Task Force:
    Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
  - NESTOR:
    Catalogue of Criteria of Trusted Digital Repositories
  - DCC/DPE:
    DRAMBORA: Digital Repository Audit Method Based on Risk Assessment

Digital Preservation

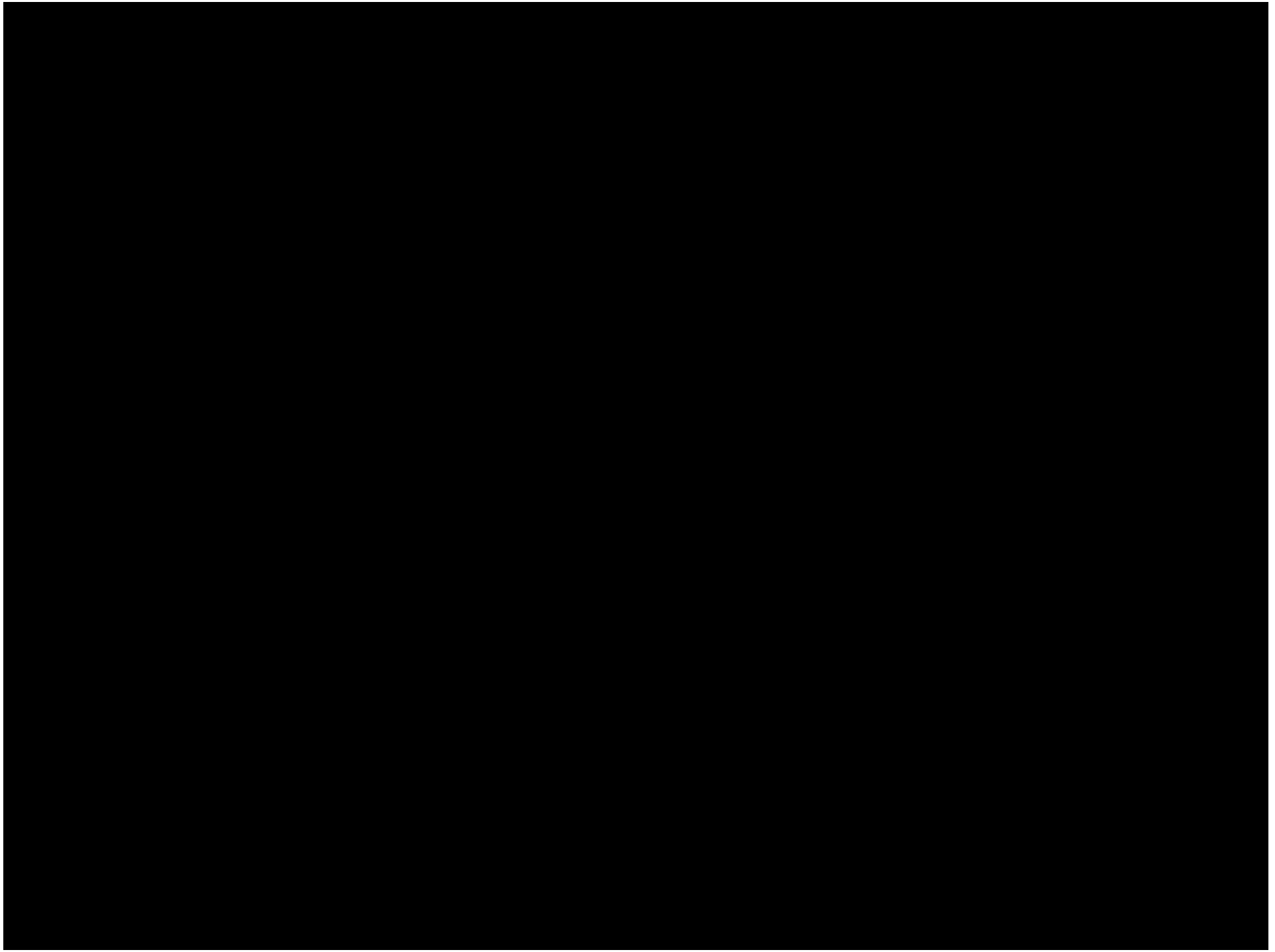
Questions / Discussion:

- At what levels are digital objects threatened?
- What are the time intervals at each level?
- How can we identify objects at risk?
- What can we do to mitigate the risk?
- How can we recover if mitigation fails / is missed?
- What competences do we need?
- How would a training/education program look like?
- How do we know if somebody is doing a good job at DP?

Overview

Part 1: Introduction

- What is Digital Preservation?
- Break? - Video?
- What is the OAIS Reference model?
- How do we build a preservation plan?
- From Data to Processes
- Other issues in DP?



Overview

Part 1: Introduction

- What is Digital Preservation?
- What is the OAIS Reference model?
- How do we build a preservation plan?
- From Data to Processes
- Other issues in DP?

OAIS

- NASA: National Space Science Data Center
  - NASA's first digital archive
  - Experienced many technological changes since 1966
- Consultative Committee for Space Data Systems
  - International group of space agencies
  - Developed range of discipline-independent standards
  - Evolved into ISO TC 20/ SC 13 working group around 1990
  - TC20: Aircraft and Space Vehicles
  - SC13: Space Data and Information Transfer Systems

OAIS

- Reference Model for an Open Archival Information System (OAIS), Blue Book, CCSDS 650.0-B-1, January 2002
- ISO 14721:2003; ISO 17721:2012
- Pink book, June 2012
  http://public.ccsds.org/publications/archive/650x0m2.pdf

- slides based on Blue Book and:
  - Don Sawyer, Lou Reich: ISO Reference Model for an Open Archival Information System (OAIS) Tutorial Presentation, LOC, June 13 2003

- http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html

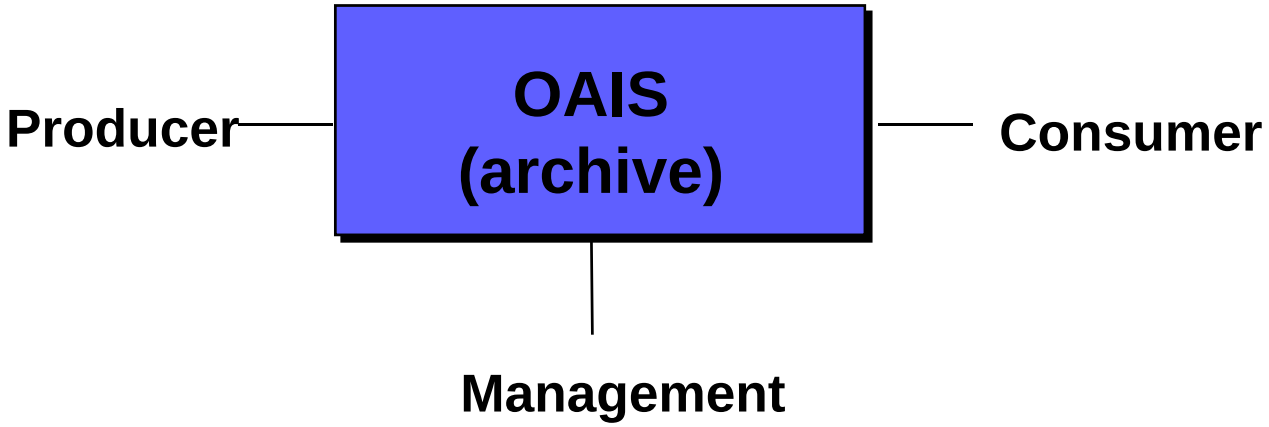
OAIS

- Framework for understanding and applying concepts needed for long-term digital information preservation
  - Long-term: long enough to be concerned about changing technologies
  - Starting point for model addressing non-digital information
- Provides set of minimal responsibilities to distinguish an OAIS from other uses of ‘archive’
- Framework for comparing alcohol levels architectures and operations of existing and future archives
- Addresses a full range of archival functions
- Applicable to all long-term archives and those organizations and individuals dealing with information that may need long-term preservation
- Does NOT specify an implementation

OAIS

- OAIS helps understanding / structuring DP
- Is not “perfect”
  - Conflicting models, different views
- Does NOT specify an implementation model !!!
- Difficult balance between high-level structure and detailed guidelines, not consistently solved
- Has to be understood wrt. its time of origin and purpose
- Standards create their own dynamics

OAIS

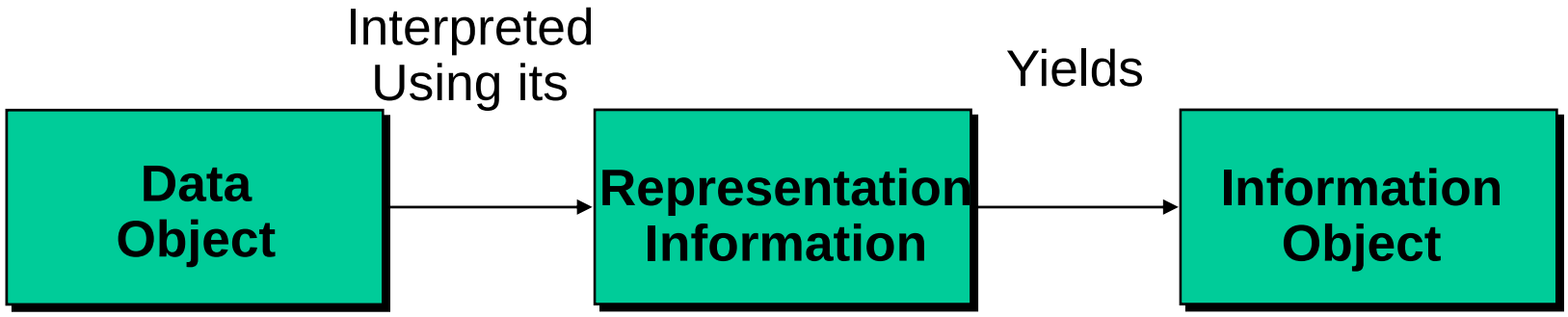


- Producer is the role played by those persons, or client systems, who provide the information to be preserved
- Management is the role played by those who set overall OAIS policy as one component in a broader policy domain
- Consumer is the role played by those persons, or client systems, who interact with OAIS services to find and acquire preserved information of interest

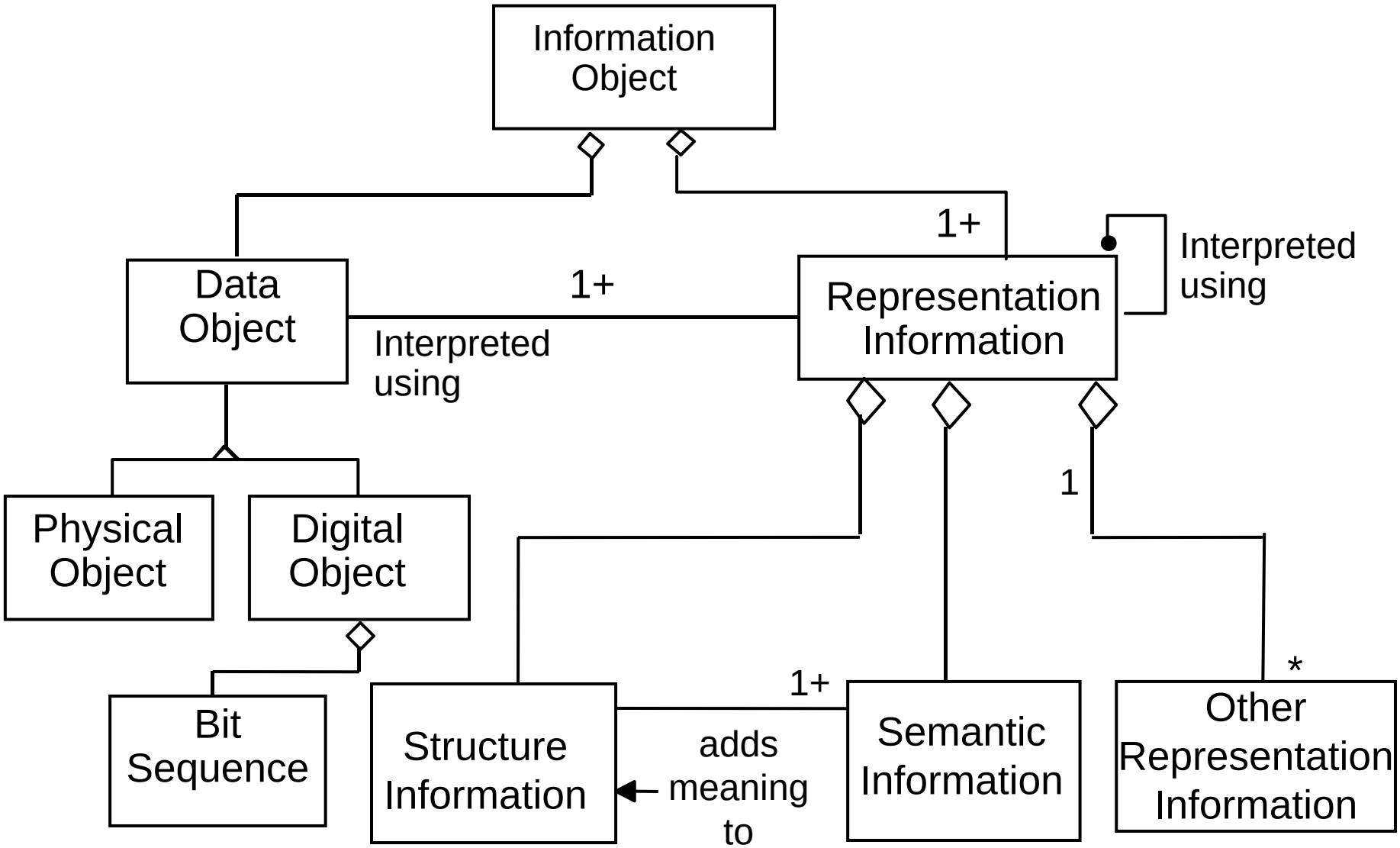
OAIS

OAIS Information Definition

- Information is always expressed (i.e., represented) by some type of data
- Data interpreted using its Representation Information yields Information
- Information Object preservation requires clear identification and understanding of the Data Object and its associated Representation Information



OAIS

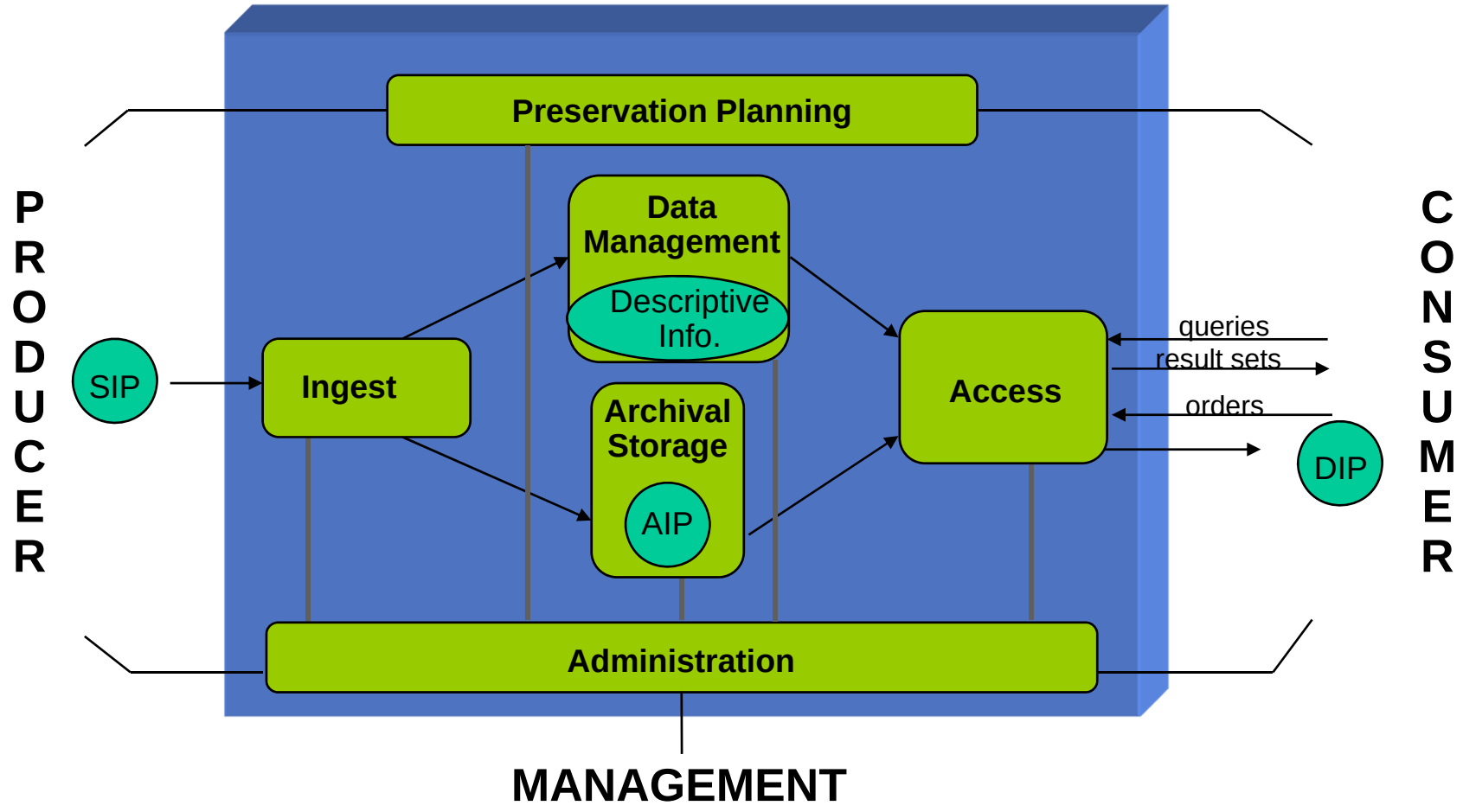

Information Object
Data Object
Representation Information
1+
1+
Interpreted using
Interpreted using
Physical Object
Digital Object
1
Bit Sequence
Structure Information
1+
adds meaning to
Semantic Information
*
Other Representation Information

OAIS

Information Package Variants

- **SIP**: Submission Information Package
  - Negotiated between Producer and OAIS
  - Sent to OAIS by a Producer
- **AIP**: Archival Information Package
  - Information Package used for preservation
  - Includes complete set of Preservation Description Information (PDI) for the Content Information
- **DIP**: Dissemination Information Package
  - Includes part or all of one or more Archival Information Packages
  - Sent to a Consumer by the OAIS

OAIS

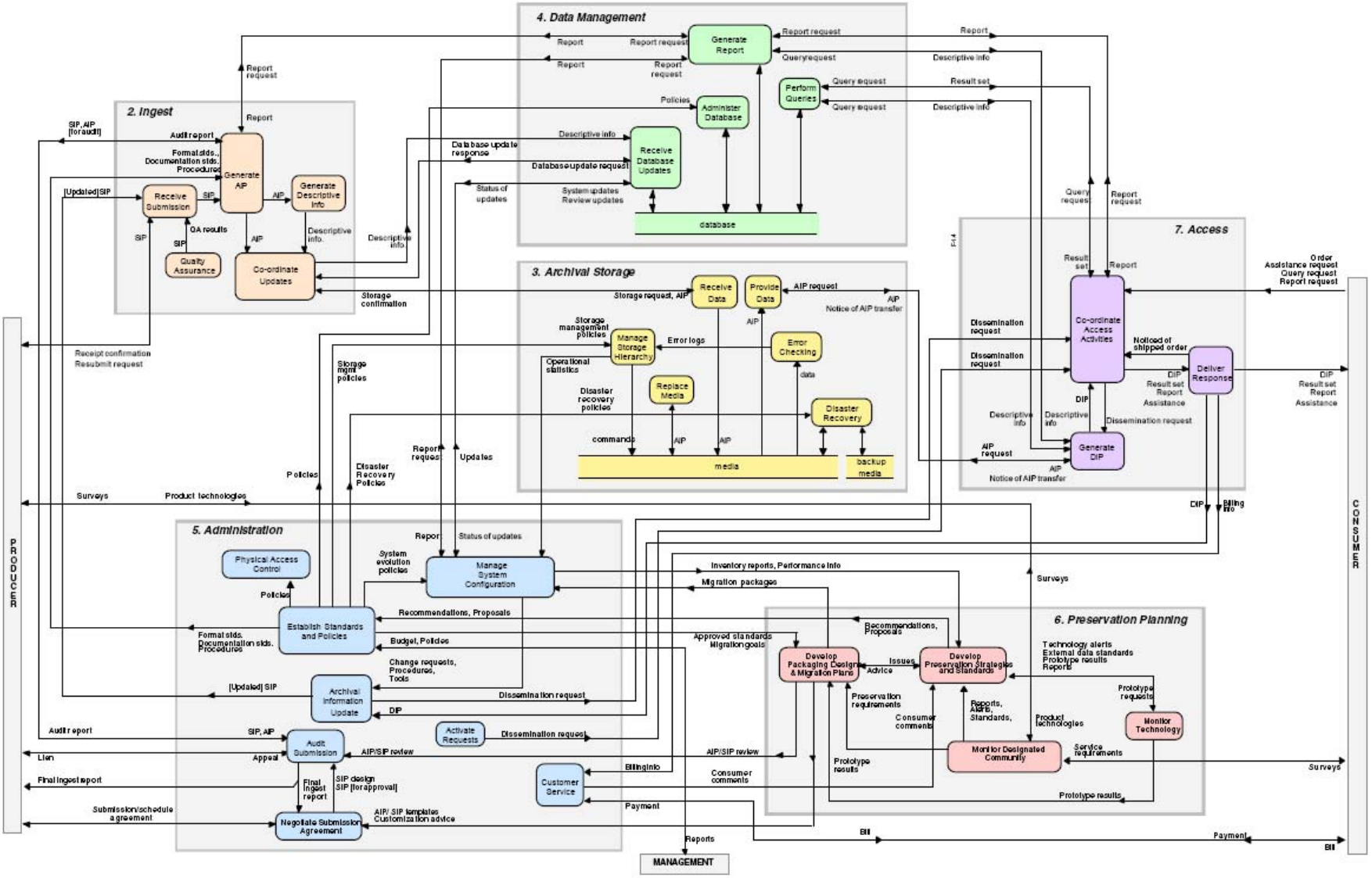


SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package

OAIS


4. Data Management
Generate Report
Report request
Report
Query request
Descriptive info
Perform Queries
Result set
Administer Database
Policies
Receive Database Updates
Database update request
Database update response
System updates
Review updates
Status of updates
database
2. Ingest
SIP, AIP [for audit]
Audit report
Format stds., Documentation stds., Procedures
Generate AIP
Generate Descriptive Info
[Updated] SIP
Receive Submission
QA results
Quality Assurance
Co-ordinate Updates
Descriptive info.
Storage confirmation
3. Archival Storage
Receive Data
Provide Data
AIP request
Storage request, AIP
Notice of AIP transfer
Storage management policies
Manage Storage Hierarchy
Error logs
Error Checking
Operational statistics
Replace Media
Disaster recovery policies
Disaster Recovery
commands
media
backup media
7. Access
Co-ordinate Access Activities
Deliver Response
Dissemination request
Noticed of shipped order
DIP
Result set
Report
Assistance
Generate DIP
Order
Assistance request
Query request
Report request
Receipt confirmation
Resubmit request
Storage mgmt policies
Disaster Recovery Policies
Surveys
Product technologies
PRODUCER
CONSUMER
5. Administration
Physical Access Control
System evolution policies
Manage System Configuration
Establish Standards and Policies
Format stds. Documentation stds. Procedures
Recommendations, Proposals
Budget, Policies
Change requests, Procedures, Tools
Archival Information Update
Activate Requests
Audit Submission
Audit report
SIP, AIP
Appeal
Lien
Final ingest report
SIP design
SIP [for approval]
Negotiate Submission Agreement
Submission/schedule agreement
AIP/ SIP templates
Customization advice
Customer Service
Billing info
Payment
Bill
Reports
MANAGEMENT
6. Preservation Planning
Develop Packaging Design & Migration Plans
Develop Preservation Strategies and Standards
Monitor Designated Community
Monitor Technology
Inventory reports, Performance info
Migration packages
Approved standards
Migration goals
Issues
Advice
Preservation requirements
Consumer comments
Reports, Alerts, Standards,
Product technologies
Prototype requests
Prototype results
Service requirements

Overview

Part 1: Introduction

- What is Digital Preservation?
- What is the OAIS Reference model?
- How do we build a preservation plan?
- From Data to Processes
- Other issues in DP?

Preservation Planning

Why Preservation Planning?

- Several preservation strategies developed
  - For each strategy: several tools available
    - For each tool: several parameter settings available
- How do you know which one is most suitable?
- What are the needs of your users? Now? In the future?
- Which aspects of an object do you want to preserve?
- What are the requirements?
- How to prove in 10, 20, 50, 100 years, that the decision was correct / acceptable at the time it was made?

Preservation Planning

What is Preservation Planning?

- Consistent workflow leading to a preservation plan
- Analyses, which solution to adopt
- Considers
  - preservation policies
  - legal obligations
  - organisational and technical constraints
  - user requirements and preservation goals
- Describes the
  - preservation context
  - evaluated preservation strategies
  - resulting decision including the reasoning
- Repeatable, solid evidence

Digital Preservation

What is a preservation plan?

- 10 Sections
  - Identification
  - Status
  - Description of Institutional Setting
  - Description of Collection
  - Requirements for Preservation
  - Evidence for Preservation Strategy
  - Cost
  - Trigger for Re-evaluation
  - Roles and Responsibilities
  - Preservation Action Plan

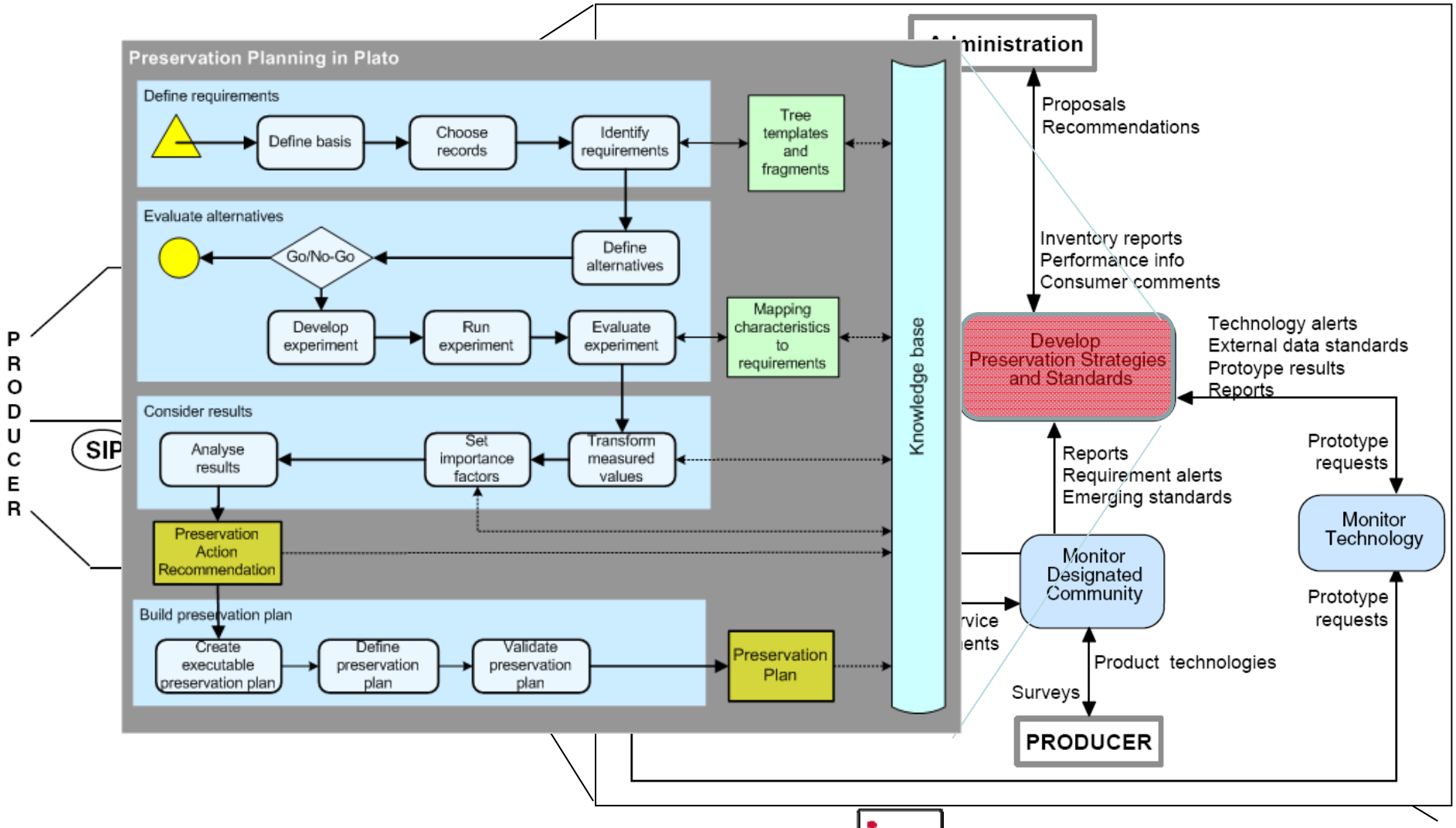
Preservation Plan Template

Preservation Planning

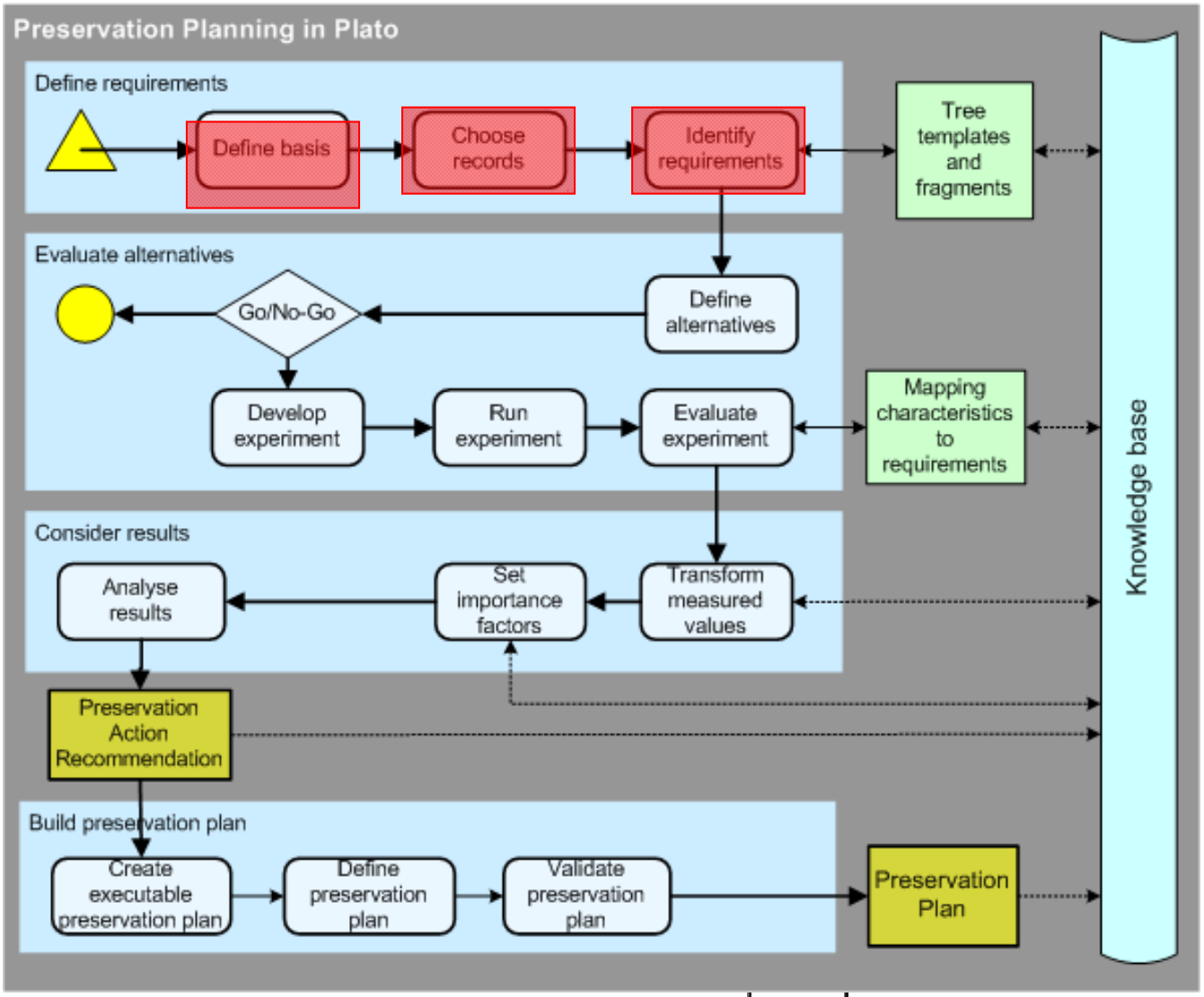
Preservation Planning Workflow

- Originally developed within the DELOS DP Cluster
  now refined and integrated within PLANETS,
  extended within SCAPE
- Based on
  - Preservation Planning approach based on Utility Analysis, developed at TU Vienna
  - Testbed/lab for evaluation developed at Nationalarchief, The Netherlands
- Follows the OAIS model
- Consistent with requirements specified by ORLC/TRAC and Nestor criteria catalogue

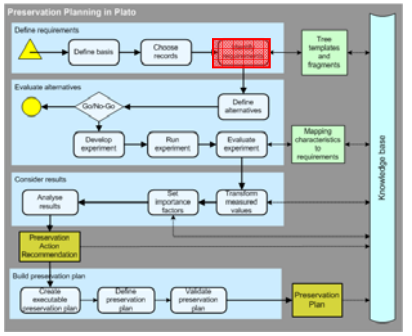
Preservation Planning

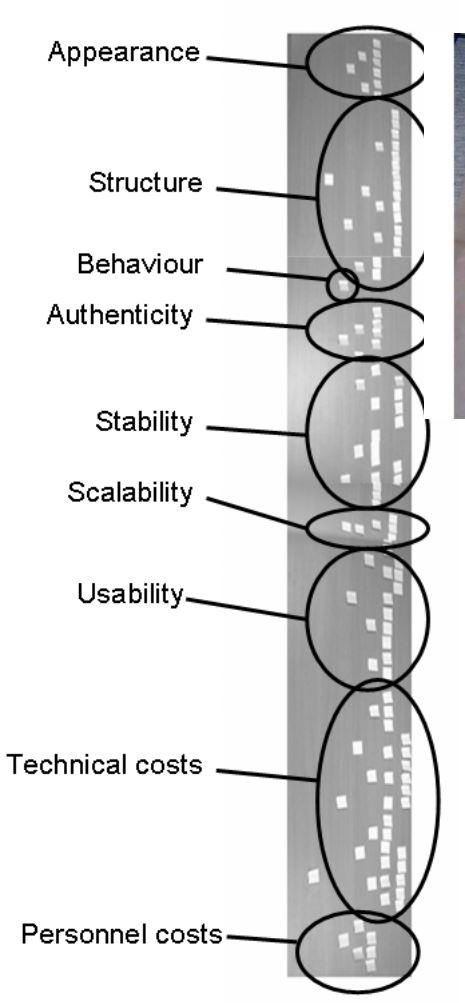


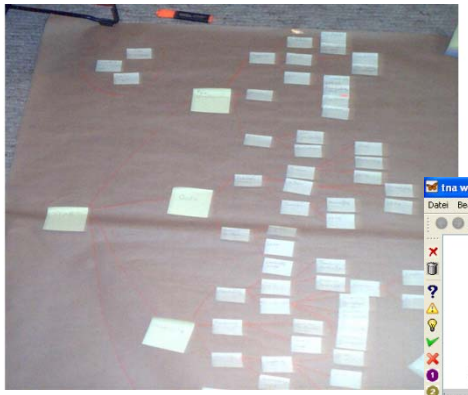
Preservation Planning Workflow



Identify requirements

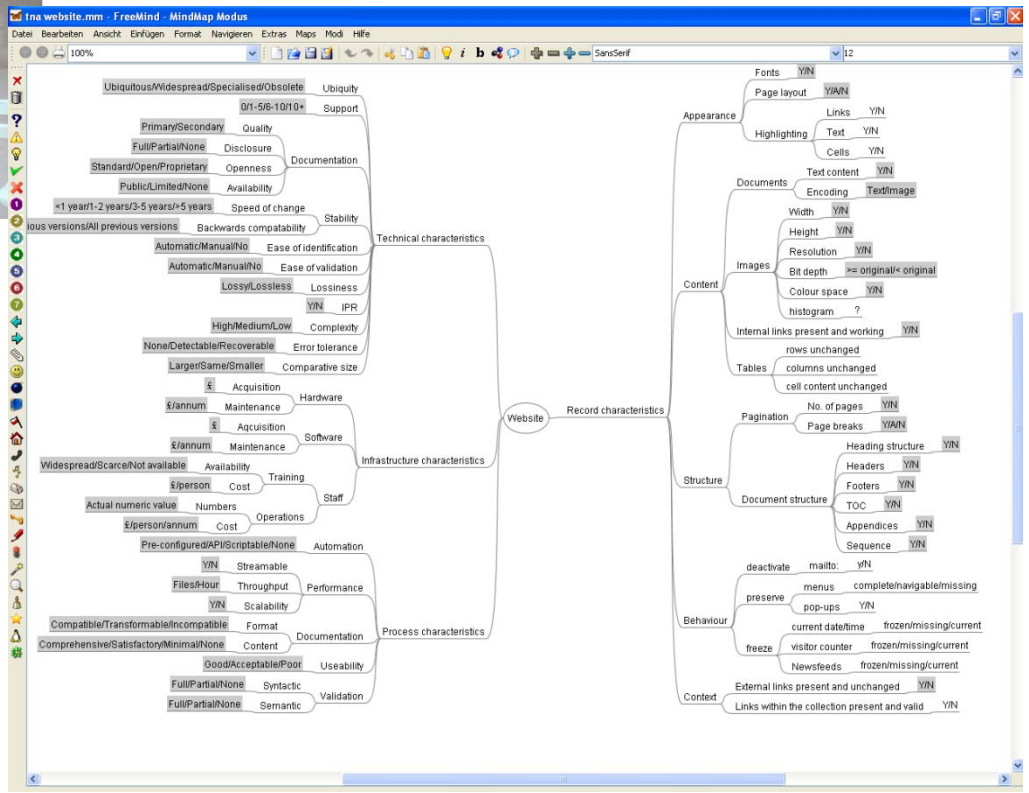




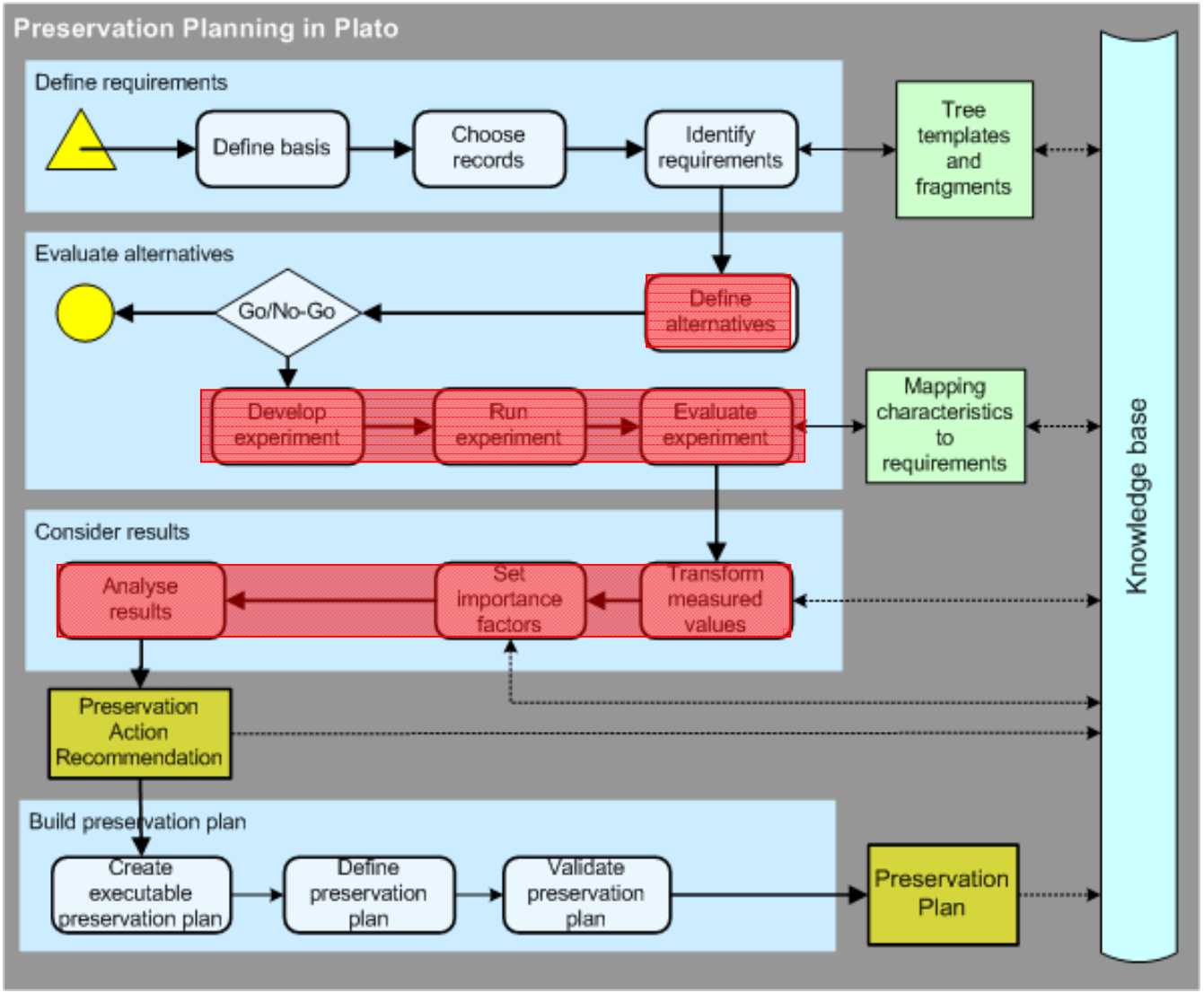


Analog…

… or born digital



Preservation Planning Workflow



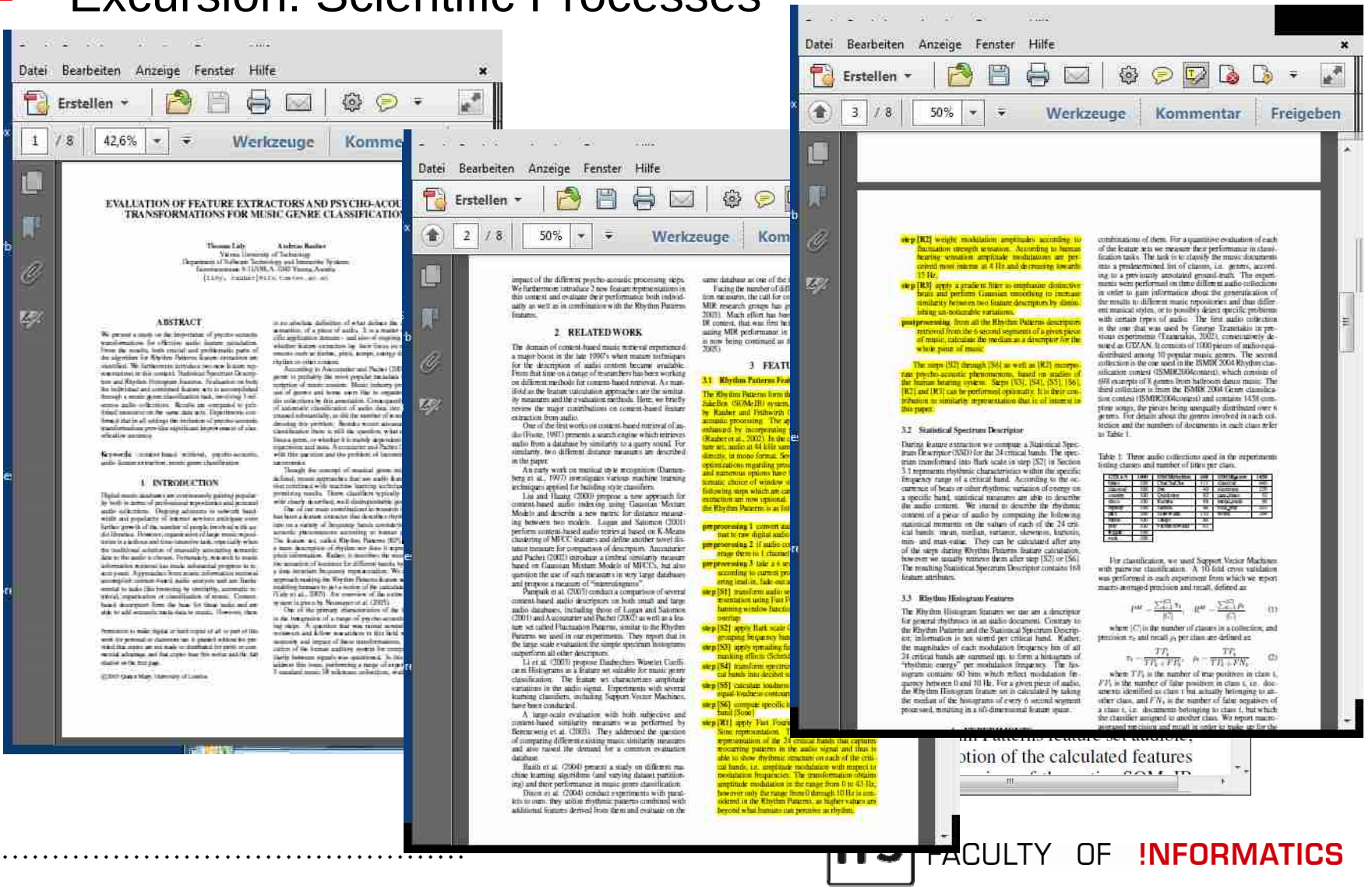
Overview

Part 1: Introduction

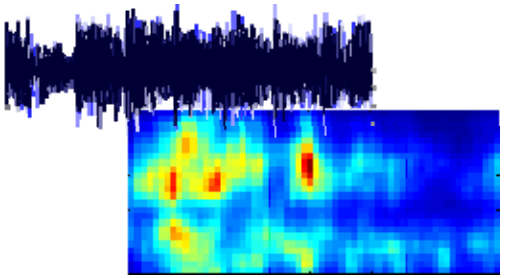
- What is Digital Preservation?
- What is the OAIS Reference model?
- How do we build a preservation plan?
- From Data to Processes
- Other issues in DP?

From Data to Processes

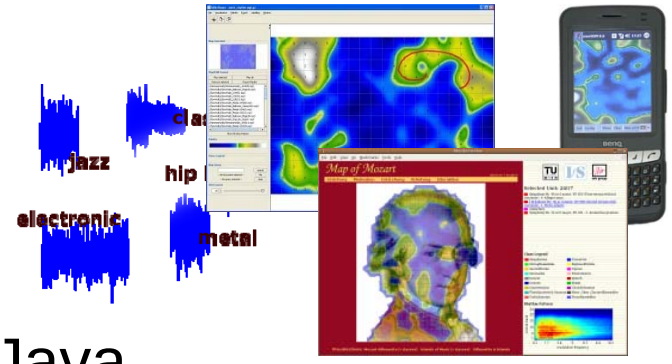
- Excursion: Scientific Processes



From Data to Processes

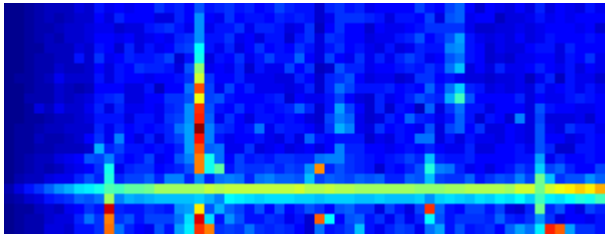


- Rhythm Pattern Feature Set
  - extracts numeric descriptors from audio
  - basically 2 Fourier Transforms
  - some psycho-acoustic modelling
  - some filters (gaussian, gradient) to make features more robust
- Used for
  - music genre classification
  - clustering of music by similarity
  - retrieval
- Implemented first in Matlab, then in Java
  - both publicly available on website
  - same same but different...

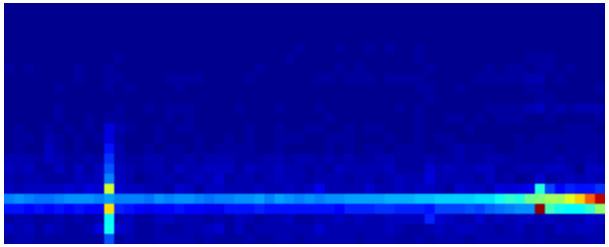


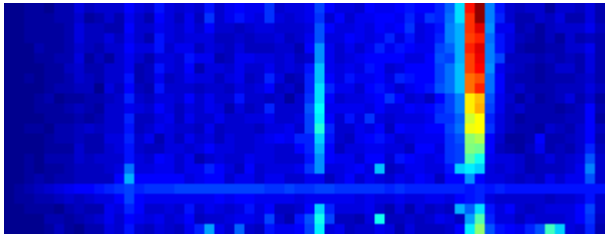
From Data to Processes

- Excursion: scientific processes

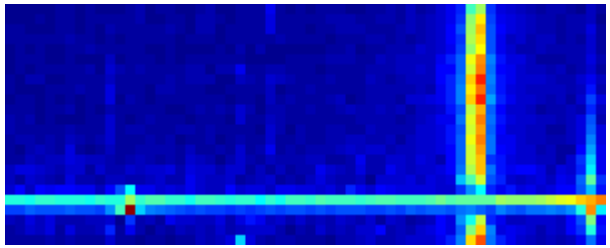


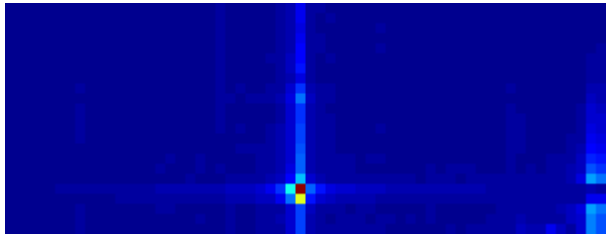
set1_freq440Hz_Am11.0Hz



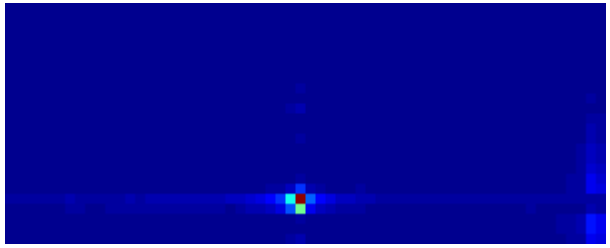


set1_freq440Hz_Am12.0Hz





set1_freq440Hz_Am05.5Hz

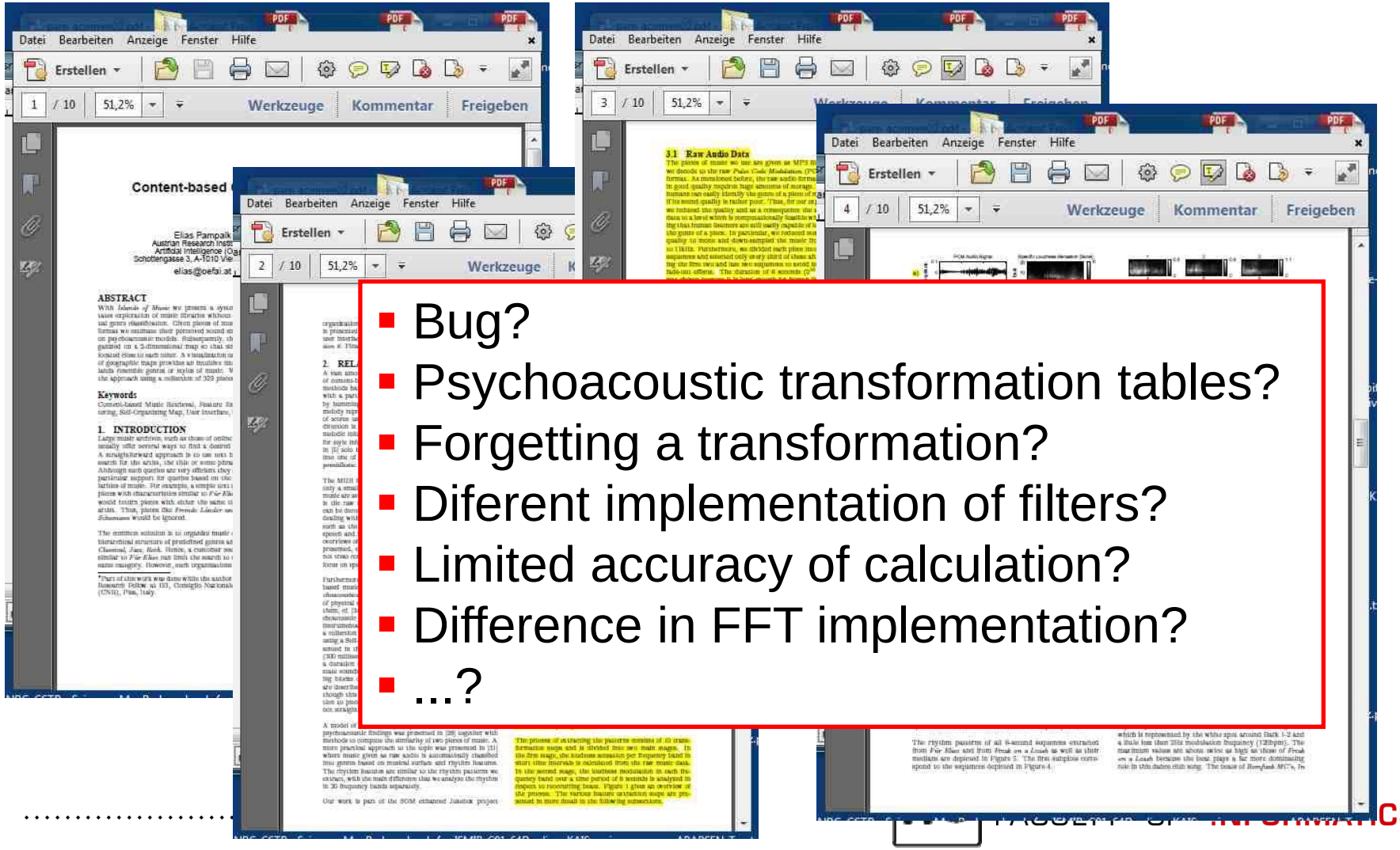


Java

Matlab

From Data to Processes

- Excursion: Scientific Processes



From Data to Processes

- Processes are important to understand data!
- Processes include
  - sensor capture (type, A/D conversion, calibration, operating conditions)
  - data (pre)processing: filtering, transformation
  - data integration: sources, transformations, treatment of missing values, outlier detection, ...
  - data analysis: tools, parameters, determinism
  - human operator activities
  - external services, web services
- End-to-end chain of activities underlying scientific experimentation
- Data as (interim) results

From Data to Processes

- Different disciplines of science, different means of validation
  - formal / proof
  - discourse
  - experimental evidence
- Many ICT-driven research areas experiment-driven
- How good are we in terms of repeatability/verifyability?
- Can we re-use earlier studies? verify code? share data?
- Need to ensure better procedures .....
  .....to support better science!

From Data to Processes

- How to curate processes?
  - how to capture and describe them?
  - what about proprietary elements?
  - how to evaluate if curation/re-activation is successful?

    (sig-props for processes and how to measure)
  - how can we cite data used in experiments?

Overview

Part 1: Introduction

- What is Digital Preservation?
- What is the OAIS Reference model?
- How do we build a preservation plan?
- From Data to Processes
- Other issues in DP?

Current Issues

- Personal & SOHO Archiving
  - What are DP requirements of SMEs? consumers?
  - Are there options for a service-based model?
  - Trust?
- Web Archiving: DP, IR & Ethics
  - How to capture the web?
  - Shall we do it? Privacy? Cultural heritage?
  - A basis for understanding society, knowledge, ... everything?
- From Documents to Interactive Content to Processes
  - Do static documents still exist?
  - Death of the file format?
  - How to preserve business processes?

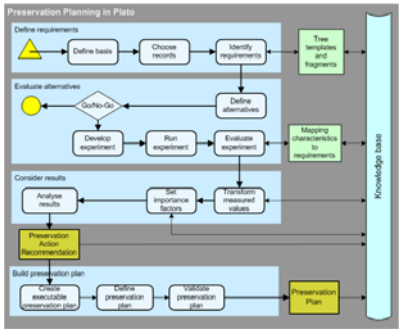
Current Issues

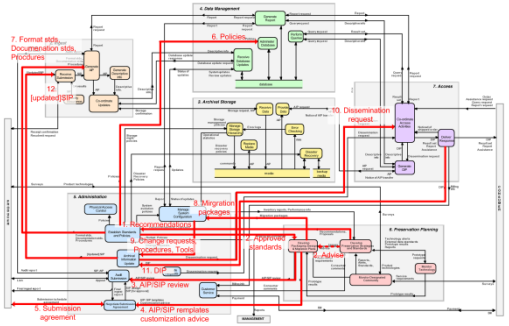
- Context of objects
  - What is a digital object?
  - What is the context of an object?
  - What is the context of a process?
- Security
  - What are the challenges in long-term signatures? Why does a simple signature not work?
  - How can we prove authenticity?
  - How does secure logging work?
- Domain-specific challenges
  - What are the needs of construction industry?
  - Airline industry?
  - Medical domain? (DICOM,...)

Current Issues

- Atomic file formats, stability of file formats
  - What are the atomic building blocks of information?
  - Can we split information objects?
  - Can we synthesize them? - Help for benchmarking?
- Scalability, Semantics
- Digital forgetting
  - how to decide what to keep and what to forget?
  - keep all? just storage? how to find? utilize? unterstand?
- Sustainable Systems Engineering
  - How can we build preservation-ready systems?
  - How to integrate DP-considerations into software engineering?
- Costs: what does DP cost?
  - cost factors?
  - How to model? evaluate?

Thank you!





http://www.ifs.tuwien.ac.at/dp