

Data Management Plans - How to Treat Digital Sources

The imminent future for repositories and their management

Paolo Budroni

Library and Archive Services, University of Vienna

Tomasz Miksa

Secure Business Austria

Andreas Rauber

Vienna University of Technology

Why a Data Management Plan?

[The management of of digital objects (e.g. articles, datasets, images, stream of data, multimedia objects) and non-digital objects (namely real-world entities, like authors, institutions but also teams, geographic locations and so on) is becoming a crucial issue for the whole scientific and administrative process within an academic institution. The need to unambiguously locate and access digital resources, associate them with the related authors and other relevant entities (e.g. institutions, research groups, projects, administrative organisational units) is becoming essential to allow the management, accessing, reuse, retrieval and preservation of huge amounts of cultural and intellectual resources. The rapid increase of digital assets in recent years has increased the relevance of this issue, making clear that **academic institutions need to plan the way on how they will achieve these goals, because the making of a data management plan helps to save time and energy and makes any kind of process easier.**]

DMP – Define a framework for creating data management

The main purpose of these reflection is to define and offer a model in order to define a comprehensive data management plan that

- can be used by the content providers (even at the stage of a project proposal),
- by the technical management,
- by the repository management and
- by the reusing parties

DMP – Define a framework for creating data management

It concerns...

Preingest and ingest workflow

Repository management

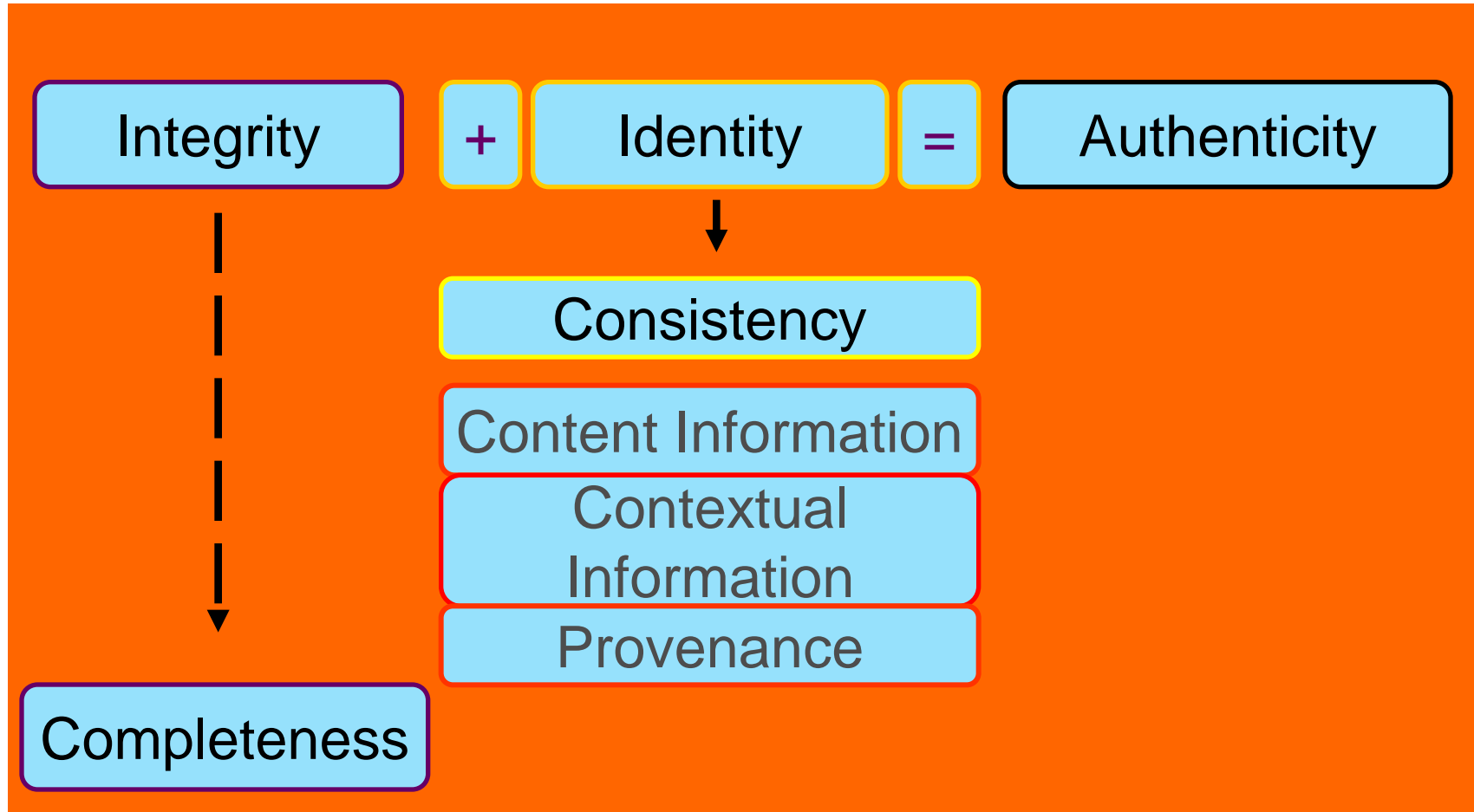
Reusing parties (export, visualization, representation)

Throughout activities, mainly all kind of processes

->> PROCESS MANAGEMENT PLAN

1st: Focus on the Digital Object. The Purpose of Preservation.

Main Goals Concerning the Object



Main Goals of the Preservation Process

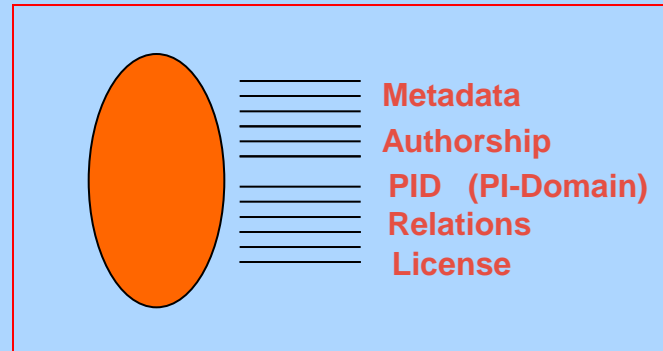
The „Document“ (a Digital Object)

1. Find, identify, select, obtain the document

2. Preserve the identity and the integrity of the document

3. Share knowledge

Properties of a digital object



At least one *Persistent* Identifier

Generated by an author

Relation to further digital objects

Relation to a policy

Deposit Licence

Metadata are assigned

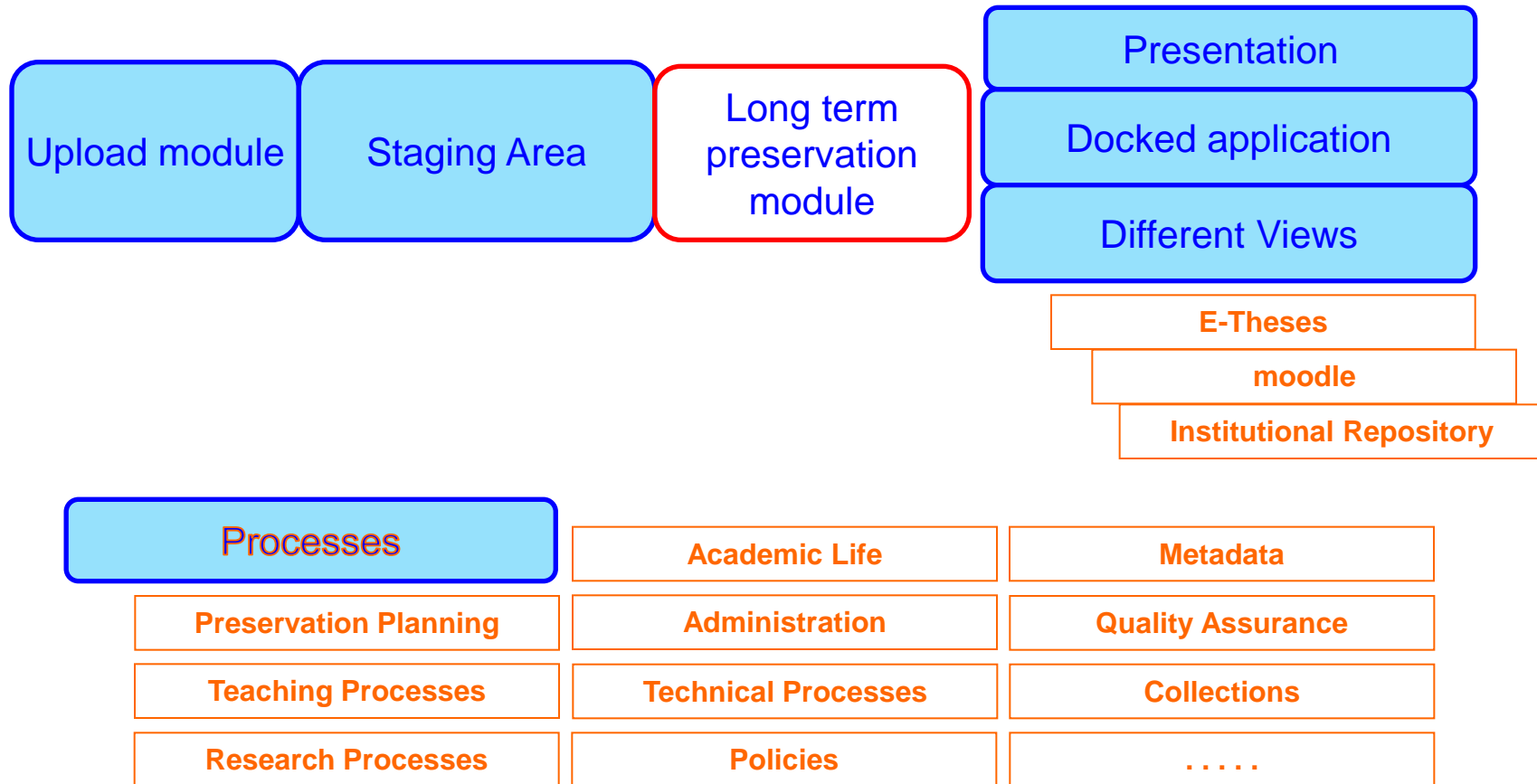
The „Document“ (a Digital Object)

2nd: Preservation Processes

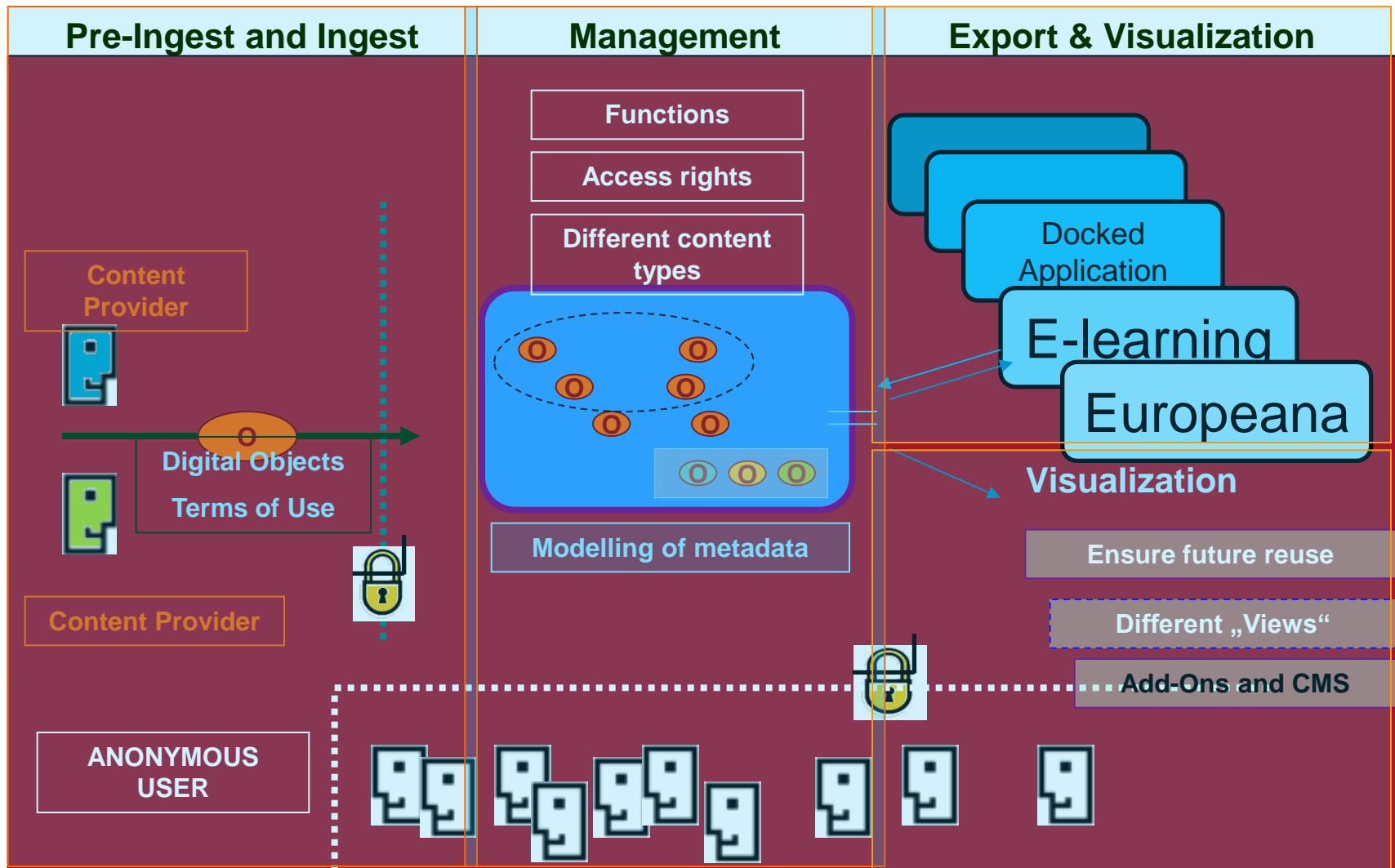
Process Management Plans

- quickly emerging concept
- process centric view
- data is the result of the processes such as
 - capture, (pre-) processing, transformation, integration and analyses
- facilitates the exchange and re-use of
 - whole processes or parts in other experiments
 - decreasing development time
 - lowering costs
- Design of solutions for DMPs should allow easy introduction of PMPs

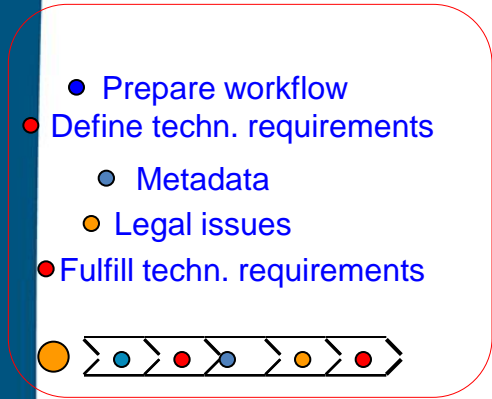
The Processes /1



Processes /2



Process Management Plan, a Use Case: „Use and reuse of digital resources, objects, assets“



Standards, procedures, code of conduct

Ingest

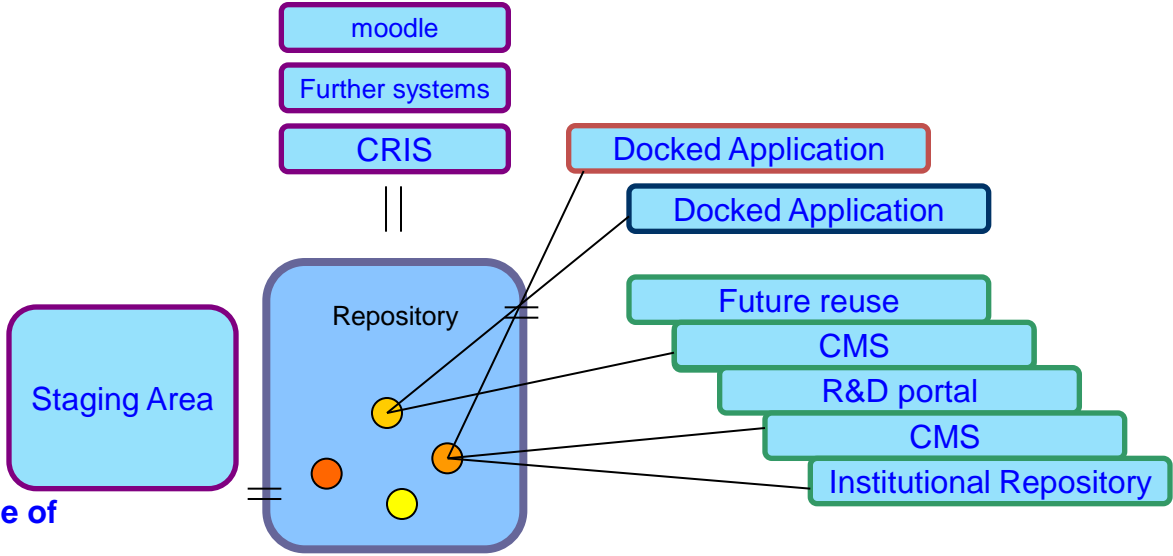
- Open Source
- Generic Metadata-editor
- Validation process
- Creation of templates
- Modules

New definition of project proposals

- Definition workflow
- Resources Pres. Planning + LTP
- Reuse of data
- (design, modelling of metadaten, development)
- Resources f. assimilation of data
- Resources for publication services

Visualization

- Functionalities
- Project based visualization
- Visualization of relations
- Different possibilities of enrichment



3rd: Data Management Plan The Points of View

A Comprehensive Data Management Plan: Views

Point of view of a project proposal

Point of view of the content provider/ owner of the object

Point of view of the funding agency

Point of view of the repository management

Point of view of the service providing institution

Point of view of the reusing parties

Point of view of the anonymous user

> **One user, different roles** <

4th: Further Developments

Further Developments /1

Further Issues Concerning Digital Objects
and DMP:

*Ontology of Objects, Formats, Content Model,
Content Types,
Presentation Type (Visualization Issues)*

Further Developments /2

Questions Concerning:
Preservation Issues
Intellectual Entity and Digital Preservation
Life Cycle Management

Further Developments /3

Arising Question:
*Purpose of Preservation vs.
Purpose of Presentation ?*

Further Developments /4

Which Representation is the Target Data Management Plan /Data Sharing Plan ?
(What Kind of Presentation is Intended in Five-Years' Time?)

Further Developments /5

Analysis of Existing Approaches to DMPs Checklists

Existing approaches to DMPs

National Science Foundation (NSF) in US

DataRes Project

<http://datamanagement.unt.edu/>

Australian National Data Service (ANDS) in
Australia

<http://ands.org.au/guides/data-management-planning-awareness.html>

Digital Curation Centre (DCC) in Great Britain

<http://www.dcc.ac.uk/resources/data-management-plans>

DCC Checklist

1. Introduction and Context
2. Data Types, Formats, Standards and Capture Methods
3. Ethical and Privacy Issues
4. Access, Data Sharing and Reuse
5. Short-Term Storage and Data Management
6. Deposit and Long-Term Preservation
7. Resourcing
8. Adherence and Review
9. Statement of Agreement
10. Annexes

5th: What is Needed ?

Go Further Existing Approaches

- *What has to be deposited?*
- *Data sharing*
 - *What has to be shared?*
 - *Output only (produced result of calculation) ?*
- *Metadata*
 - *Software environment*
 - *Software version*
 - *Etc.*
- *Process of data creation is vaguely described*

Multi-phase DMP Development

Example from: Data Management Services at
John Hopkins University in Baltimore (US)

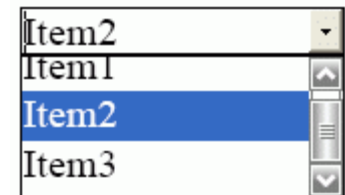
<http://dmp.data.jhu.edu/>

Phases

- identification of researcher needs
- Preparation of DMP for grant proposal
- Post-award DMP detailed specification and mapping of services provided by JHU
- Revisions of DMP during project's lifetime

ComboBox-Style DMP

- Researcher selects answers from a defined set of possible answers (as much as possible)
- Decrease in preparation time
- Higher coherence – less verbal ambiguities
- Higher precision – no facts omitted
- Machine readable
- Automatic checking
- Query possibility



Actionable DMP

- More than ComboBox-Style
- Not only a digital version of print-out
- Partially executable and interpreted by machine
- Basic example
 - DMP specifies two back-up locations
 - Actionable DMP has links to these locations and can verify their availability

Enforceable DMP

- Enforced compliance
- Technical
 - Automatic checking
 - Actionable DMP
- Organizational
 - Penalties
 - Withholding of funding
 - Incentives

6th: Conclusions

Consider the Views

Point of view of a project proposal

Point of view of the content provider/ owner of the object

Point of view of the funding agency

Point of view of the repository management

Point of view of the service providing institution

Point of view of the reusing parties

Point of view of the anonymous user

> **One user, different roles** <

Conclusions

- Base the DMP on existing research
- Keep in mind
 - Actionability
 - Enforceability
- Consider in parallel the design of
 - Data Management Plan
 - Process Management Plan
 - Policy
 - Assessment Criteria

Conclusions

- Think digital and consider the custody chain
- Each content type has different metadata sets
- Each content type should be represented differently
- Which representation of the object is the target of preservation?
- Data plans are an integral part of archiving policies (starting at the stage of the grant application)

Further issues

- Different Role of Users
- Processes and Tasks
- DMP: Analysis of the whole Life Cycle
- Process based reengineering of the workflow

COAR Annual Meeting, May 2013

Presented by: Paolo Budroni in Istanbul,
Library and Archive Services, University of Vienna
Head of Project „Phaidra“

Thanks

Questions?