Track 3 Workshop

ELIANE BLUMER
JAN MELICHAR
HUGUES CAZEAUX

June 8, 2016

DLCM | swissuniversities

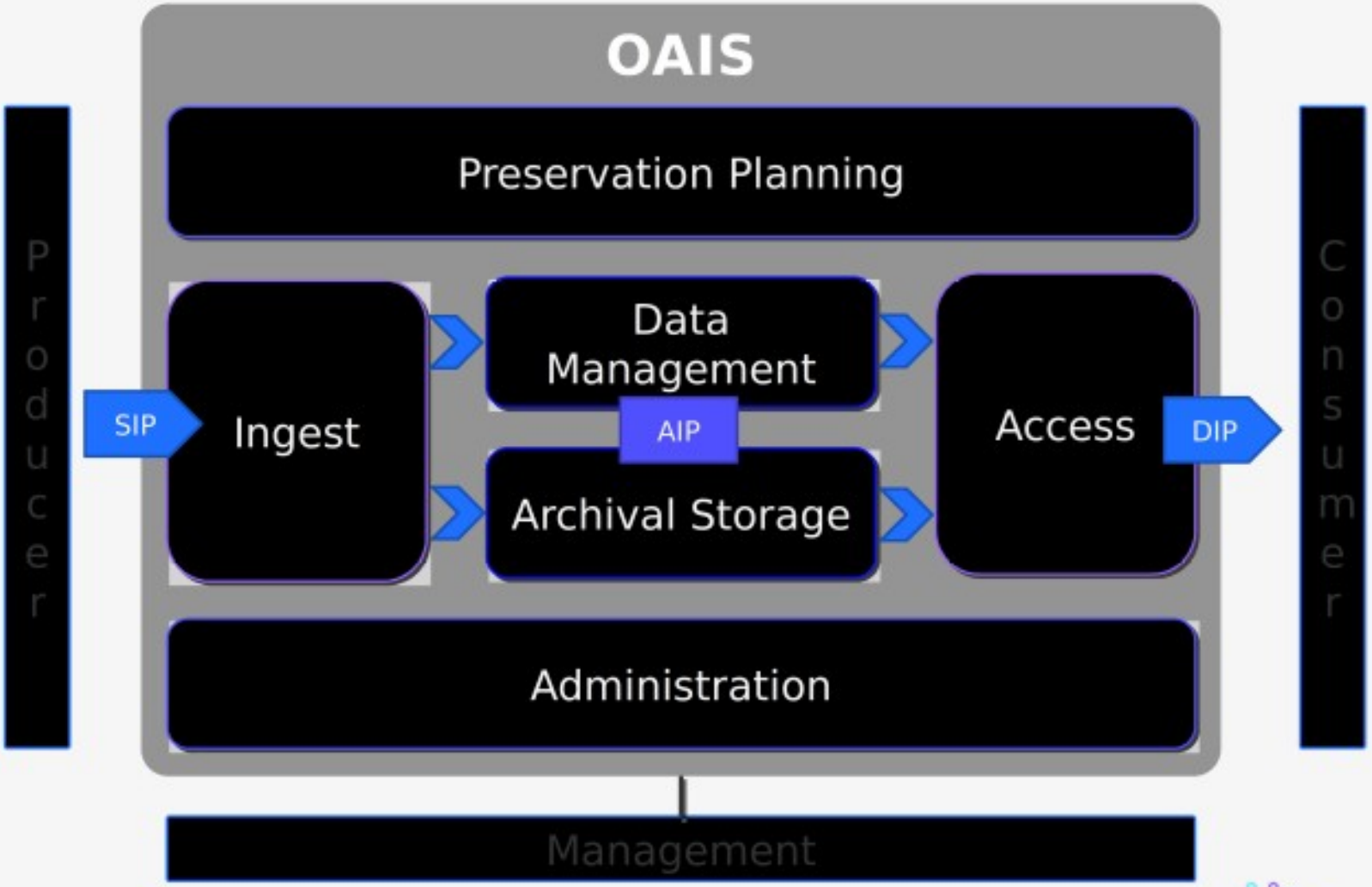
Workshop Details

- **Practical Information**
  - *Date:* June 8, 2016
  - *Time:* From 10:15 to 16:15
  - *Location:* University of Bern, room 331 3.OG/West
- **Participant Homework**
  - Read different gap analysis
  - Define “What is Archiving for you?”
  - Read slides
    - *OAIS Model (#4)*
    - *Glossary (#5)*
  - Read Section 6 of OAIS standard (http://public.ccsds.org/publications/archive/650x0m2.pdf)
- **Agenda**
  - Introduction (15 min)
  - Part 1 - Gap Analysis Review (1h30)
  - Lunch (12h-13h)
  - Part 2 - Project Alignment (2h)
    - *Archiving Positioning*
    - *DLCM Collaboration Discussion*
    - *Metadata Discussion*
  - Part 3 – Project Management (1h)
    - *Deliverables*
    - *Planning*
    - *Tasks*

Agenda Details

- **Introduction**
  - Participant round table
    - *Each participant introduces herself/himself*
  - Communication Channels
    - *Single point of contact by institution*
- **Part 1 - Gap Analysis Review**
  - Comments/Questions
  - Future expectations, evolutions or requirements?
- **Part 2 - Project Alignment**
  - Archiving Positioning
    - *Common & Shared definition of Archiving term*
    - *Archiving Actors?*
  - DLCM Collaboration Discussion
    - *What kind of "Archive Interoperability" do you expect in DLCM project?*
    - *Actor roles?*
  - Metadata Discussion
    - *What will be the minimal set of information to manage?*
    - *SIP / AIP / DIP?*
- **Part 3 – Project Management**
  - Deliverables
  - Planning
  - Tasks

OAIS Model


OAIS
Preservation Planning
Producer
SIP
Ingest
Data Management
AIP
Archival Storage
Access
DIP
Consumer
Administration
Management

OAIS Glossary

- **SIP = Submission Information Package**
  - An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.
- **AIP = Archival Information Package**
  - An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.
- **DIP = Dissemination Information Package**
  - An Information Package, derived from one or more AIPs, and sent by Archives to the Consumer in response to a request to the OAIS.
- ***DI = Descriptive Information***
  - *The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers.*
- **Designated Community**
  - An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time.
- **Producer**
  - The role played by those persons or client systems that provide the information to be preserved. This can include other OAISes or internal OAIS persons or systems.
- **Consumer**
  - The role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISes, as well as internal OAIS persons or systems.
- **Management**
  - The role played by those who set overall OAIS policy as one component in a broader policy domain, for example as part of a larger organization.
- **Access Aid**
  - A software program or document that allows Consumers to locate, analyze, order or retrieve information from an OAIS.
- **Finding Aid**
  - A type of Access Aid that allows a user to search for and identify Archival Information Packages of interest.

DLCM

Introduction

General Objective

To define functional scope of the track #3

Participants

Institutions	Participants
EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE	Aude Dieudé / Nathalie Lambeng Gaël Anex / Peter Laszlo Hliva
ETH zürich	Ana Sesartic / Matthias Töwe
Unil UNIL | Université de Lausanne	Hamid Hussain-Khan
Universität Zürich[UZH]	Martin Brändle / André Hoffmann / Christian Fuhrer Sergio Maffiolletti / Owen Appleton
UNIVERSITÉ DE GENÈVE	Eliane Blumer / P.-Y. Burgi / Jan Melichar Jean-Blaise Claivaz / André Jelicic / Hugues Cazeaux

DLCM

Part 1
Gap Analysis Review

Current Solutions

Institutions	Tools	Metadata
EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE	zenodo	MARC 21
ETH zürich	ExLibris Rosetta	METS
Unil UNIL | Université de Lausanne	Fedora	METS FOXML
Universität Zürich	eprints repository software	METS MODS
UNIVERSITÉ DE GENÈVE	Fedora PRESERVED IN LOCKSS	MARC 21 FOXML

POWRR Matrix

	Ingest					Processing						Access		Storage					Maintenance			Other		
digitalpowrr.niu.edu	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Bit Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
ETH Zürich	(X)	X	X		X	X	X	X	X	X	X	(X)	X	X	X	X	(X)	X	X	X			X	(X)
UNIBAS																								
UNIGE Archive/Fedora 3	X				X	X	X	X	X	X	X	X	X	X		X		X				X	X	free
UNIL Serval/Fedora 3	X	(x)	X	X	X	X	X	X	X	X	X	X	X	X		X	(x)	X	X	(x)		X	X	free
UZH/ZORA Eprints	(X)		X		X		X	X	X	X		X			(X)	X	X					X	X	free
Zenodo (EPFL)	S	X	X	U	X			X	X	X	S+U	X	S+U	S+U	X	X		X		X	X	X	X	free

DLCM

Part 2
Project Alignment

Archiving Positioning

“ What is archiving? ”

Archiving Positioning

- **Objective**

Common & Shared definition of Archiving term

- **Points to start the discussion**
  - Archiving vs Storage vs Backup
    - *What is the difference?*
  - What is the information to archive?
    - *Information*
      - Metadata
      - Content or Data
    - *Format*
      - Open or restricted list?
      - To archive also the program to read?

Archiving Positioning

- **Objective**

Common & Shared definition of Archiving term

- **Points to start the discussion (suite)**
  - Who are archiving Actors?
    - *Producer*
    - *Consumer*
    - *Management*

Archiving Positioning

OpenAIRE / EUDAT definition

- **Backup and preservation – not the same thing!**

- **Backups**
  - Used to take periodic snapshots of data in case the current version is destroyed or lost
  - Backups are copies of files stored for short or near-long-term
  - Often performed on a somewhat frequent schedule

- **Archiving**
  - Used to preserve data for historical reference or potentially during disasters
  - Archives are usually the final version, stored for long-term, and generally not copied over
  - Often performed at the end of a project or during major milestones

Archiving Positioning

Some Definitions

> **Records management** (**RM**), also known as the **records and information management** or **RIM**, is the professional practice of managing the **records** of an organization throughout their **life cycle**, from the time they are created to their eventual disposal. This includes **identifying**, **classifying**, **storing**, **securing**, **retrieving**, **tracking** and **destroying** or **permanently** preserving records.
>
> *Records management definition From Wikipedia*

> [1] **Documents** within the meaning of this Act are all **recorded information**, **irrespective of the medium**, that is received or produced in the fulfilment of the public duties of the Confederation, as well as all finding aids and supplementary data that are required in order to understand and **use this information**.
> [2] Archive records are documents that have been accepted by the Federal Archives for safekeeping or that are independently archived by other bodies in accordance with the principles laid down in this Act.
> [3] Documents that are of archival value are documents of legal or administrative importance or which **contain valuable information**.
>
> *152.1 Federal Act on Archiving – Art. 3*

> **Digital Archival**
> Set of actions aiming to **identify**, **capture**, **classify**, **preserve**, **retrieve**, **display** and provide access to **documents** for **informational** or historical purposes, or for **duration required** to meet legal obligations.
>
> *ISO 14641-1 Electronic archiving - §3.13 (NF Z42-020)*

DLCM Collaboration Discussion

“What kind of “Archive Interoperability” do you expect in DLCM project?”

DLCM Collaboration Positioning

- **Objective**

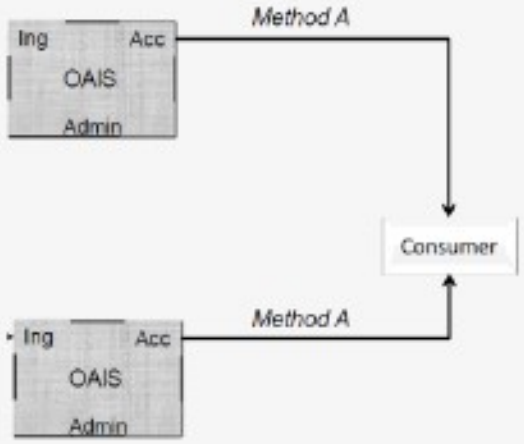
Discuss how archiving system will inter-operate or collaborate?

- **Points to start the discussion**
  - Different types of collaboration (OAIS – Section 6)
  - Other examples of collaboration
  - Actor roles

DLCM Collaboration Discussion

Interoperability Types

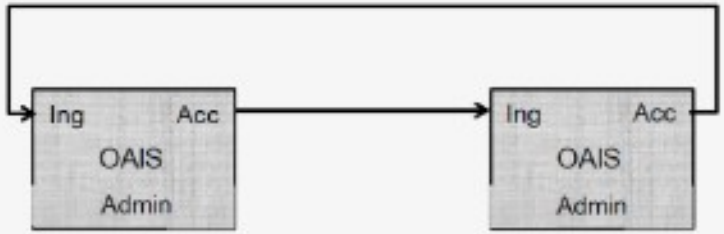
- Independent Archives (OAIS Page 6-2)
  - *Local concerns*
  - *Designed community*
  - *Common DIP & finding aids*
  - *Archive independent*

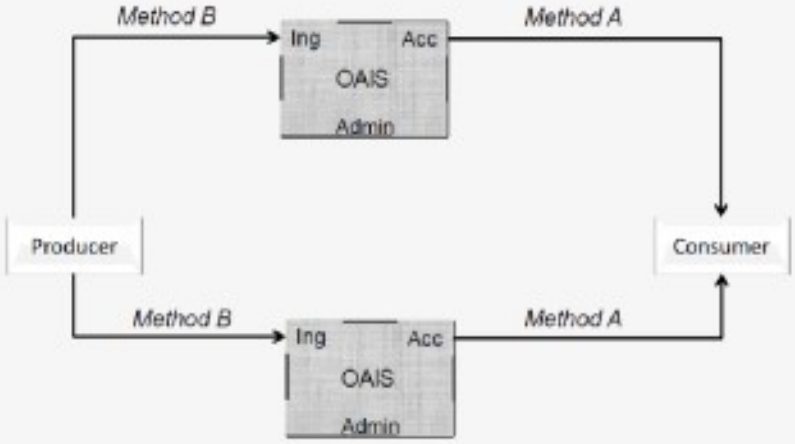


DLCM Collaboration Discussion

Interoperability Types

- **Cooperating Archives (OAIS Page 6-3)**
  - Common producers, submission standards or dissemination standards
  - No common finding aids
  - Standards agreements among archives
  - No common access, submission or dissemination
  - One compatible SIP/DIP

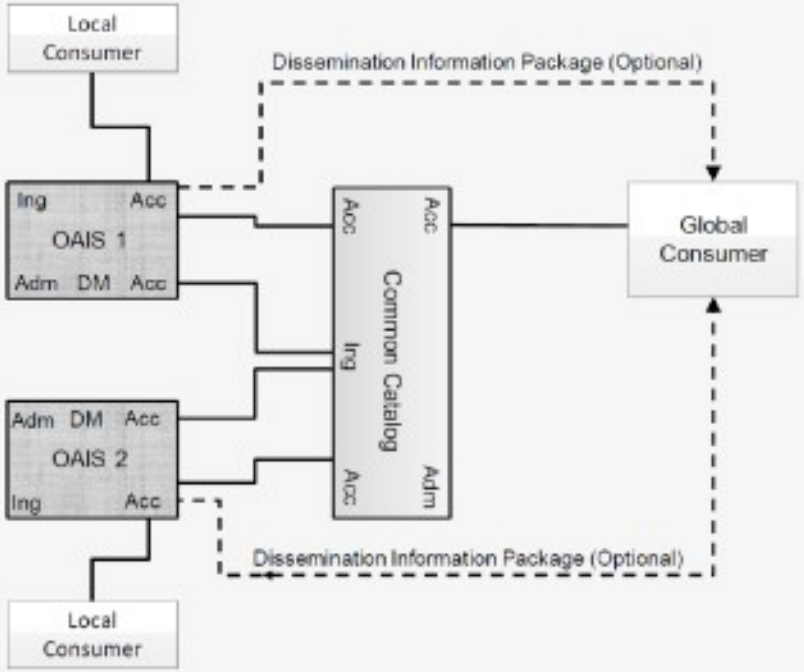




DLCM Collaboration Discussion

Interoperability Types

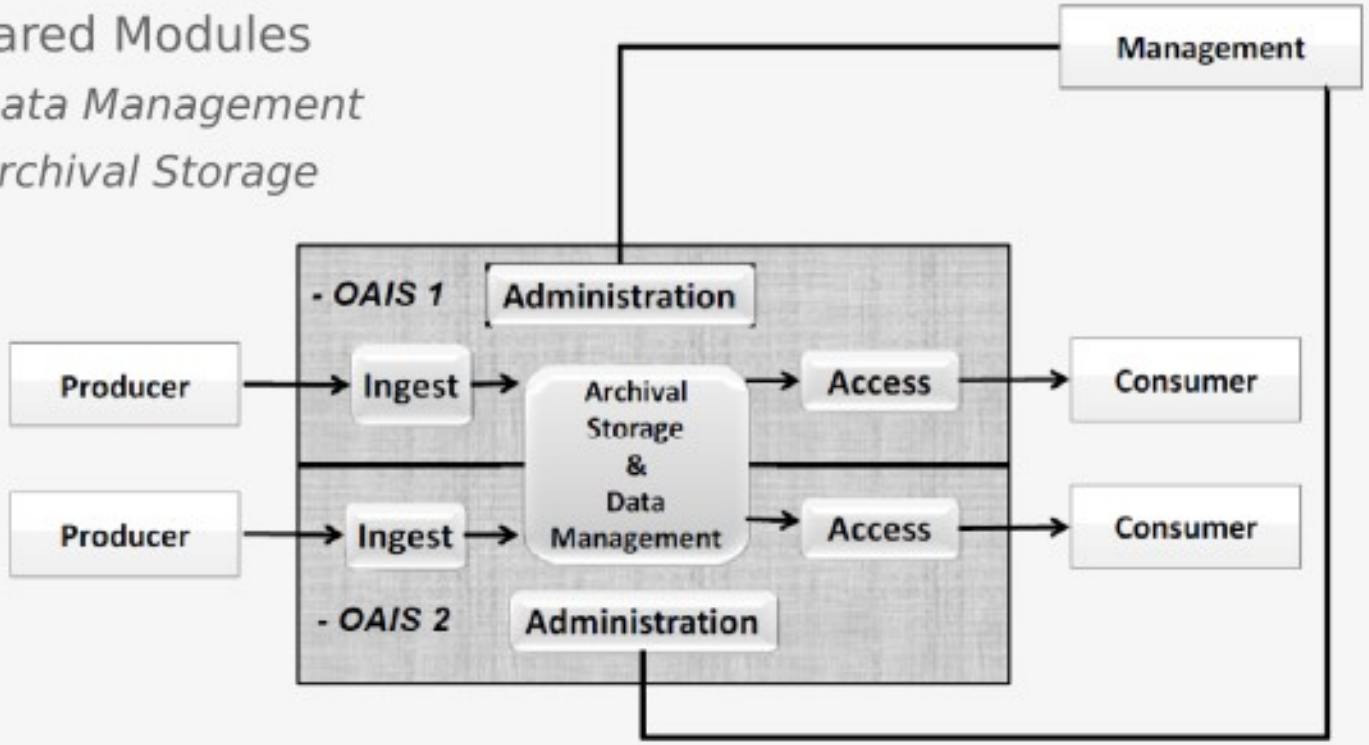
- **Federated Archives (OAIS Page 6-4)**
  - Local & global communities
  - Common finding aids
  - Global dissemination & Ingest optional
  - Consumer-oriented
  - Local Community
    - *Original Designated Community*
  - Global Community
    - *Extended Designated Community*
  - Challenges
    - *Unique AIP Names*
    - *Duplicate AIPs*
    - *Preservation of Federation Access to AIPs*
    - *User Authentication and Access Managem*



DLCM Collaboration Discussion

Interoperability Types

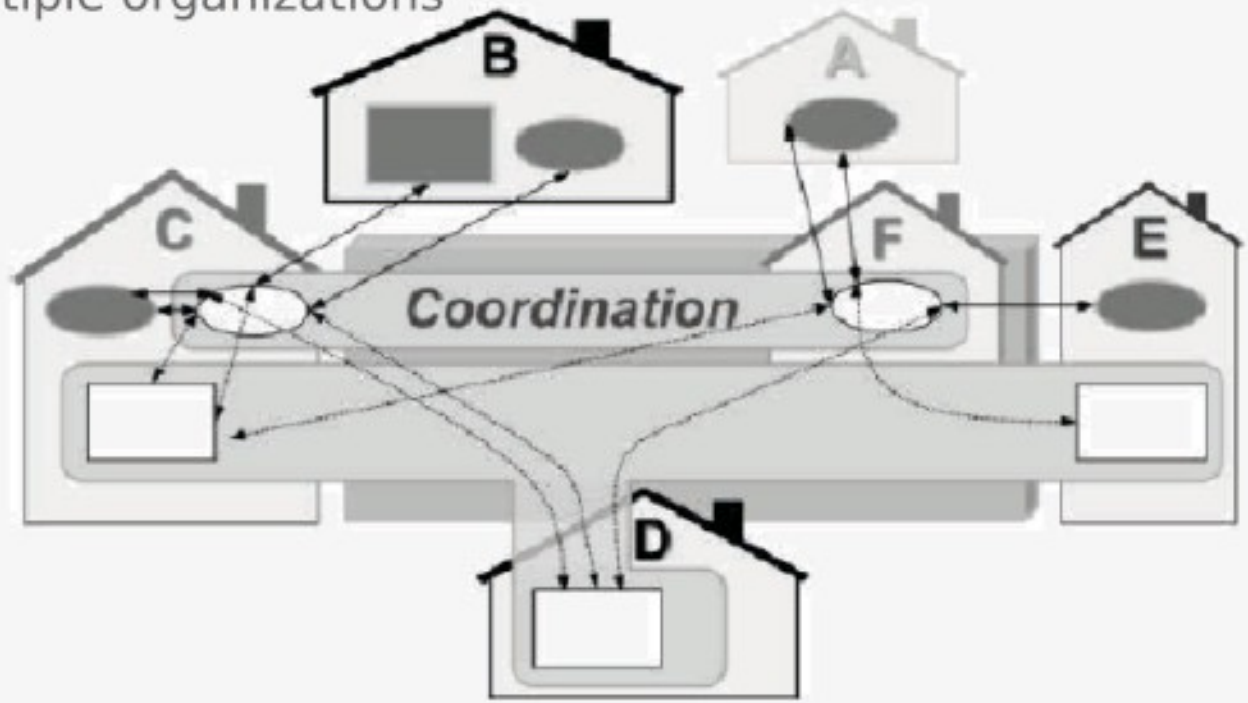
- **Shared Resources (OAIS Page 6-7)**
  - Ingest-storage and access-storage interface standards
  - No change on user community
  - Shared Management
  - Shared Modules
    - *Data Management*
    - *Archival Storage*



DLCM Collaboration Discussion

Interoperability Types

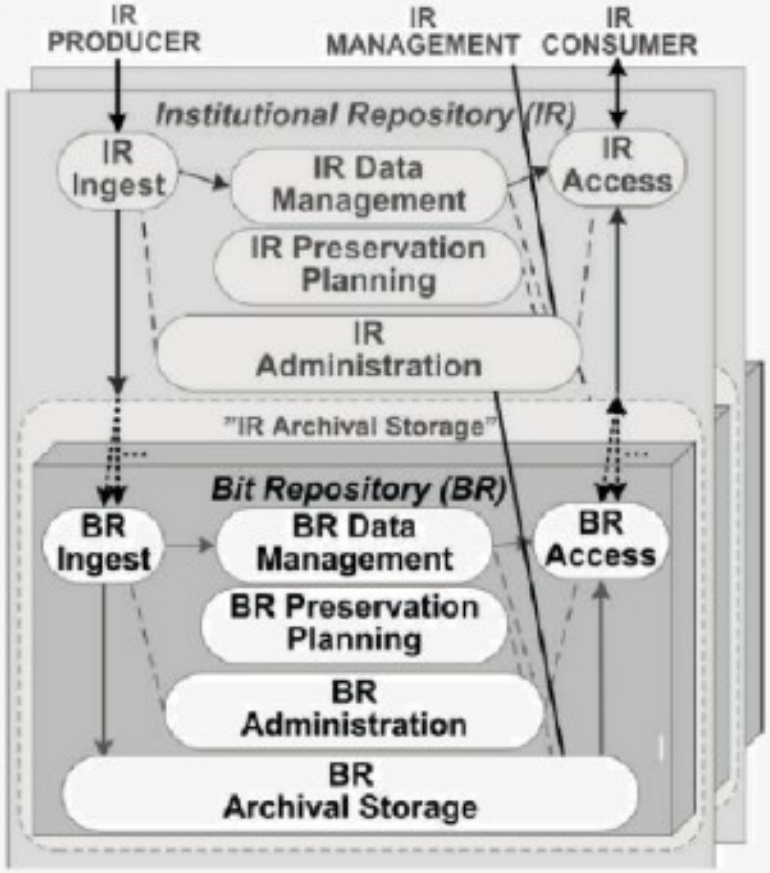
- **Distributed Organization**
  - Danish Bit repository platform
  - Multiple geographic locations
  - Multiple organizations



DLCM Collaboration Discussion

Interoperability Types

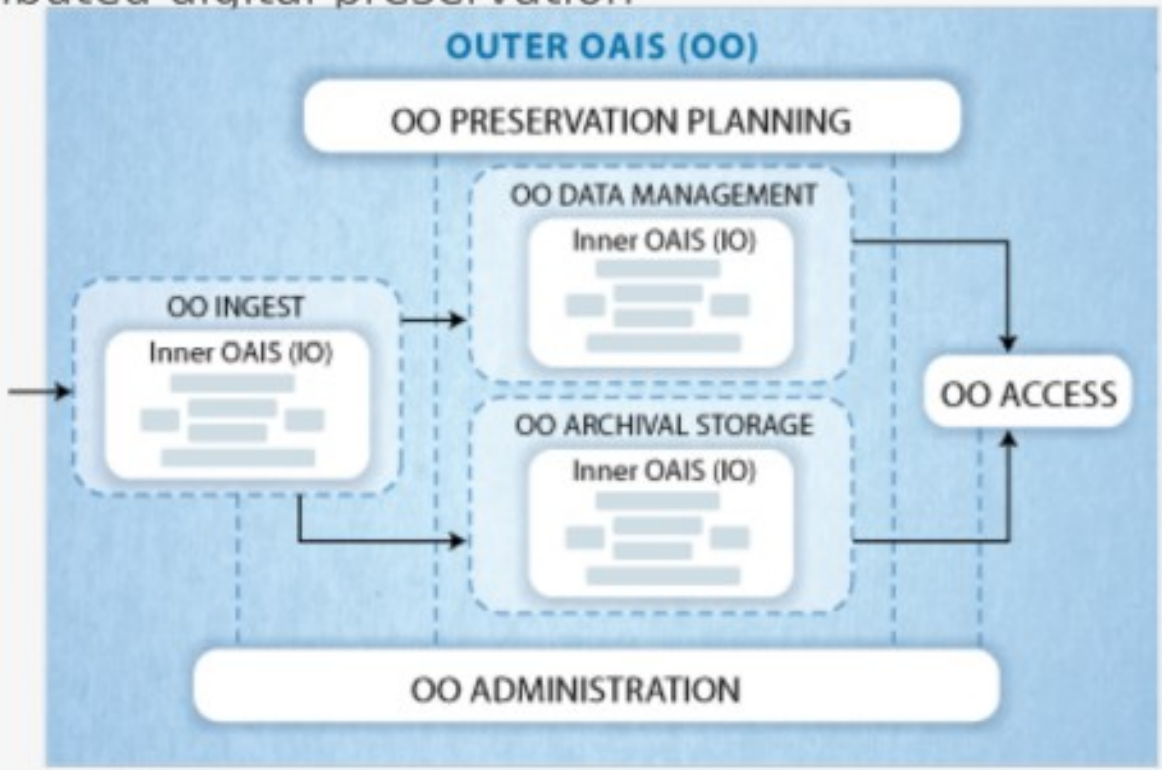
- **Institutional Repository – Bit Repository Model**
  - Danish Bit repository platform
  - Different implementation by institution
  - Storage managed at BR level



DLCM Collaboration Discussion

Interoperability Types

- **Outer OAIS-Inner OAIS Model**
  - Danish BitRepository.org
  - Built on IR-BR model
  - Support the specification and audit of collaborative interactions between multiple OAIS implementations for distributed digital preservation



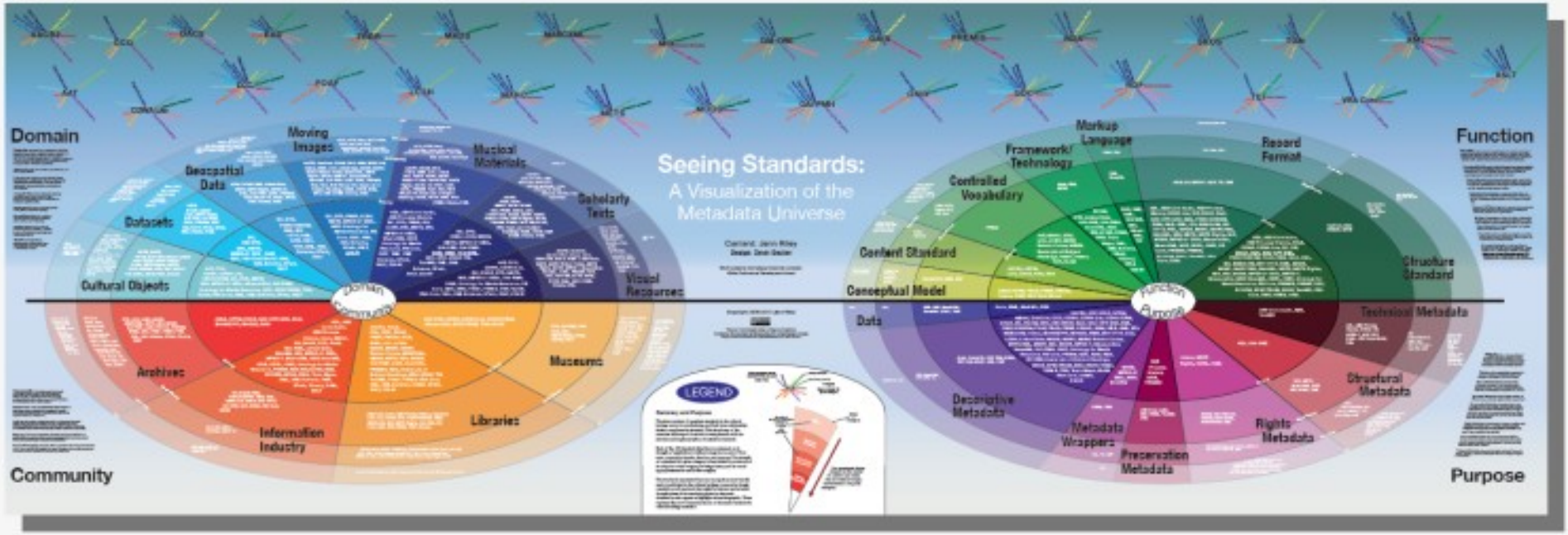
DLCM Collaboration Discussion

Interoperability Proposal

- Independent Archives?
- Cooperating Archives?
- Federated Archives?
- Shared Resources?
- Distributed Organization?
- IR-BR Model?
- OO-IO Model?
- Mixed?
- Hybrid?

Metadata Discussion

“Which Information?”



Metadata Positioning

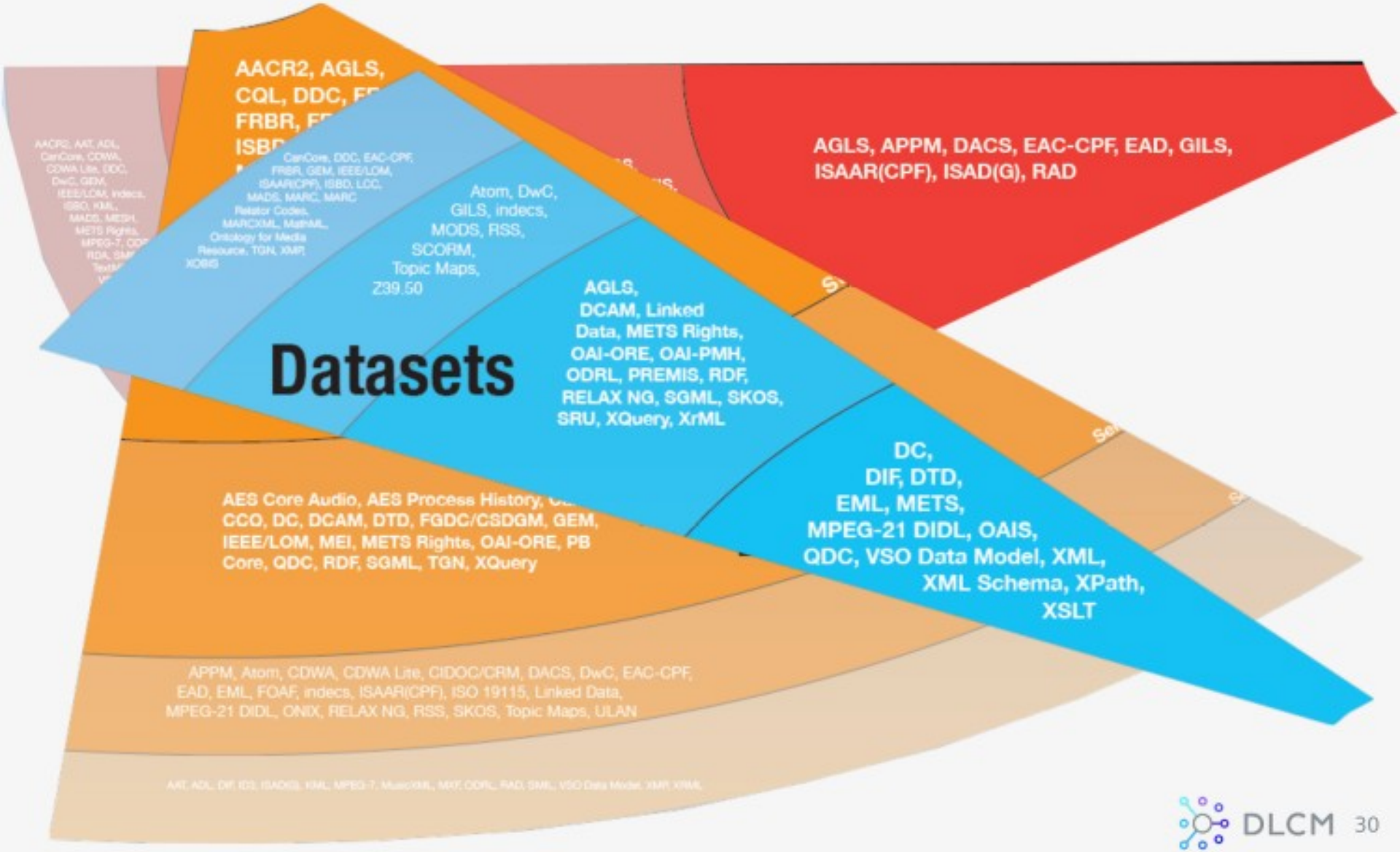
- **Objective**

Define short list of metadata set to support

- **Points to start the discussion**
  - View metadata poster
    - *By domain/function/format*
  - FAIR best practice

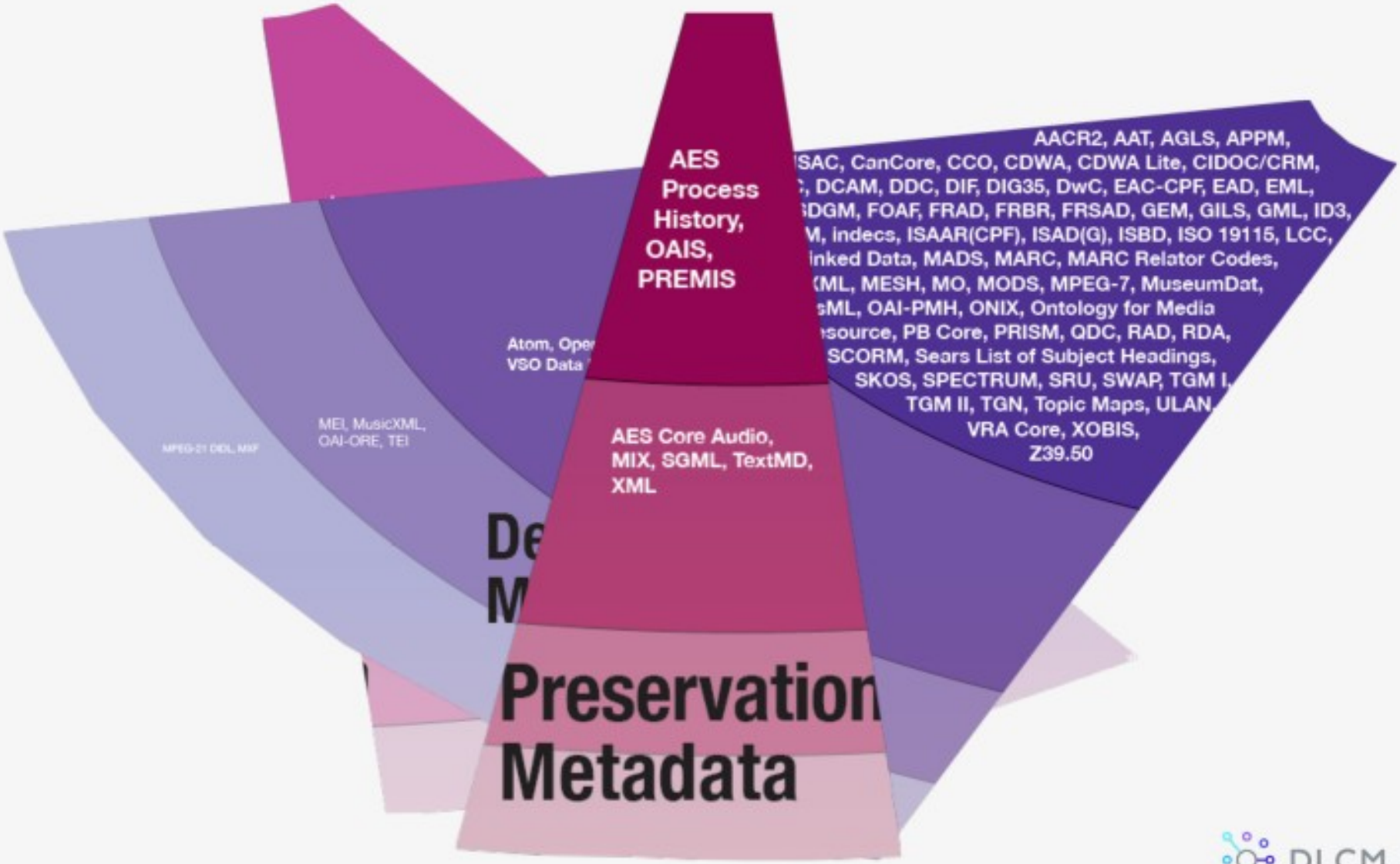
Metadata Discussion

by Domain



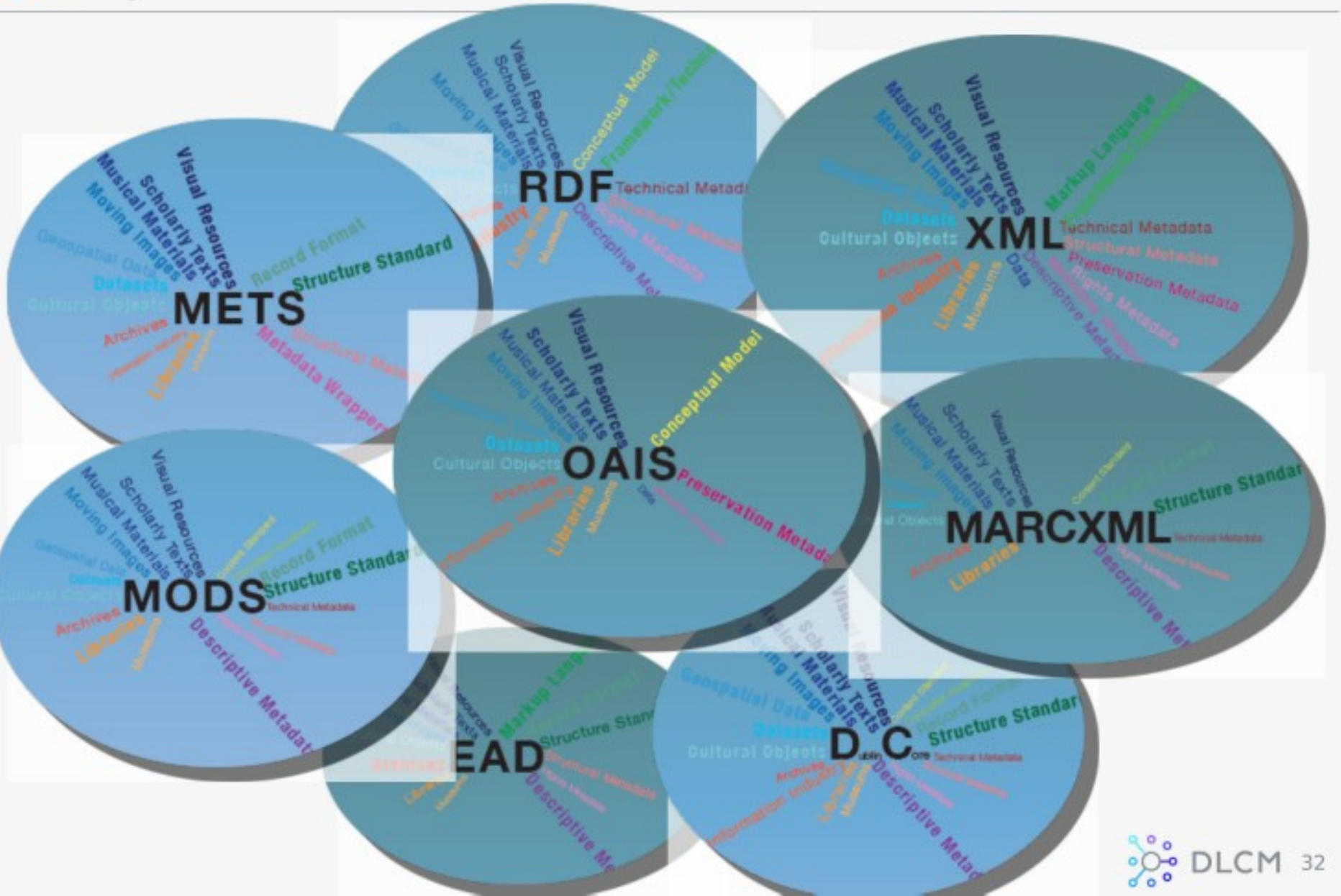
Metadata Discussion

by Function



Metadata Discussion

by Format



Metadata Discussion

- **FAIR Data Principles**
  - *Reference*: http://www.force11.org/group/fairgroup/fairprinciples
  - *Definition*: A set of guiding principles to make data **F**indable, **A**ccessible, **I**nteroperable, and **R**e-usable
- **To be Findable**
  - F1. (meta)data are assigned a globally unique and eternally persistent identifier
  - F2. data are described with rich metadata
  - F3. (meta)data are registered or indexed in a searchable resource
  - F4. metadata specify the data identifier
- **To be Accessible**
  - A1 (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
  - A2 metadata are accessible, even when the data are no longer available
- **To be Interoperable**
  - I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
  - I2. (meta)data use vocabularies that follow FAIR principles
  - I3. (meta)data include qualified references to other (meta)data
- **To be Re-usable**
  - R1. meta(data) have a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with their provenance
  - R1.3. (meta)data meet domain-relevant community standards

Metadata Discussion

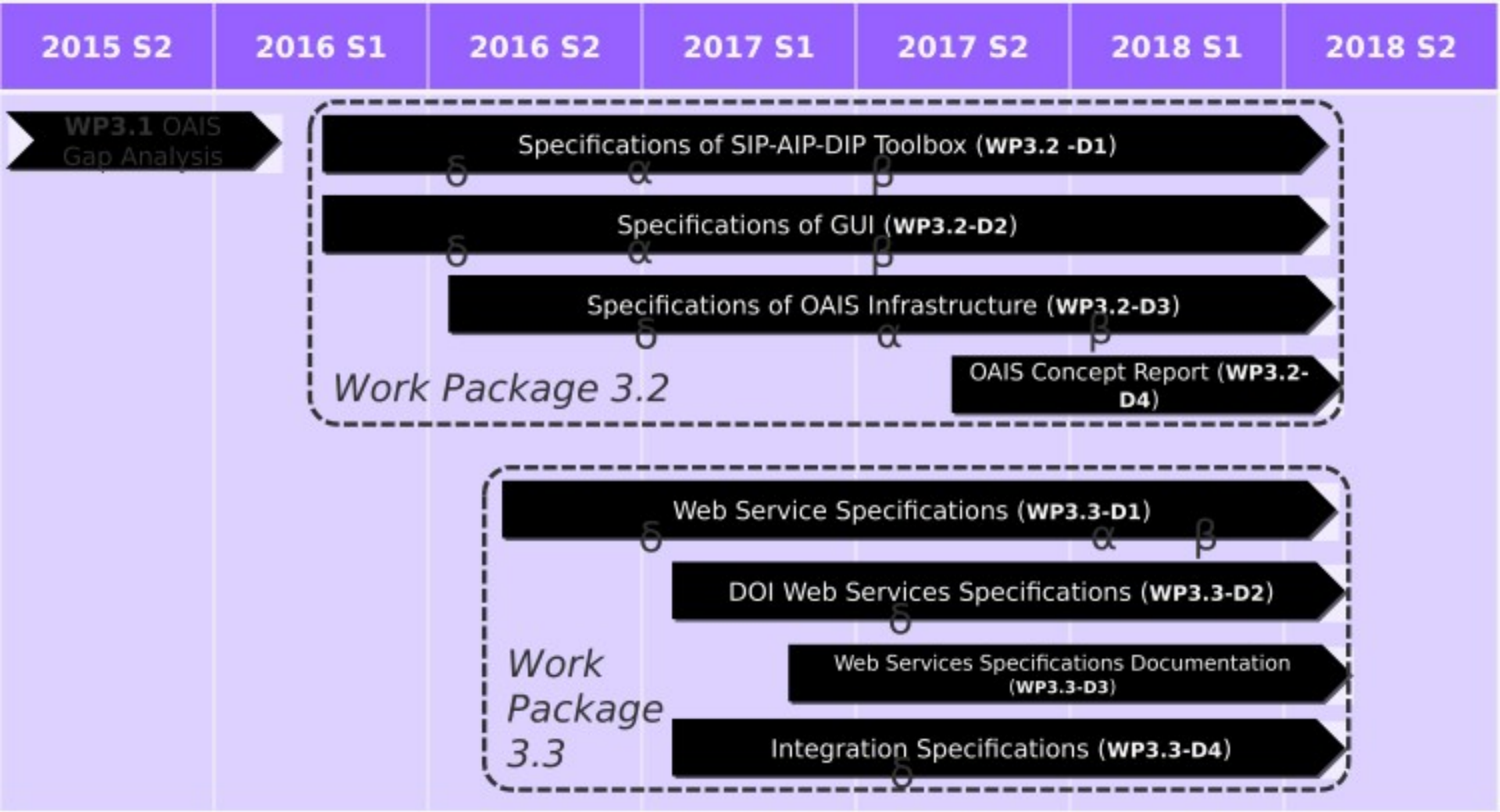
- **Format**
  - XML
  - RDF
  - JSON
- **What will be the minimal set of information to manage?**
  - MARC, METS, MODS, DC
  - PREMIS, MIX, EAD
  - LIDO, EDM
  - ...
- **OAIS Packages**
  - SIP / AIP / DIP
  - ZIP
  - BagIt
  - ...
- **Standard Compliance**
  - OAIS, ISO 23081-1, MoReq, ICA-Req
  - ...
- **How to exchange information?**
  - OAI-PMH, ResourceSync, SWORD, CMIS
  - ...
- **Dictionnary**
  - https://rd-alliance.github.io/metadata-directory/
  - http://metadataregistry.org/

Part 3
Project Management

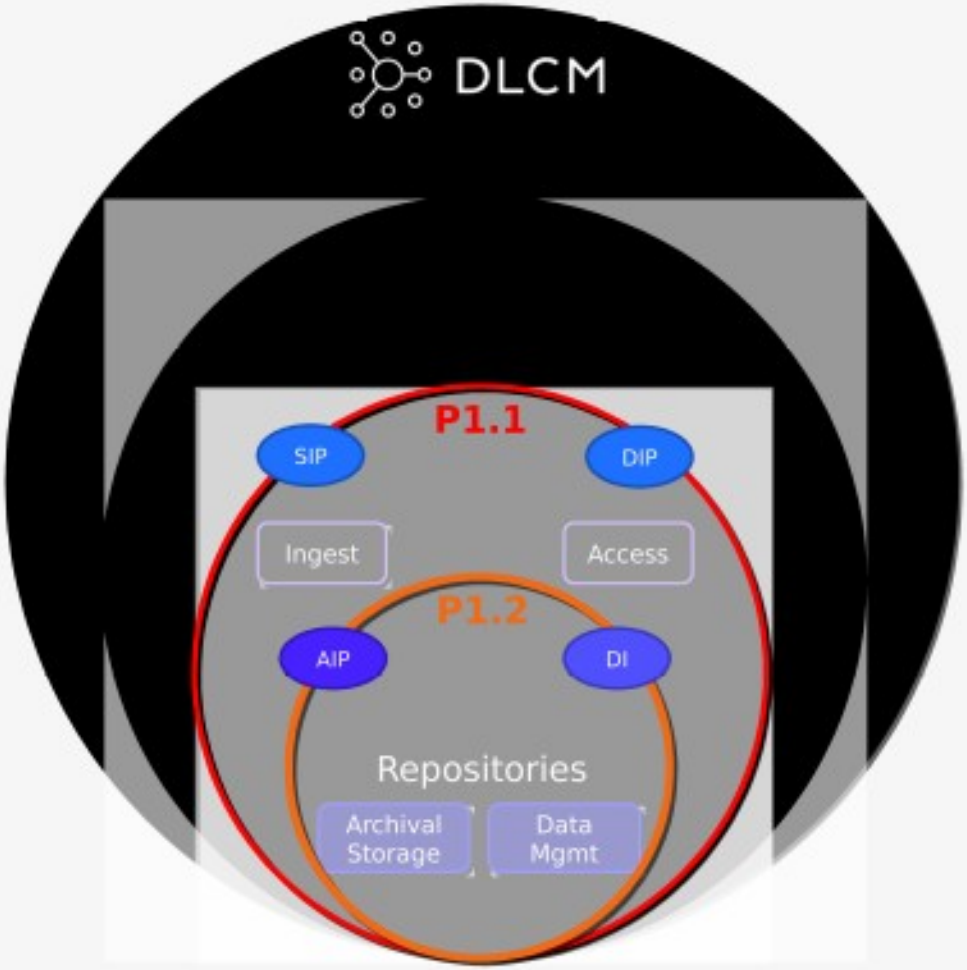
Work Packages

Work Package	Current Description	Proposal
WP3.2	**Toolbox to address requirements of OAIS compliance**	**DLCM Design**
WP3.2-D1	*Toolbox for building SIP - AIP - DIP from subsets of research data*	*Specifications of SIP-AIP-DIP Toolbox*
WP3.2-D2	*GUI adapted to different tools*	*Specifications of GUI*
WP3.2-D3	*Prototype of an OAIS-compliant infrastructure*	*Specifications & prototype of OAIS Infrastructure*
WP3.2-D4	*Report on concepts for a nationally distributed storage infrastructure compliant with the OAIS requirements*	*OAIS Concept Report*
WP3.3	**Define interfaces (APIs) to connect with the active data storage**	**DLCM Development**
WP3.3-D1	*Web services to connect the active storage system and OpenData repositories with the LTP- systems*	*Web Service Specifications & Implementation*
WP3.3-D2	*Web services to manage DOI*	*DOI Web Services Specifications & Implementation*
WP3.3-D3	*Specifications of the available Web services and standardization*	*Web Services Specifications Documentation*
WP3.3-D4	*Facilitating communication of metadata from LTP-solutions to ORD@CH, DataCite Metadata Store etc...*	*Integration Specifications & Implementation*

Initial Planning

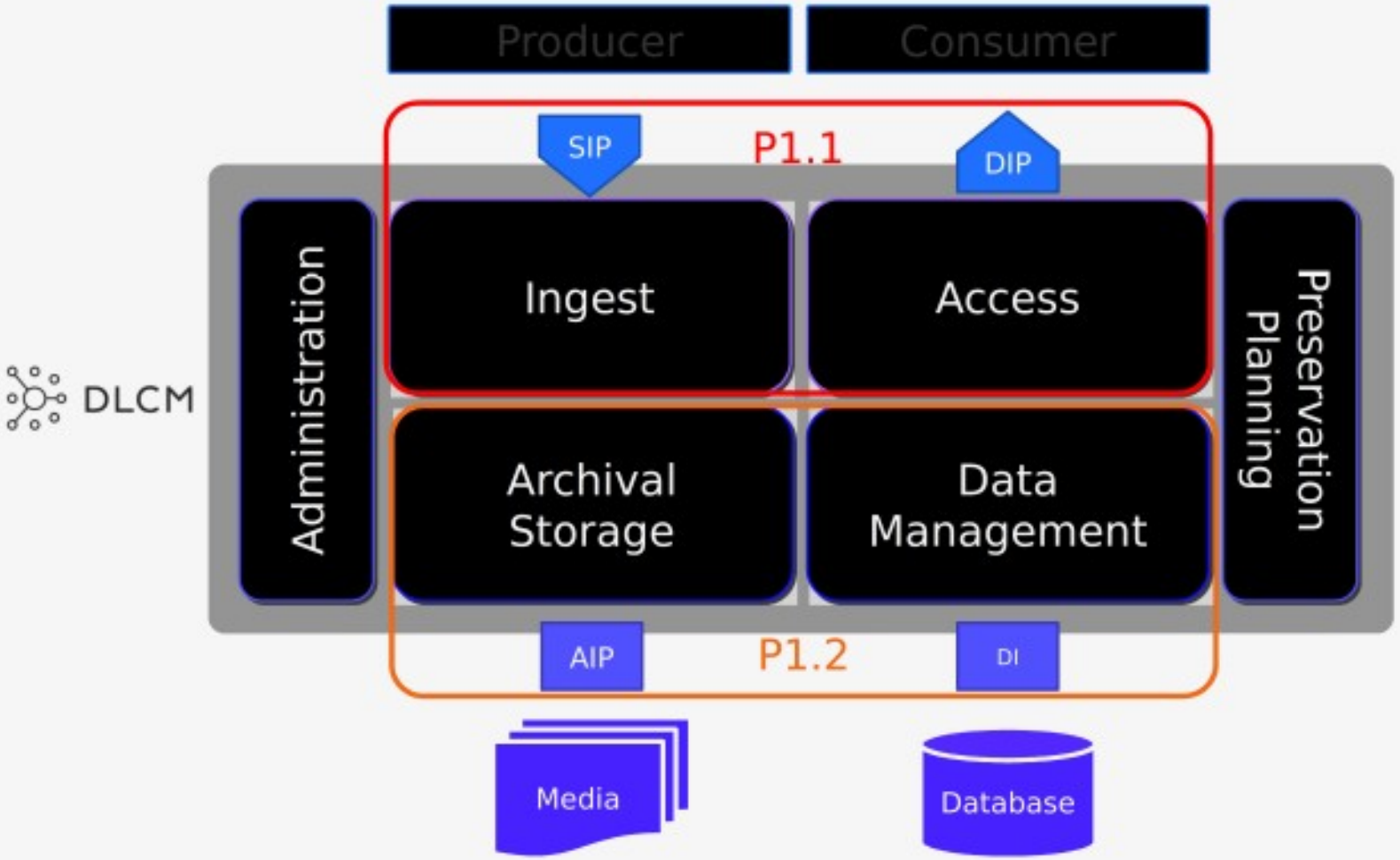


Iterative Approach



- ***Layers***
  - *Repositories*
    - *AIP*
    - *Archive Storage & Data Management*
  - *Services*
    - *SIP & DIP*
    - *Ingest & Access*
  - *Macro-Services*
    - *Tools for SIP & DIP*
  - *Applications*
    - *End-Users tools*
- ***Priorities***
  - *P1*
    - *Services*
    - *Repositories*
  - *P2*
    - *Macro-Services*
  - *P3*
    - *Applications*

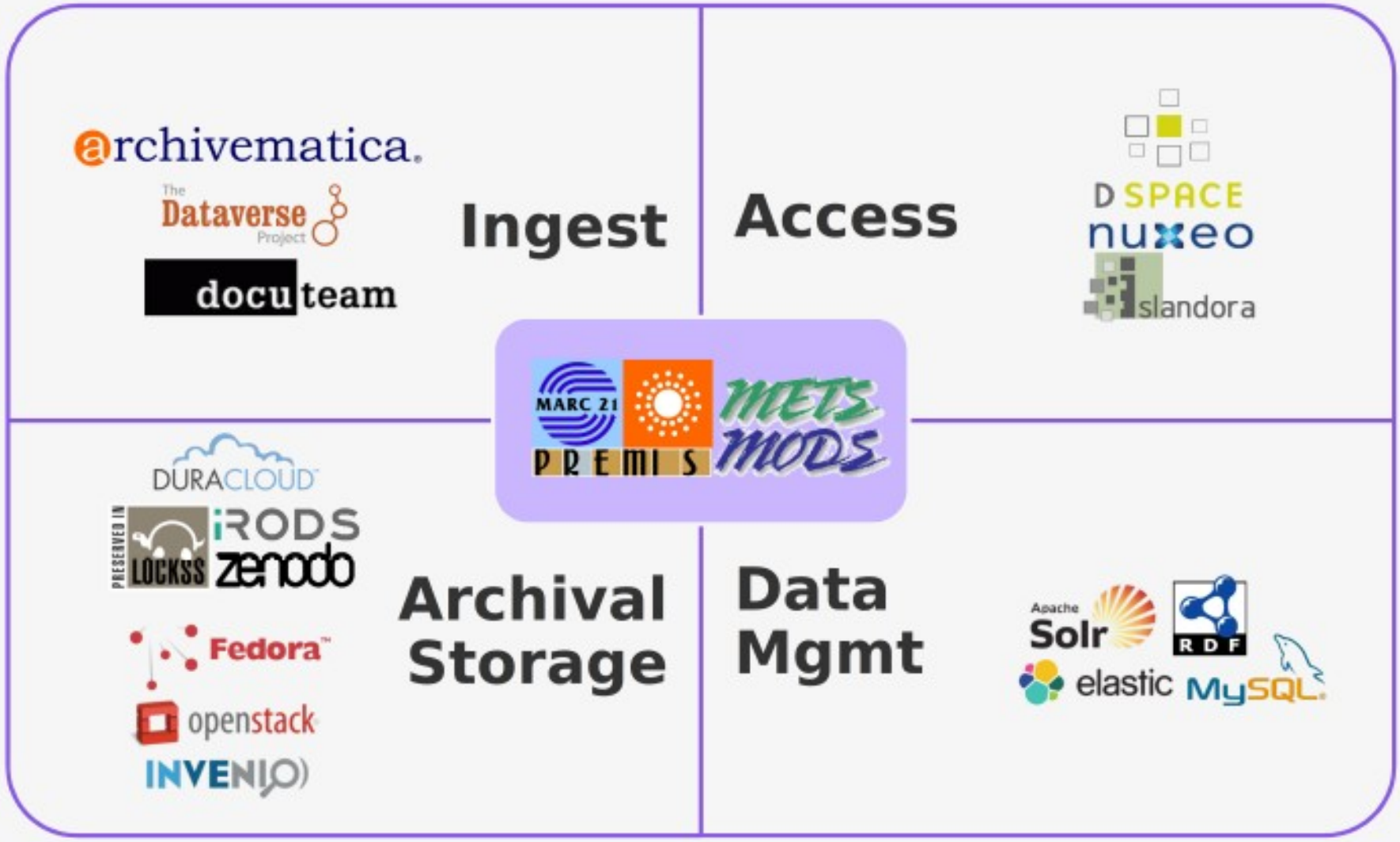
DLCM Solution (OAIS inspired)



P1 Task List

- **P1.1 tasks**
  - SIP Specifications
  - DIP Specifications
  - Who?
    - *Group “library-oriented”*
- **P1.2 Task**
  - AIP Specifications
  - DI Specification
  - Who?
    - *Group “IT-oriented”*
- **Package Specification Template**
  - Metadata
    - *Fields & Format*
  - Package
    - *Metadata & Data*
  - Protocol
    - *Standard or not*

DLCM Landscape



List not exhaustive

Next Steps

- Gather & summarize all workshop input
- Define specifications deadlines
- Write specifications
- Schedule a follow-up meeting
- Decide next workshop period

DLCM

Research Data Management Hub

Merci

Project Institutions

UNIVERSITÉ DE GENÈVE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

UNIL | Université de Lausanne

h e g Haute école de gestion Genève

UNI BASEL

Universität Zürich

ETH zürich

SWITCH

DLCM | swissuniversities