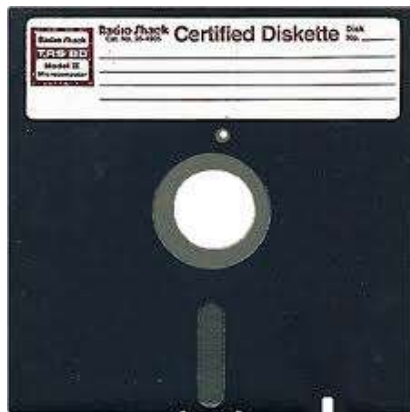


Archipel: Digitale Lange Termijn Preservatie

Inge Van Nieuwerburgh
(Universiteitsbibliotheek Gent)
Sam Coppens (Multimedia Lab)

Inhoud

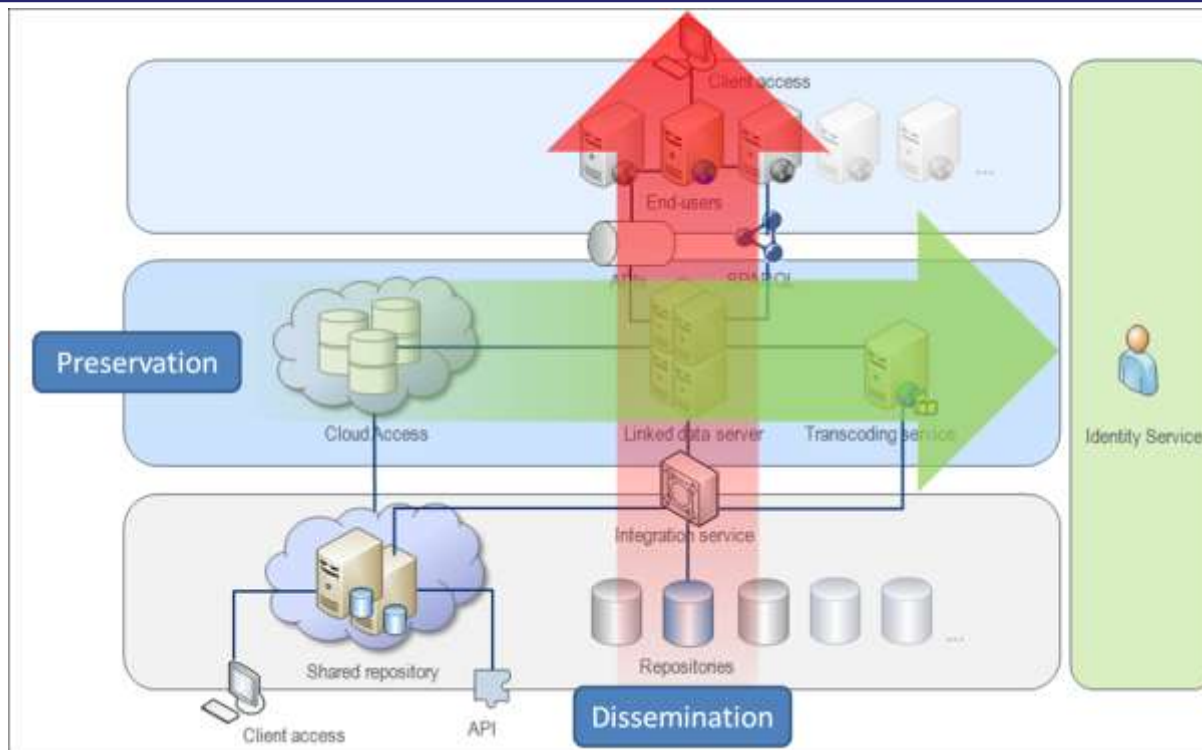
- Doelstellingen en Uitdagingen
- Data Model
- PREMIS OWL
- Architectuur
- Voordelen content providers
- Demonstratie



Doelstellingen & Uitdagingen

Doelstellingen:

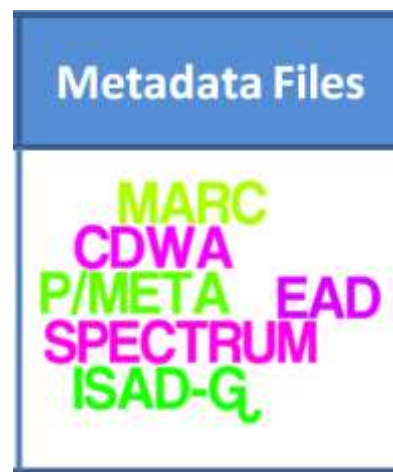
- OAIS Lange termijn preservatie van de data (metadata en multimedia assets)
- Publicatie van alle informatie als Linked Open Data



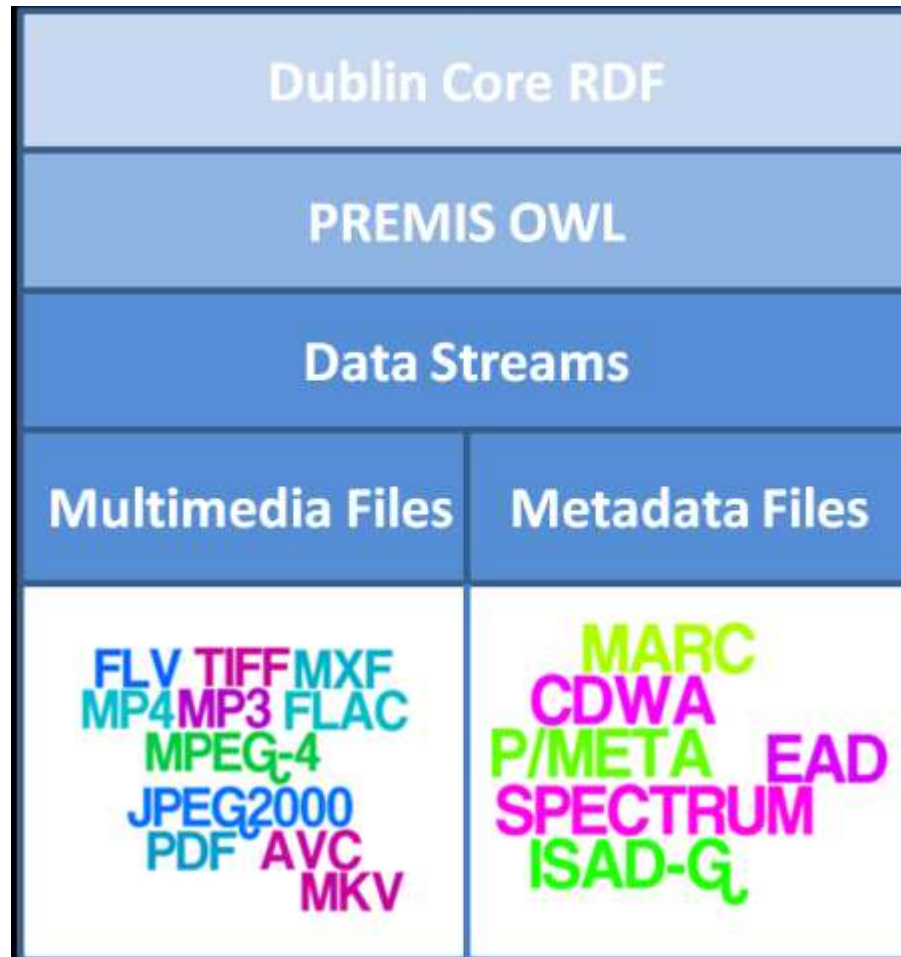
Doelstellingen & Uitdagingen

Uitdagingen:

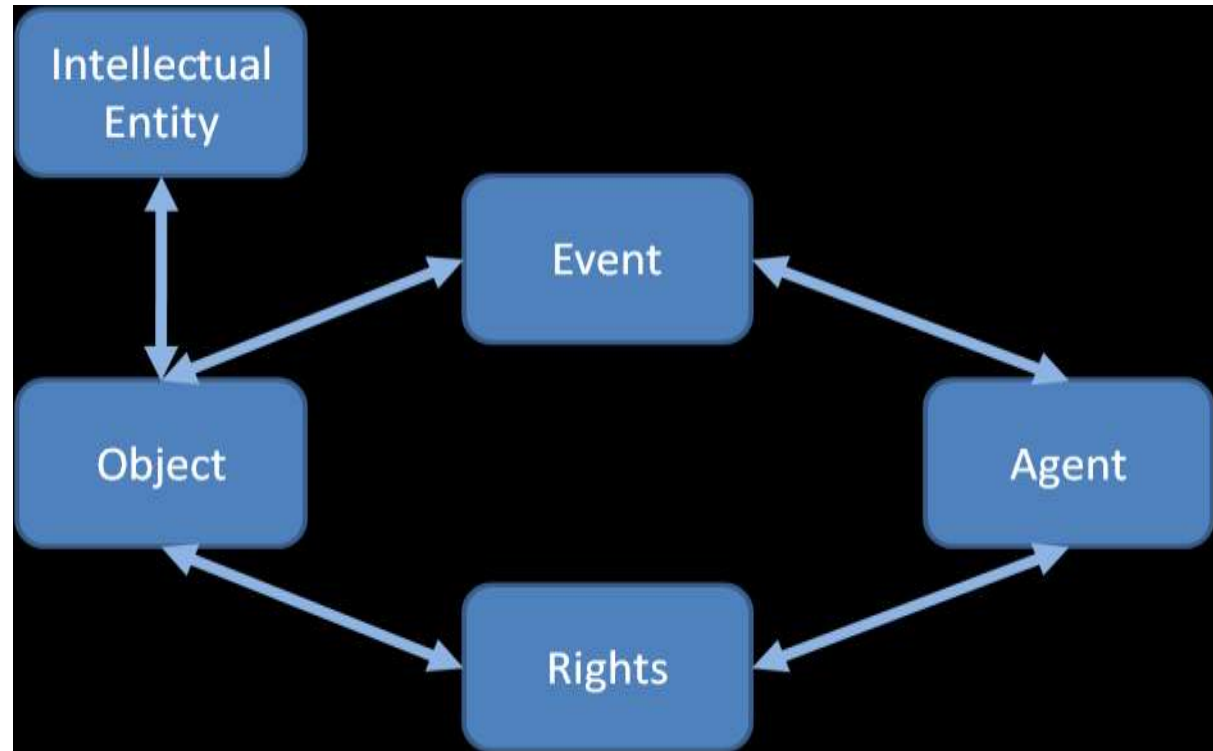
- Multi-layered metadata
 - Verschillende sector-specifieke metadata
 - Gemeenschappelijk metadata model
 - Preservatie metadata
 - Linked Open Data publicatie
- OAIS in een gedistribueerde, semantische omgeving



Data Model



PREMIS OWL



PREMIS OWL

- Semantische implementatie van de PREMIS 2.1 Data Dictionary.
- Kan gebruikt worden om preservatie informatie te publiceren als Linked Open Data.
- Gelinkt aan de preservatie vocabularies van het LOC.

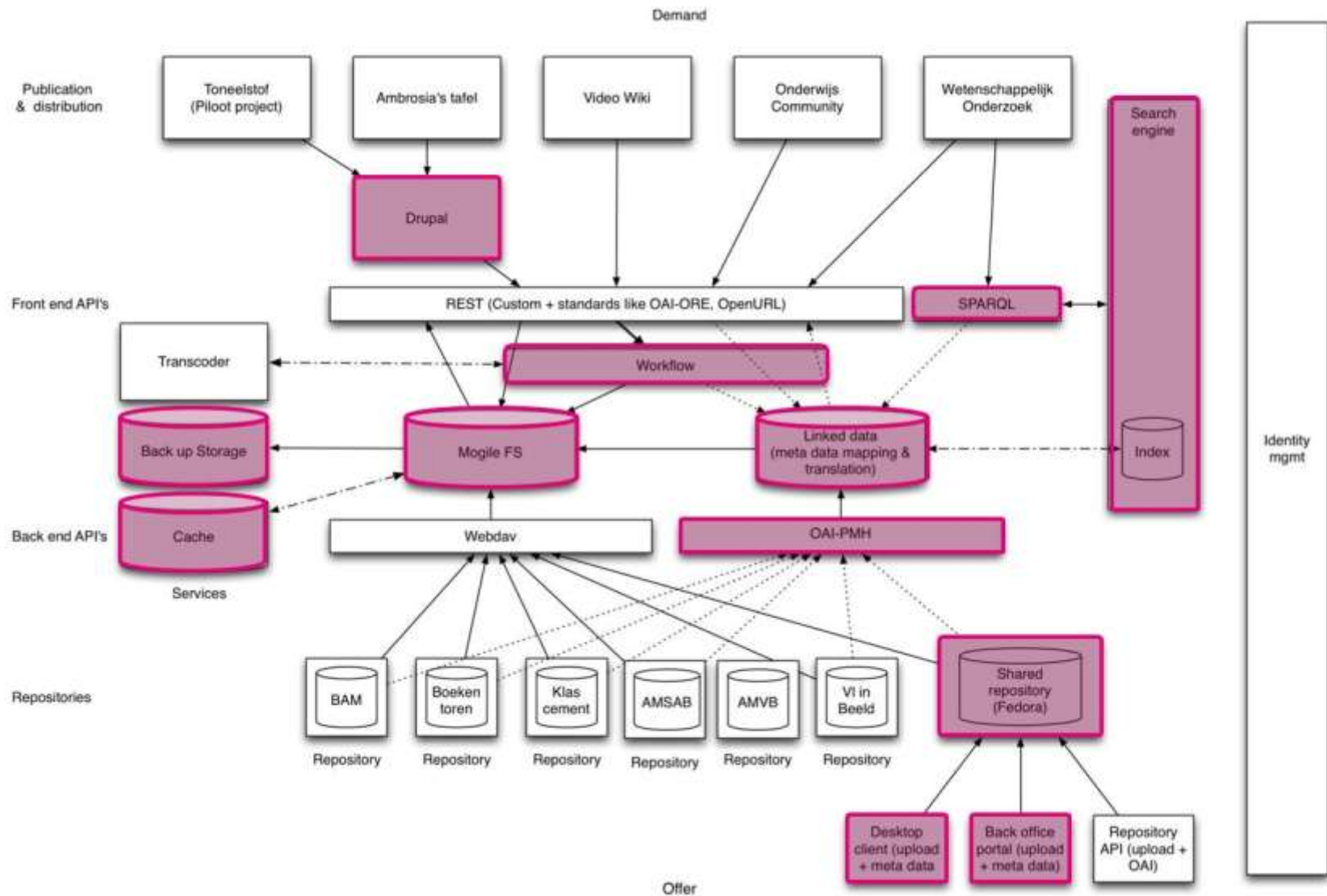
Namespace:

<http://multimedialab.elis.ugent.be/users/samcoppe/ontologies/Premis/premis.owl>

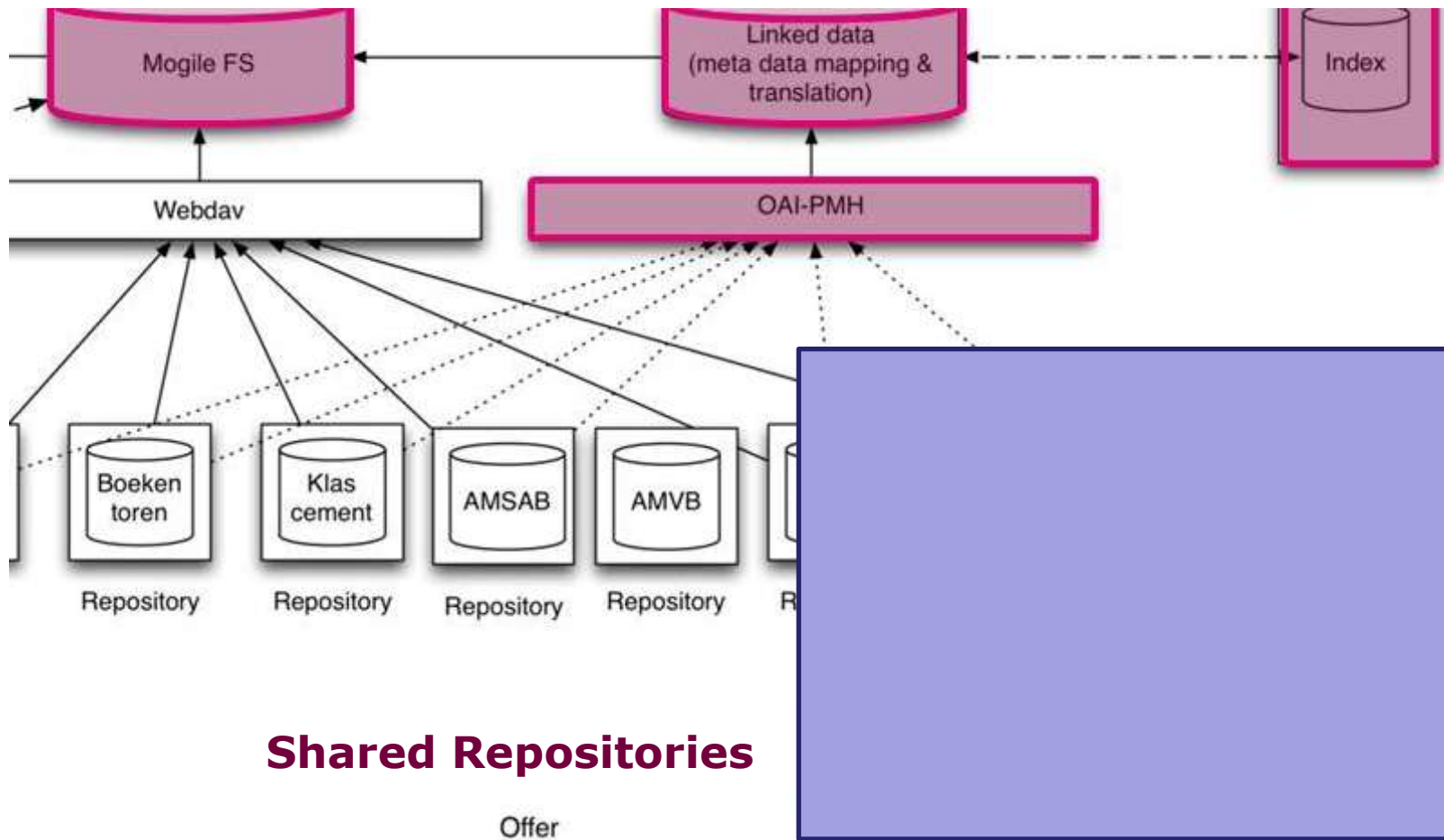
Ondergaat een standardisatie proces:

- Archipel – LOC - BNF
- Zal weldra aangekondigd worden door het Library of Congress en naar een LOC namespace verhuizen.

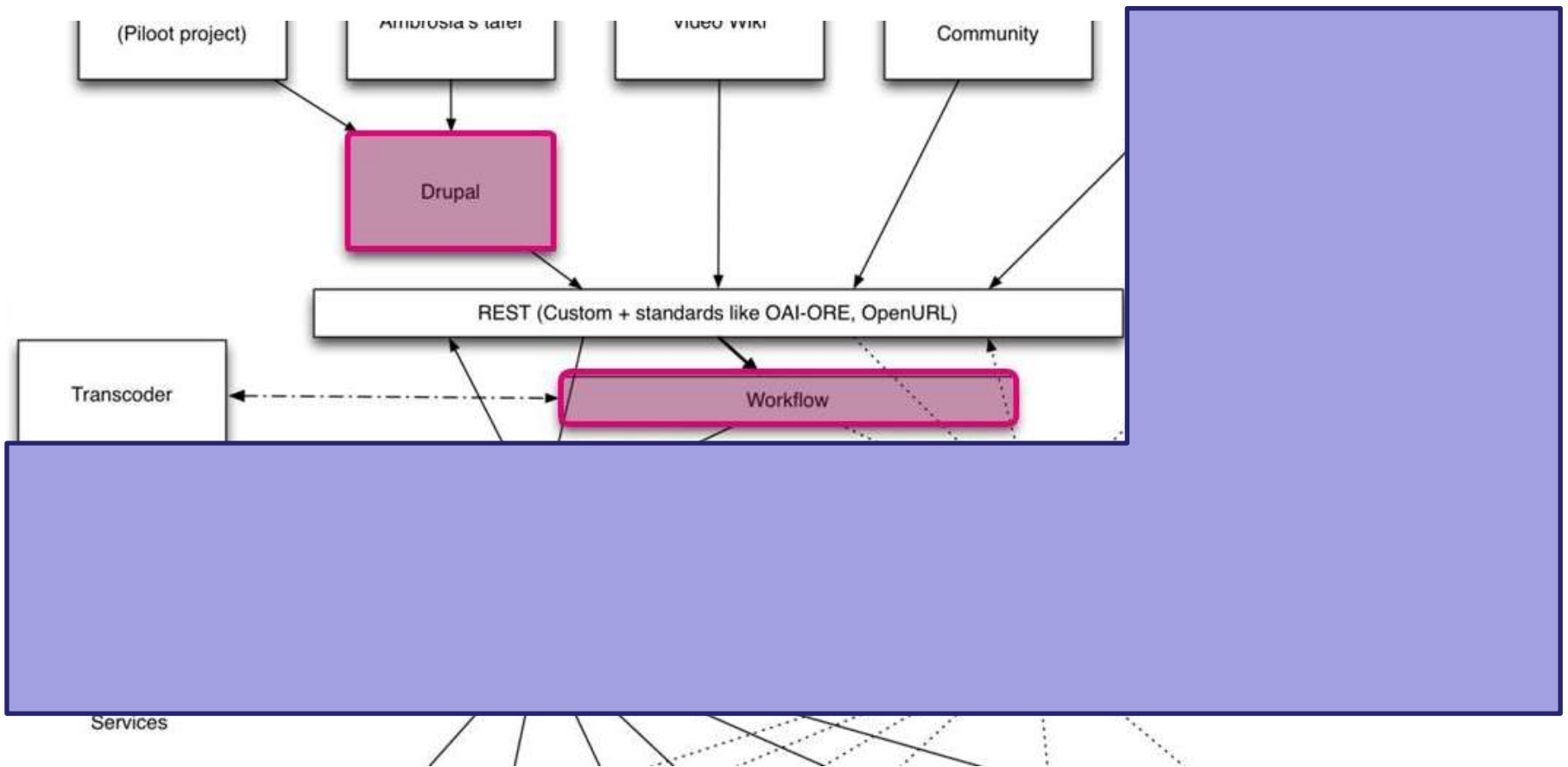
Architectuur



Architectuur



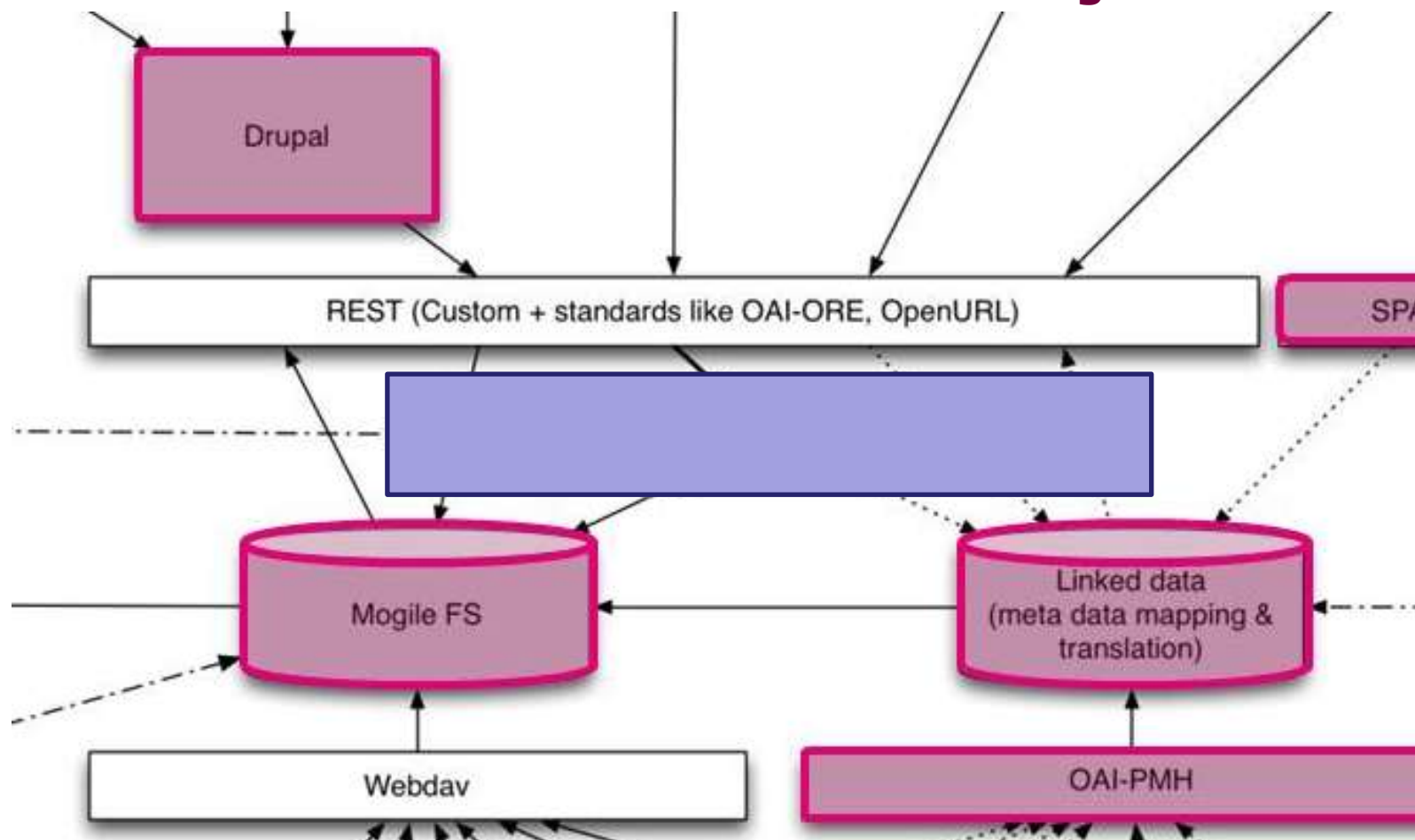
Architectuur



**LOD Server –
Preservatie Platform**

Architectuur

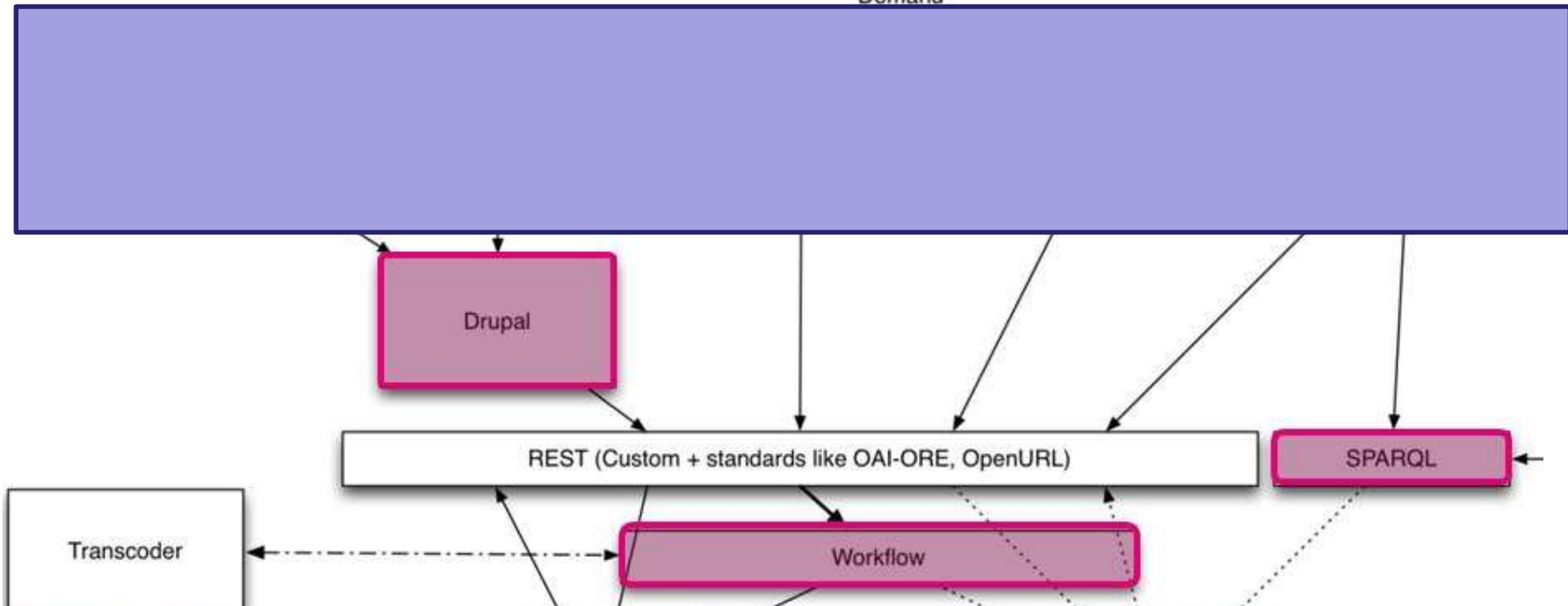
Integratie Server



Architectuur

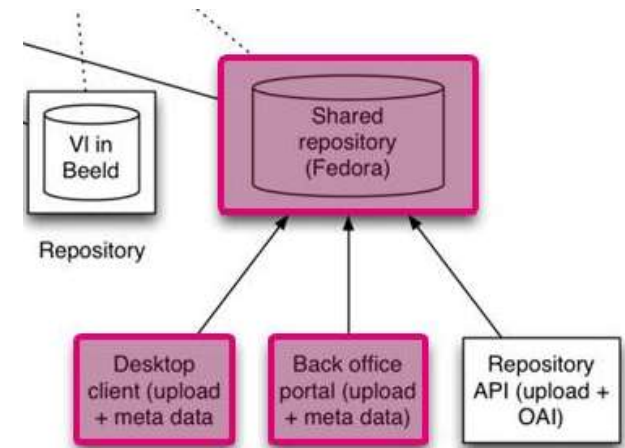
Disseminatie modules

Demand



Architectuur - Ingest

1) Shared Repository via OAI-PMH:



- Publiceert content online en maakt de content beschikbaar voor harvesting (OAI-PMH)

MediaMosa (<http://www.mediamosa.org/>)

Andere overwogen tools:

Omeka (<http://omeka.org>)

Fedora Islandora (<https://wiki.duraspace.org/display/ISLANDORA/Islandora>)

Architectuur - Ingest

2) BagIts via FTP:

- ZIP-archieven met metadata, data en extra informatie (checksums)
BagIt (<https://wiki.ucop.edu/display/Curation/BagIt>)

```
exampleBag/
```

```
| manifest-md5.txt
```

```
| (49afbd86a1ca9f34b677a3f09655eae9 data/27613-h/images/q172.png)
```

```
| (408ad21d50cef31da4df6d9ed81b01a7 data/27613-h/images/q172.txt)
```

```
| bagit.txt
```

```
| (BagIt-version: 0.96)
```

```
| (Tag-File-Character-Encoding: UTF-8 )
```

```
| \--- data/
```

```
| | 27613-h/images/q172.png
```

```
| | (... image bytes ...)
```

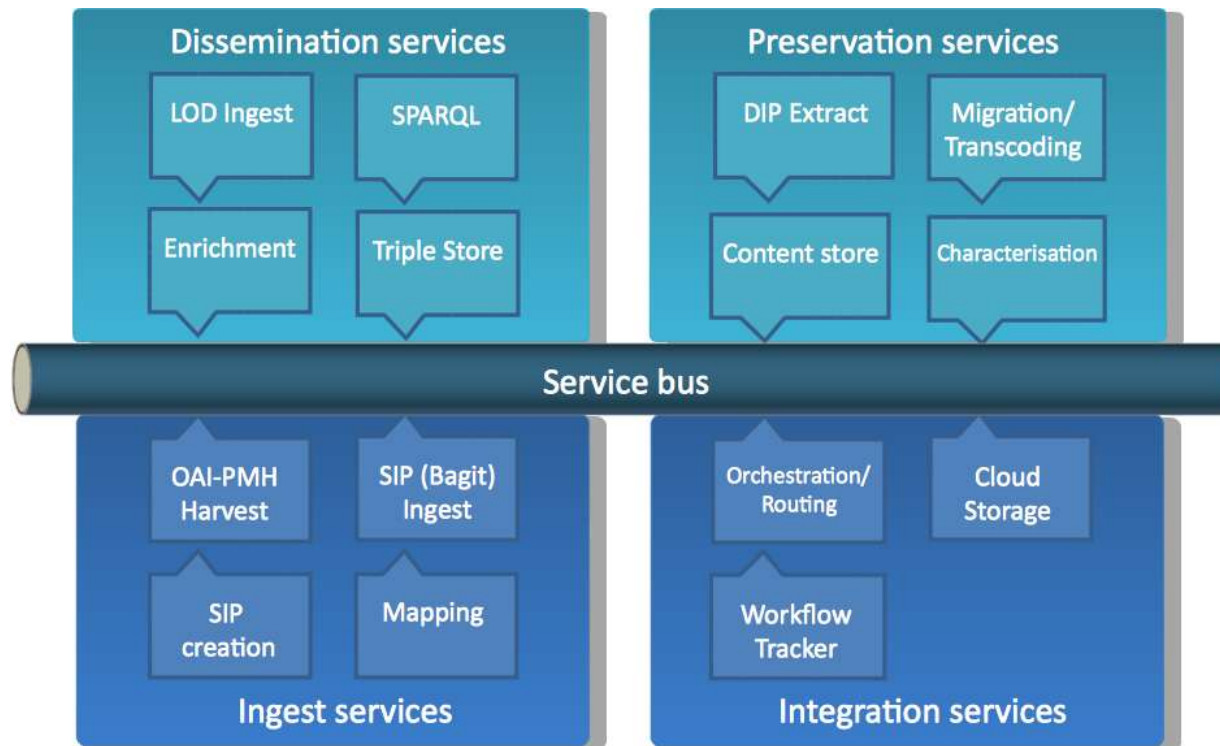
```
| | 27613-h/images/q172.txt
```

```
| | (... OCR text ...)
```

Architectuur - LODServer

LODServer & Preservatie Platform:

- Disseminatie van de content als Linked Open Data
- Preservatie van de content via semantisch webtechnologieën



Architectuur - LODServer

LODServer & Preservatie Platform:

- Ingest: OAI-PMH of via BagIts
- Characterisatie / Validatie: DROID + PRONOM + Preserv2
- Triple Store: Openlink Virtuoso Open Source
- Transcoding:
 - Video/sound: FFMpeg
 - Beeld: ImageMagick
 - Documenten: JODConverter (OpenOffice converter)
- Disseminatie:
 - SPARQL
 - Lucene Indexing → Full Text Search
 - Linked Open Data
 - BagIts

Architectuur – Integratie Server

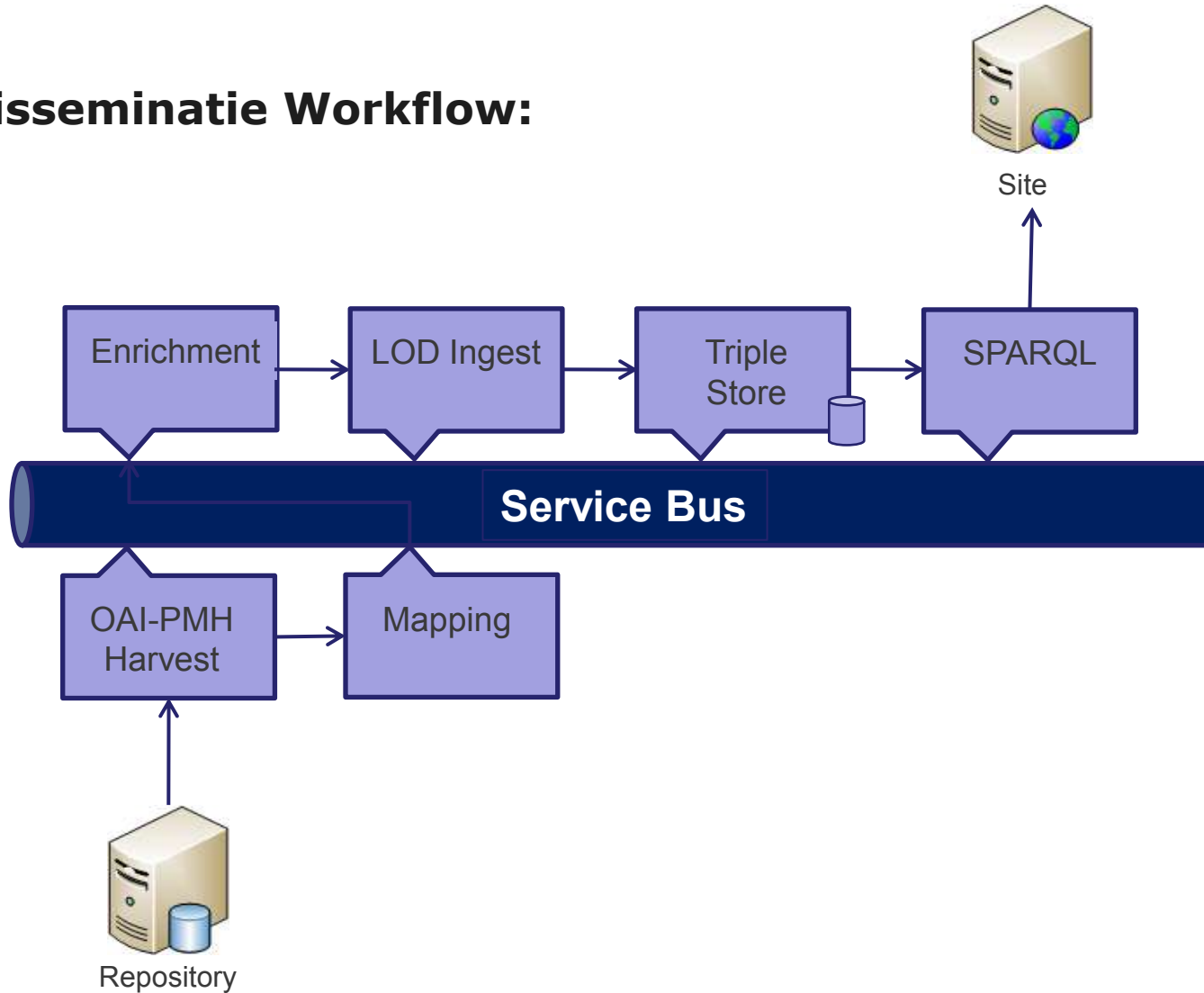
Integratie Server:

- Orchestreert alle preservatie- en disseminatieprocessen.

Porthus .Net Integration Server

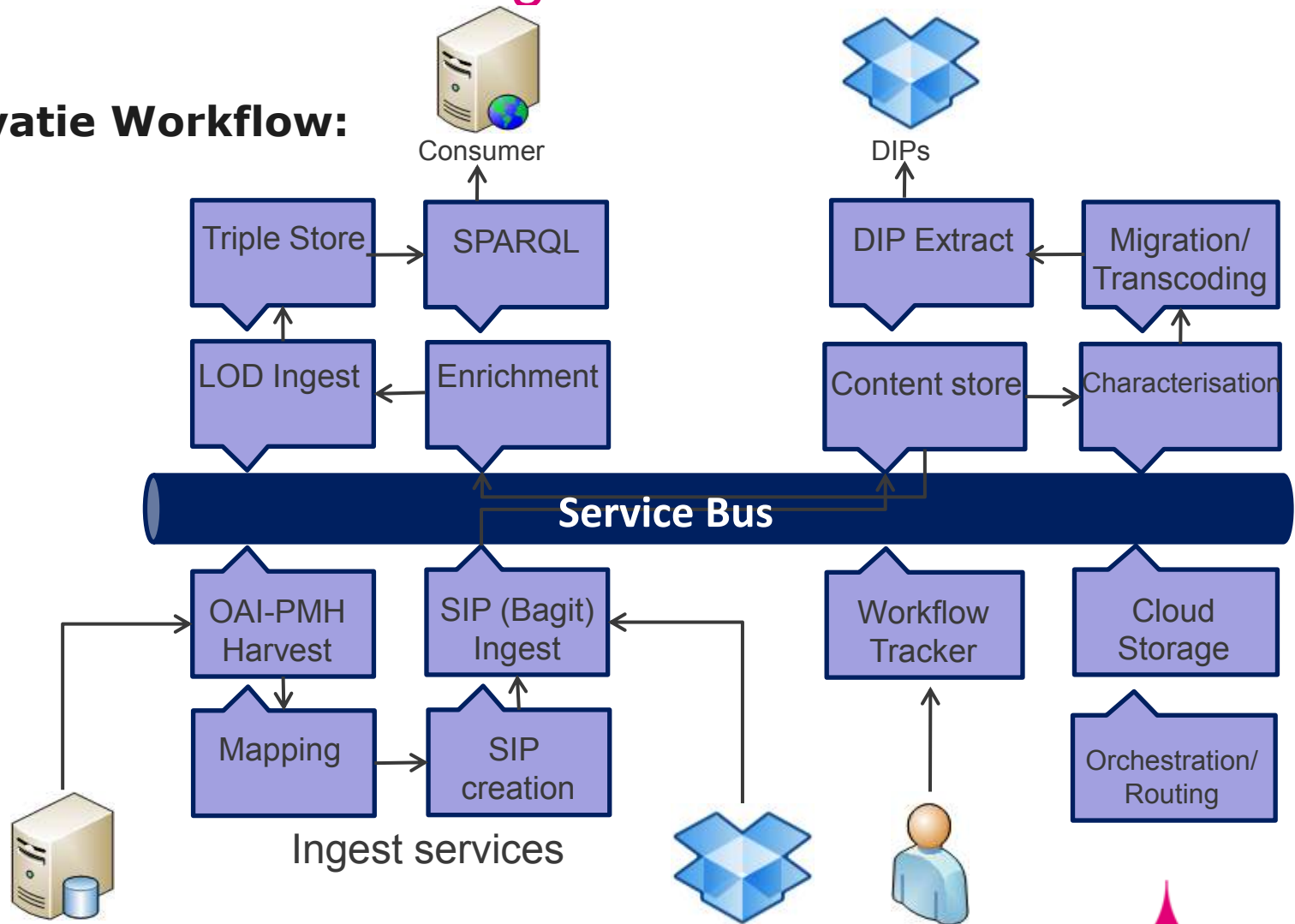
Architectuur – Integratie Server

Disseminatie Workflow:



Architectuur – Integratie Server

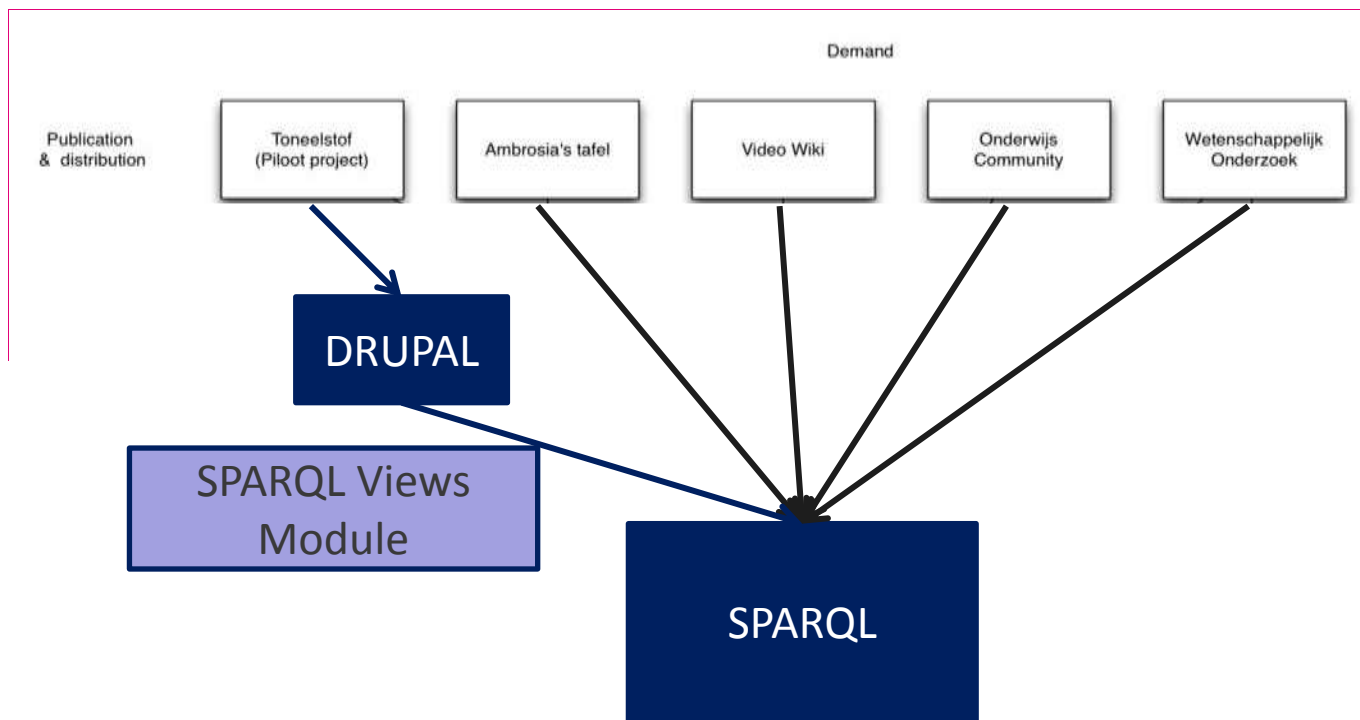
Preservatie Workflow:



Architectuur – Disseminatie Modules

Disseminatie Modules:

- Alle disseminatie modules maken gebruik van het SPARQL endpoint van de LOD Server



Voordelen Content Providers

- Lange termijn preservatie van de content.
- Linked Open Data publicatie van de content.
- Reconciled metadata:
 - Semantische metadata
 - Gelinkte metadata
 - Resource based:
 - SKOS vocabularies
 - Personen- en Organisatie-register
 - Verrijkingen

Demo

Medewerkers Architectuur

- Descartes (Porthus)
- Inuits
- Krimson
- VTi
- Boekentoren
- Multimedia Lab

Demonstrators

- Toneelstof: theater
- Ambrosia's Tafel: mediawijsheid bij jongeren
- Constant vzw: Active Archives Video Wiki
- Klascement: portaal voor leerkrachten
- Boekentoren – UGent: portaal voor Onderzoek en hoger onderwijs

Wensen

- Zoveel digitale audiovisuele data, zo moeilijk toegankelijk
- 1 portaal voor onderzoek en onderwijs UGent: zoveel mogelijk externe bronnen
- Data data data data

Voorwaarden

- Simpele maar krachtige zoekmogelijkheden
- Snelheid
- Eenvoudige downloadmogelijkheden
- Data mining (maar: CR)

Technische voorwaarden

- Gebaseerd op standaarden (zie ook onderzoek in BOM_v1)
- Mogelijk integratie met andere systemen zoals ELO
- Gebaseerd op bestaande technologie

Bestaande technologie

- Solr: zoekmachine
- Lucene index
- Representatie:
 - Hangt af van bron-interface
 - Zoomer beelden: IIPimage
 - Streamer video en audio: CRTMP server

File types

- Lokaal beschikbare files:
 - TIFF = input
 - Pyramid TIF en JPEG2000 = zoomer (afgeleid van tiff)
 - mp4 (wordt bijgewerkt), flv, mp3 (dit heeft streamer nodig)
- Afgeleiden:
 - JPEG, GIF, PNG

Search&find

- Krachtige search:
 - Google-like, los van velden
 - Fuzzy (ondermeer spellingsfouten worden opgevangen)
 - Boolean search
- Zoeken via een kaart (= feature)
- Openurl gebaseerd: bv. verschillende formaten eenvoudig tonen
- Toegang tot de xml-beschrijving of Json-beschrijving, indien beschikbaar

Resultaten

- Sterke zoomer
- Bladerinterface: kan complexe objecten aan
- Persistente URLs: zowel per complex object als per onderdeel (bv. boek – blz)
- Selectie en mandje (vooral organisatorische keuze wat kan)
- Toegangscontrole mogelijk

- 16 november: eindevenement archipel in deSingel in Antwerpen
- Registreren via <http://archipelproject.be>

Medewerkers

- Technisch: Nicolas Franck en Patrick Hochtstenbach
- Inhoudelijk: Mui-ling Verbist en Inge Van Nieuwerburgh