Mtas in TLA-FLAT

Menzo Windhouwer

menzo.windhouwer@meertens.knaw.nl

Overview

1. TLA-FLAT: a CLARIN compatible repository solution based on Fedora Commons and Islandora

2. Integration of Mtas in TLA-FLAT
   1. Current status
   2. Demo
   3. Future work

TLA-FLAT base line

- Meet the, technical, CLARIN B centre requirements
- Meet the, technical, Data Seal of Approval (DSA) requirements
- Meet organization specific requirements
  - Meertens Institute
  - Max Planck Institute for Psycholinguistics

    The Language Archive

- Meet, at least the CLARIN B centre and DSA, requirements, as much as possible, with the Fedora Commons backend
  - frontend (technology) come and go quickly
- How far can we get using available components, configuration and a limited level of tailor made software?
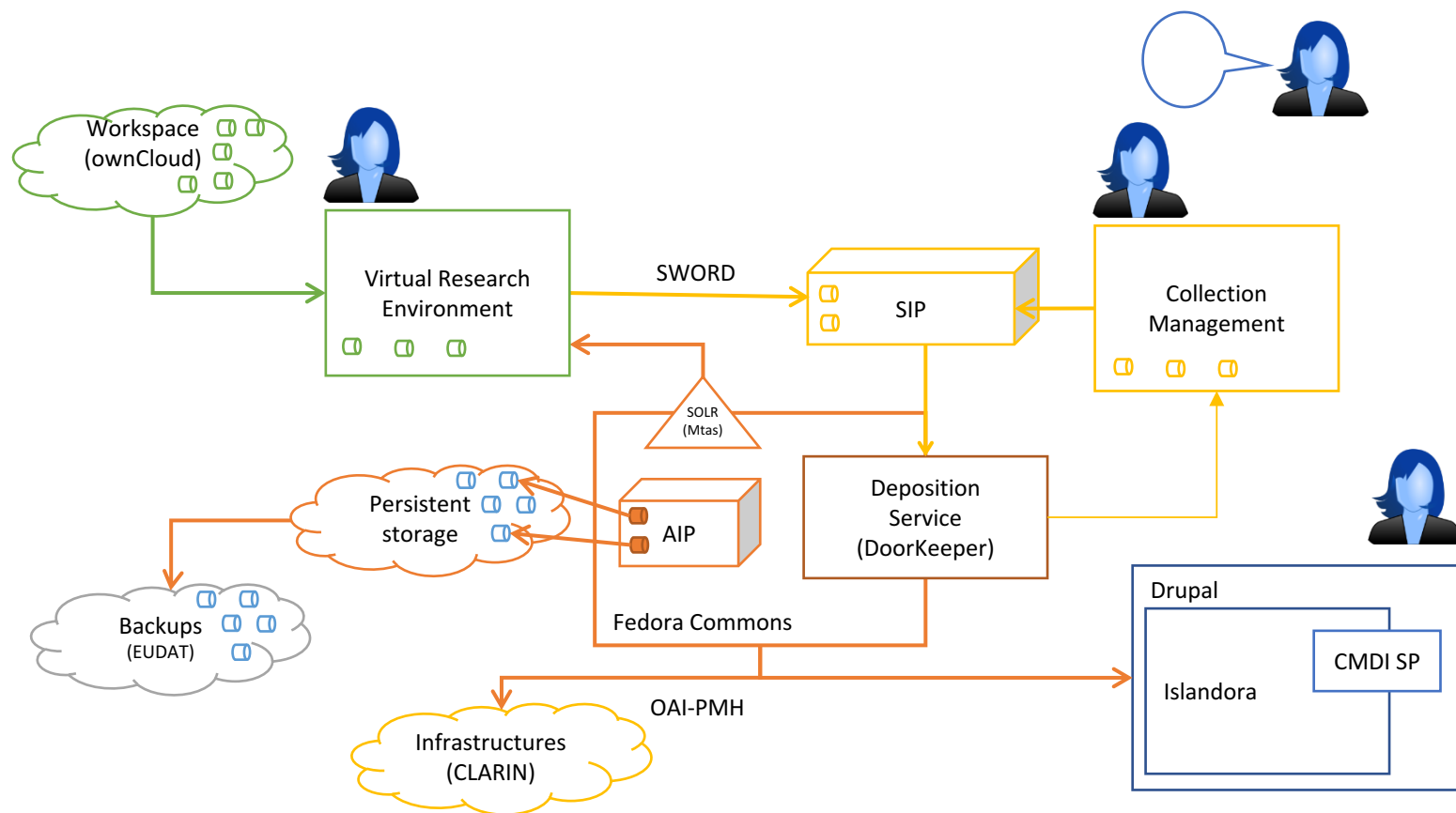 - Mainly to add support for CMDI

TLA-FLAT modules

- Core
  - Fedora Commons and Islandora setup
  - CMDI Solution Pack
  - CMD to FOXML conversion (batch import)
  - Proai setup

- Indexing (SOLR)
  - VLO/gsearch-based solution for CMDI
  - Meertens' CMDI indexer

- SWORD 2.0
  - Reuses a deposit via SWORD approach and implementation by DANS

- DoorKeeper
  - Lots of actions to validate and properly ingest a bundle of CMD record + resources

- Shibboleth

- Deposition UI

- IMDI conversion

Shibboleth setup is very server specific, so there is a module that illustrates the Drupal setup and can be combined with a test IdP.

TLA-(Mtas)-FLAT's place at the Meertens Institute

CMDI indexing for facetted search

1. Meertens CMD indexer
   - Analyzes the profiles in your CMD records
   - Creates facets for all semantic paths it finds
     - Facet names based on concept links (plus context)
   - At runtime switch between facets for querying and rendering
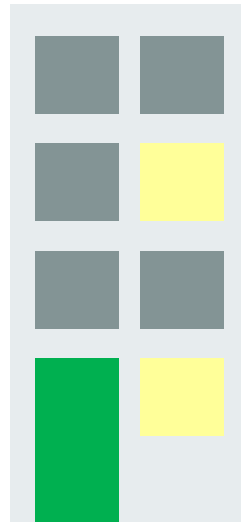   - Also indexes the FOXML to meet Islandora facet needs

Integration of Mtas

➢First example of content-based search in TLA-FLAT

1. Upgraded the SOLR used by TLA-FLAT from 4.x to 6.x
2. Replaced the SOLR Core template for the Meertens CMD indexer
   1. 6.x-based
   2. Includes the Mtas JARs
   3. solrschema.xml adapted to support Mtas
   4. schema.xml adapted to support Mtas
   5. Includes basic config for FoLiA resources
3. Generate the Mtas facet if a FoLiA resource is encountered
   1. The file path to the resource

Let's visit the Meerten's TLA-Mtas-FLAT!

TODO's

- Now only batch import supports indexing, also a single (DoorKeeper) deposit should trigger indexing (TODO, but not specific to Mtas)
- Meertens indexer now works on the CMDI resource, which is a compound, but the Mtas-indexed resource is actually a member of this compound
  - Becomes especially problematic if the compound contains multiple Mtas-indexed resources
  - But allows mixing metadata and content queries
    - Joins to the rescue?
- Support different Mtas configurations for different collections
  - Beware: keep the index uniform!
- CQL query box/builder + Mtas specific result views
- Viewer for FoLiA/Mtas-indexed resources