

openMIN7ED
@openminted_eu

Data Extraction and Annotation from Viticulture Publications (PDF files)

Agroknow



The background of the slide is a complex digital visualization. It features a central perspective view of a tunnel-like structure formed by numerous glowing, multi-colored lines (blue, green, yellow, and orange) that radiate outwards. Interspersed among these lines are vertical columns of binary code (0s and 1s) and various geometric shapes, creating a sense of depth and data flow. The overall aesthetic is high-tech and data-oriented.

Data Extraction and Annotation from Viticulture Publications (PDF files)

The presentation was done in the context of the [1st Agrohackathon event](#) that was organized in Montpellier between 29/6/2016 and 1/7/2016. The scope of the presentation is to demonstrate how Information can be extracted from publications on viticulture research

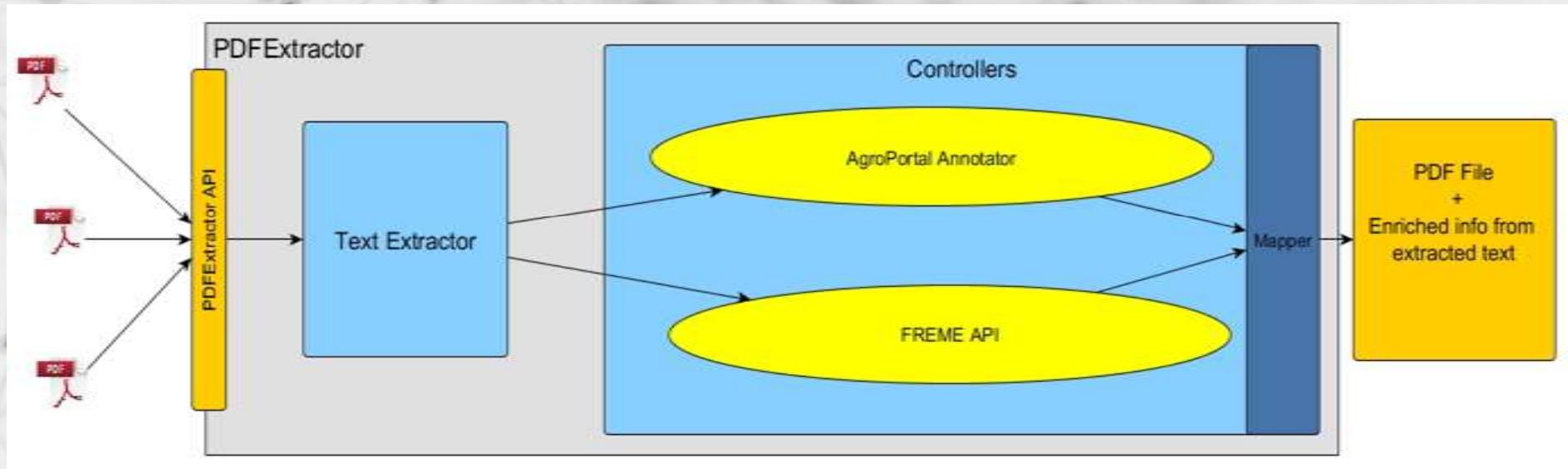
Main Goals

- Extract textual information contained in pdf files of (agricultural) bibliographic resources.
- Use extracted information against known endpoints for the semantic annotation.
- Design a modular system easily configurable and extendable.
- Have an intuitive way for end users to semantically annotate their pdf files.

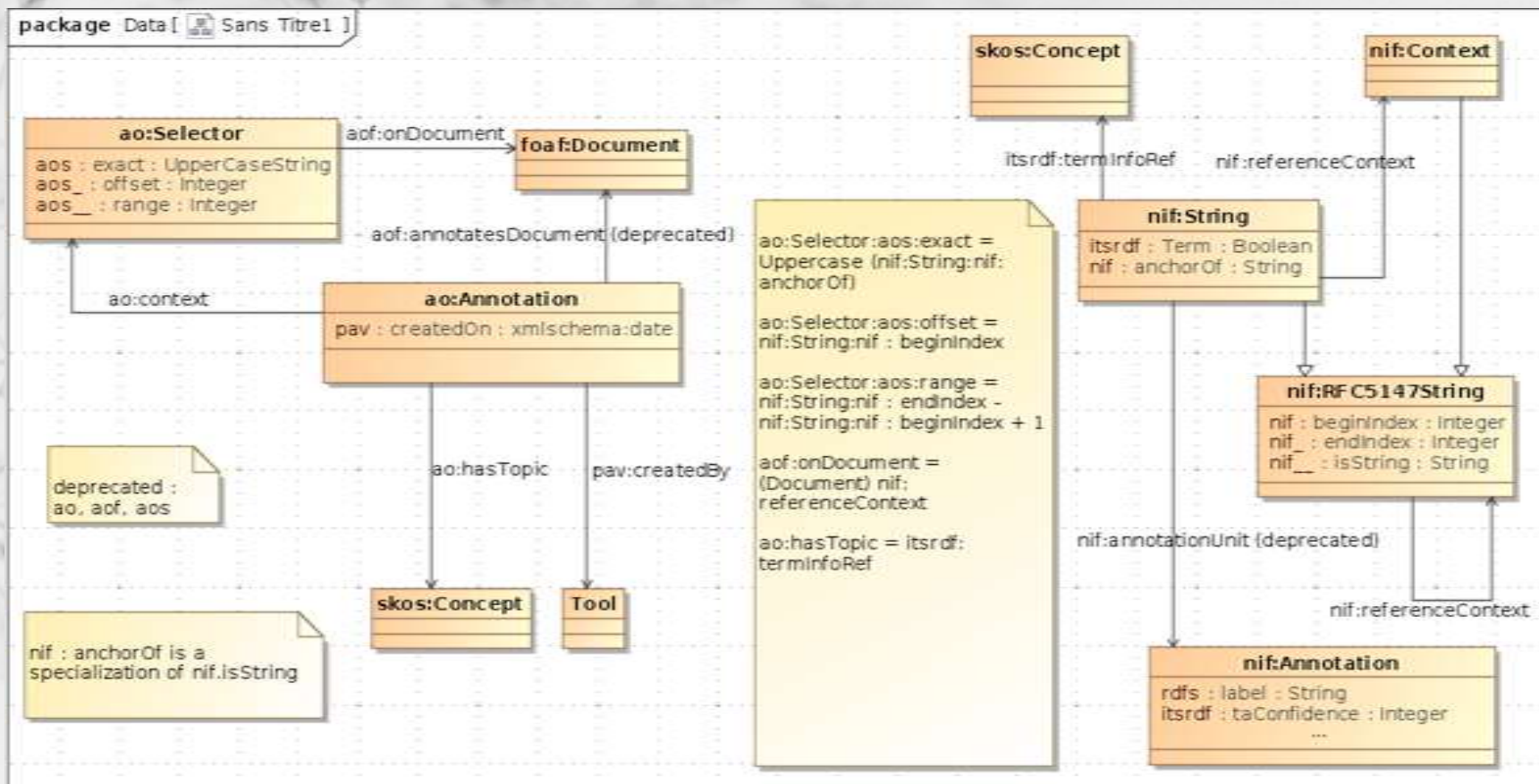
The process – Key Features

- We divided the system into 4 separate components:
 - ✓ API Endpoint, where all the calls are made, using as input the url of a pdf file.
 - ✓ TextExtractor, where the actual extraction of textual information takes place.
 - ✓ Controllers, where the extracted information is passed and then external endpoints are called.
 - ✓ Mapper, where the returned annotated result is presented back in a unifying manner.

PDFExtractor Workflow

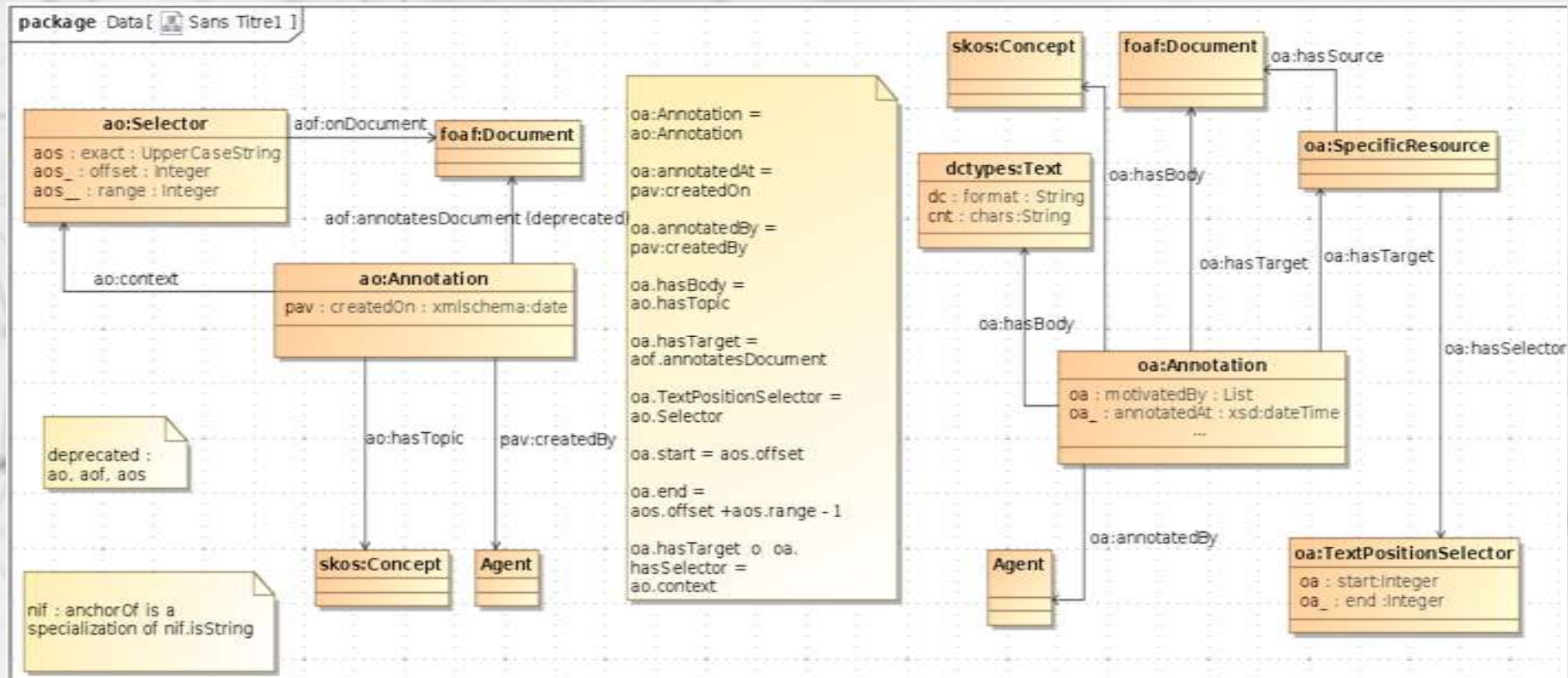


Mappings – AgroPortal with FREME

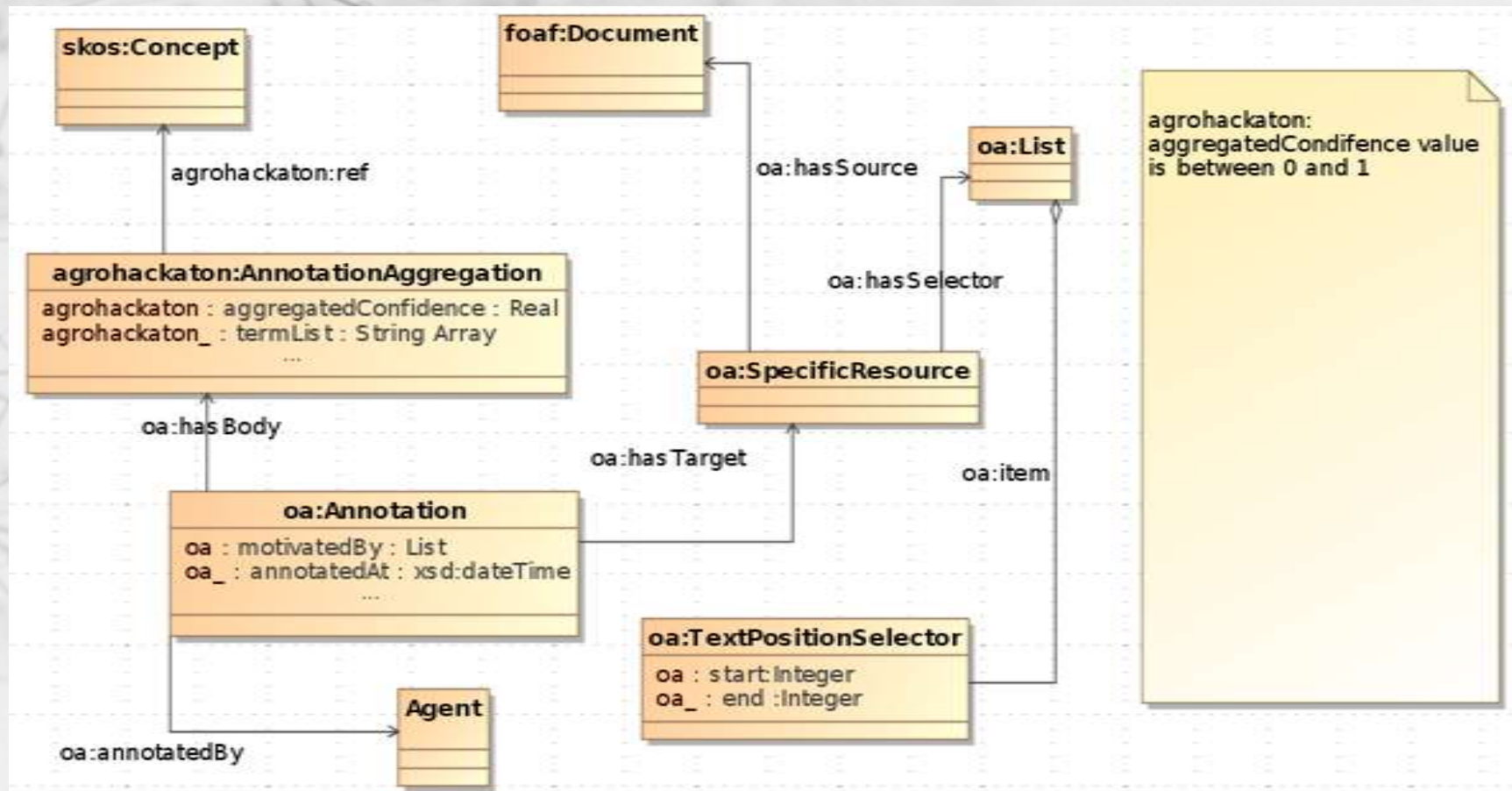


OPENMINTED - The Open Mining Infrastructure for Text and Data

Mappings – AgroPortal with OA



Proposed Evolution of OA based on FREME



Future Work

- Integrate more controllers into the workflow.
- Design a cleansing component clearing out redundant information.
- Provide a richer API endpoint for the system.
- Benchmark various endpoints called by the controllers.
- Build a front-end web app on top to help end users.

openMIN7ED



twitter.com/openminted_eu



facebook.com/openminted



bit.do/openmintedlinkedin



vimeo.com/openminted



bit.do/openmintedplus

Contact us

www.openminted.eu