INTEGRATION OF ALVIS WITHIN THE OPENMINTED PLATFORM AS GALAXY UTILITIES

Jean-Baptiste Bohuon*, Mouhamadou Ba*, Estelle Chaix, Robert Bossy, Claire Nédellec
MaLAGE, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France
firstname.lastname@inra.fr, *: corresponding authors

WORKFLOW ENGINES OF ALVIS AND GALAXY

- OpenMinTeD has chosen Galaxy as workflow engine
- The workflow engines of Alvis and Galaxy differ, each Alvis module can be wrapped as a Galaxy tool

<table>
<thead>
<tr>
<th>Feature</th>
<th>Alvis</th>
<th>Galaxy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execution of workflows</td>
<td>Sequential</td>
<td>Parallel</td>
</tr>
<tr>
<td>Workflow description</td>
<td>XML parameter-oriented</td>
<td>JSON with tool state</td>
</tr>
<tr>
<td>Data interoperability</td>
<td>Shared data structure</td>
<td>Native formats, shims</td>
</tr>
<tr>
<td>Data persistency</td>
<td>RAM, dedicated modules</td>
<td>Filesystem</td>
</tr>
<tr>
<td>Interfaces</td>
<td>CLI, REST API</td>
<td>Java API, GUI, REST API</td>
</tr>
</tbody>
</table>

ALVIS MODULES IN GALAXY

- We created a generic Alvis docker image
- Data interoperability within OpenMinTeD: RDF for documents and XMI for text annotations
- An RDF/XMI export module is under development
- Most Alvis workflows use a docker container giving standalone Galaxy tools
- Modules will be progressively wrapped as independent tools

COMMUNITY-DRIVEN WORKFLOWS

- Existing state of the art TDM workflows will be released on the OpenMinTeD platform as Galaxy workflows
- These workflows take domain-specific resources as input (corpora and semantic resources)
- Workflows in Agricultural Science and Biodiversity Science include:
  - Plant development
  - Wheat data
  - Microbial diversity
  - Client application
  - FlagId++
  - GnpIS, WheatIS
  - Florilège
  - Partner
  - IJPS2 (INRA)
  - UriGI (INRA)
  - E-infra partners
  - Main TDM resource
  - TAIR
  - Wipo
  - OntobioType
- The bioinformatics applications associated with workflows (FlagId+++, GnpIS, Florilège) will integrate TDM results and bioinformatics data
- Bioinformatics applications will execute the preconfigured workflows
- Periodical execution: corpora and resource update and user feedback will trigger execution
- CORE, OpenAire, AgroPortal will provide access to data

REFERENCES

- Ba et al. (2016). Interoperability of corpus processing workflow engines: the case of AlvisNLPL in OpenMinTeD. Presented at Meeting of working Group Medicago sativa, Portoroz, SVN.
- Alvis git repository http://github.com/Bibliome/alvisnlp
- OpenMinTeD git repository http://github.com/openminted
- Alvis generic docker image http://github.com/openminted/alvis-docker

www.openminted.eu