CRIS & VIVO

Promoting and facilitating interoperability with VIVO
Improving data load in VIVO
Uploading data in VIVO

Introduction

Goal:
So, the goal was, can we do the same process that VIVO provides, but automatically and in one step? The answer is Yes and is more simple that it seems!

Problem:

VIVO manages manually the uploading of information. This is a big effort.

SIGMA, as a Spanish universities consortium, needs to provide an easy way to upload data to VIVO for more than one university.
Uploading data in VIVO

Description of the current way to upload information in VIVO

As we said, uploading data in VIVO it’s a manual process…. When we want to upload data we must manage manually 2 forms:

1. **csv to RDF**
   - **Input:** CSV file URL (e.g., file:///)
   - **Options:**
     - Comma separated
     - Tab separated
   - **Namespace:**
     - Specify namespace in which to generate class and properties
   - **Class Name for Resources**
2. **RDF to Ontology (VIVO)**
   - **Input:** RDF file URL
   - **Options:**
     - Specify destination model
     - Specify destination model for Tbox

---

**Note:**
- For "Ingest Menu > Convert CSV to RDF" and "RDF to Ontology (VIVO)", you can select different output formats such as RS_TEXT, CSV, TSV, RS.XML, RS.GSON, N-Triple, RDF/XML, N3, Turtle, JSON-LD.
SIGMA’s approach to upload data in VIVO
Uploading data in VIVO

Description of the SIGMA proposal

We avoided the need of the manual data entry filling the parameters needed by the forms and execute off-line a process that “simulates” the manual process.

Prepare the Servlet

Fill the servlet with parameters needed to execute the form 1 and 2

+ list of URL's, for the Hub

Validates if the upload is consistent

Yes

Process ends

No

VIVO

- Indexes
- Delete previous data (*)
- Upload the new information

(*) The previous published data is not deleted until we don’t have the new upload correctly validated.
SIGMA’s approach for uploading data in VIVO

Description of an automatized way to upload information in VIVO

Example of csv file
Where we have all the information. The first line are the metadata and the rest are the data.
SIGMA’s approach for uploading data in VIVO

Description of an automatized way to upload information in VIVO

This file uploads the information to the form 1. We inform each of the parameters of the form (csv to RDF)
SIGMA’s approach for uploading data in VIVO

Description of an automatized way to upload information in VIVO
SIGMA’s approach for uploading data in VIVO

Description of an automatized way to upload information in VIVO

We have also a file that contains the relations that we must verify to ensure the upload is correct.
SIGMA’s approach for uploading data in VIVO

Description of an automatized way to upload information in VIVO

What we achieve?

- We start the process and the system runs automatically until the information is loaded to VIVO, and it can be displayed.
- We minimally modified the VIVO process and it’s valid for any university and different sources, because we only need that they provide the correct csv.
- It’s easy and not depends on the number of data files do you want to upload.

Problems:

- Not runs for very big files (more than 1,000 lines) → We need to run a previous process to separate the data in many files.
- Accents and special characters in csv that not translates well in VIVO → we include also a converter in the previous process.
Timmings
SIGMA’s approach for uploading data in VIVO

Timmings

The following examples were for uploading data from some of SIGMA’s universities, but with a condition:

We first implemented the Experts guide with VIVO. We only upload the information required for the Experts Guide, i.e. we upload the researcher name, department, expertise and photograph and then, in a single field we group all the scientific activities of the researcher.

Example:
SIGMA’s approach for uploading data in VIVO

Example:

[Image: Screen capture of a website showing search results for experts and their details, including names and departments.]
SIGMA’s approach for uploading data in VIVO

Timmings

This tool, can also upload aggregated information from more than one university, we called “The HUB”. The hub works the same as the single experts guide, but also stores the university to which the researcher belongs.

Example:
SIGMA’s approach for uploading data in VIVO

Timmings
SIGMA’s approach for uploading data in VIVO

Timmings

In these conditions the timmings are:

<table>
<thead>
<tr>
<th>University</th>
<th># active researchers</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pompeu Fabra university of Barcelona</td>
<td>1.519</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Public university of Navarra</td>
<td>1.692</td>
<td>35 minutes</td>
</tr>
<tr>
<td>University of Valladolid</td>
<td>2.400</td>
<td>35 minutes</td>
</tr>
</tbody>
</table>

We observe that, as in UPF or UPNA we have at about 1.500 active researchers, in UVA we have more or less 2.400, the timmings are similar, so we can conclude that the number of researchers when we upload one university is not significant and for an Expert finding system could be at about half an hour.

To upload the hub, actually for 4 universities, in total more or less 6.300 researchers, the timmings are about 80 minutes.
SIGMA’s approach for uploading data in VIVO

Conclusions

- We believe that it’s important to work in the improving of the interoperability of VIVO with CRIS’s or other systems.

- We need to provide our universities an easier way to upload information.

- This has been a good test for the automation of the upload of the information in VIVO.

- We are working in the next step that is to have all the Scientific Portal, not only the Experts finder, using VIVO, where we will want to apply this upload too.
Next step: using CERIF
Problem:

We want to have standard files, using CERIF, that can upload automatically to VIVO and that don’t depend on the data source (or sources).

But we find that:

1. The mapping from CERIF model to VIVO model doesn’t exist.
2. CERIF model has a high complexity

Goal:

So, the goal was, can we map the CERIF ontology to VIVO ontology in an easy way that can provide an standard to upload data in VIVO? The answer is yes, though the CERIF2VIVO project.
The CERIF2VIVO project
The CERIF2VIVO Project

SIGMA is involved in the definition of the CERIF2VIVO project.

This project will be a collaboration between: euroCRIS, VIVO and SIGMA and open to other collaborations.

The goal of this project is to define an interface to upload information to VIVO always the same way independent of the source, so, the standard CERIF, seems the best option.

We have had a first experience with CERIF in the collaboration with CSUC to upload information in the PRC (Research Portal of Catalan Universities). Even though PRC only uploads a subset of the information that is stored in a CRIS (articles, books, book chapters, projects and thesis), it was a good experience.
The CERIF2VIVO Project

To do this project, we must avoid one of the current drawbacks of CERIF, that is its complexity. The CERIF model is big and complex, so, in parallel, it’s planned also the refactoring of the CERIF model.

This refactoring proposes to divide the CERIF model in a CORE model that will contain the essential entities, and then the definition of specific areas of entities that completes the model and that should be defined by experts in the area.

At the same time to do the refactoring, the mapping to VIVO will be defined.
Conclusions
VIVO

Conclusions...

- We can use **VIVO as a CRIS**, uploading all the information from some sources automatically or manually data entry.

- We can use VIVO as a **scholarly showcasing tool** (read-only), uploading the information from the CRIS of the organization (or more than one organization). Ideally we should use the CERIF model as the exchange format.

- euroCRIS will work in the **refactoring of the CERIF model** to simplify its use.

- In parallel, euroCRIS, VIVO and SIGMA will work in the **mapping of CERIF to VIVO ontologies**

- Five universities in the U.S. are working on **VIVO Scholar**, a modernized, read-only front-end for VIVO (for using as a showcasing tool)

- With these projects we will be able to upload information from any CRIS to VIVO in a consistent way (using the standard CERIF).