Mapping Existing Data Sources into VIVO

Pedro Szekely, Craig Knoblock, Maria Muslea and Shubham Gupta
University of Southern California/ISI
Outline

• Problem
• Current methods for importing data into VIVO
• Karma approach
• Demo
• Conclusions
Problem: Data Ingest

VIVO Data Ingest Guide:

Data ingest refers to any process of loading existing data into VIVO other than by direct interaction with VIVO's content editing interfaces.

Typically this involves downloading or exporting data of interest from an online database or a local system of record.
Current Methods for Importing Data into VIVO
VIVO Provided Ingest Methods

• Writing SPARQL Queries
  • Convert external data (e.g., CSV) into RDF
  • Map data onto VIVO ontology

= Programming

• Option 1: Convert data into predefined CSV format
  • Supports limited set of data fields
• Option 2: Edit existing XSL scripts for your data
# Example Data

## People

<table>
<thead>
<tr>
<th>person_ID</th>
<th>name</th>
<th>first</th>
<th>last</th>
<th>middle</th>
<th>email</th>
<th>phone</th>
<th>fax</th>
<th>title</th>
</tr>
</thead>
<tbody>
<tr>
<td>3130</td>
<td>Burks, Rosella</td>
<td>Rosella</td>
<td>Burks</td>
<td></td>
<td><a href="mailto:BurksR@univ.edu">BurksR@univ.edu</a></td>
<td>963.555.1253</td>
<td>963.777.4065</td>
<td>Professor</td>
</tr>
<tr>
<td>3297</td>
<td>Avila, Damien</td>
<td>Damien</td>
<td>Avila</td>
<td></td>
<td><a href="mailto:AvilaD@univ.edu">AvilaD@univ.edu</a></td>
<td>963.555.1532</td>
<td>963.777.7914</td>
<td>Professor</td>
</tr>
<tr>
<td>3547</td>
<td>Olsen, Robin</td>
<td>Robin</td>
<td>Olsen</td>
<td></td>
<td><a href="mailto:OlsenR@univ.edu">OlsenR@univ.edu</a></td>
<td>963.555.1378</td>
<td>963.777.9262</td>
<td>Assistant Professor</td>
</tr>
<tr>
<td>1538</td>
<td>Moises, Edgar Estes</td>
<td>Edgar</td>
<td>Moises Estes</td>
<td></td>
<td><a href="mailto:MoisesE@univ.edu">MoisesE@univ.edu</a></td>
<td>963.555.2731x3565</td>
<td>963.777.8264</td>
<td>Professor</td>
</tr>
<tr>
<td>2941</td>
<td>Brian, Heath Pruitt</td>
<td>Heath</td>
<td>Brian</td>
<td>Pruitt</td>
<td><a href="mailto:BrianH@univ.edu">BrianH@univ.edu</a></td>
<td>963.555.2800</td>
<td>963.777.7249</td>
<td>Associate Curator</td>
</tr>
<tr>
<td>2401</td>
<td>Claude, Elvin Haney</td>
<td>Elvin</td>
<td>Claude</td>
<td>Haney</td>
<td><a href="mailto:ClaudeE@univ.edu">ClaudeE@univ.edu</a></td>
<td>963.555.2902</td>
<td>963.777.9730</td>
<td>Curator</td>
</tr>
<tr>
<td>2070</td>
<td>Mosley, Edmund</td>
<td>Edmund</td>
<td>Mosley</td>
<td></td>
<td><a href="mailto:MosleyE@univ.edu">MosleyE@univ.edu</a></td>
<td>963.555.2945</td>
<td>963.777.9285</td>
<td>Assistant Professor</td>
</tr>
<tr>
<td>2561</td>
<td>Derek, Antoine McCoy</td>
<td>Antoine</td>
<td>Derek</td>
<td>McCoy</td>
<td><a href="mailto:Derek@univ.edu">Derek@univ.edu</a></td>
<td>963.555.2992</td>
<td>963.777.5454</td>
<td>Curator</td>
</tr>
<tr>
<td>1625</td>
<td>Hawkins, Callie</td>
<td>Callie</td>
<td>Hawkins</td>
<td></td>
<td><a href="mailto:Hawkins@univ.edu">Hawkins@univ.edu</a></td>
<td>963.555.3350x6480</td>
<td>963.777.4949</td>
<td>Professor</td>
</tr>
</tbody>
</table>

## Organizations

<table>
<thead>
<tr>
<th>org_ID</th>
<th>org_name</th>
<th>org_vivo_uri</th>
</tr>
</thead>
<tbody>
<tr>
<td>101028</td>
<td>Energy Sciences Department</td>
<td><a href="http://vivoweb.org/ontology/core#Department">http://vivoweb.org/ontology/core#Department</a></td>
</tr>
<tr>
<td>102017</td>
<td>Geothermal Technology Department</td>
<td><a href="http://vivoweb.org/ontology/core#Department">http://vivoweb.org/ontology/core#Department</a></td>
</tr>
<tr>
<td>100000</td>
<td>University of VIVO</td>
<td><a href="http://vivoweb.org/ontology/core#University">http://vivoweb.org/ontology/core#University</a></td>
</tr>
<tr>
<td>102000</td>
<td>College of Research</td>
<td><a href="http://vivoweb.org/ontology/core#College">http://vivoweb.org/ontology/core#College</a></td>
</tr>
<tr>
<td>101000</td>
<td>College of Science</td>
<td><a href="http://vivoweb.org/ontology/core#College">http://vivoweb.org/ontology/core#College</a></td>
</tr>
</tbody>
</table>

## Positions

<table>
<thead>
<tr>
<th>person_ID</th>
<th>job_title</th>
<th>position_type</th>
<th>department_name</th>
<th>department_id</th>
<th>start_date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2254</td>
<td>Assistant Professor I</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2003</td>
</tr>
<tr>
<td>2217</td>
<td>Assistant Professor II</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2008</td>
</tr>
<tr>
<td>1968</td>
<td>Assistant Professor III</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2002</td>
</tr>
<tr>
<td>2303</td>
<td>Assistant Professor IV</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2003</td>
</tr>
<tr>
<td>2303</td>
<td>Assistant Professor V</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2005</td>
</tr>
<tr>
<td>2146</td>
<td>Assistant Professor VI</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>1999</td>
</tr>
<tr>
<td>2895</td>
<td>Associate Curator of South Library</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2005</td>
</tr>
<tr>
<td>2682</td>
<td>Associate Curator of Central Library</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2010</td>
</tr>
<tr>
<td>3112</td>
<td>Professor of Kinetic Energy</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2008</td>
</tr>
<tr>
<td>3095</td>
<td>Professor of Elastic Energy</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2008</td>
</tr>
<tr>
<td>12</td>
<td>Curator of South Library</td>
<td>Faculty</td>
<td>Energy Sciences Department</td>
<td>101028</td>
<td>2008</td>
</tr>
</tbody>
</table>

[http://isi.edu/integration/karma](http://isi.edu/integration/karma)
VIVO Data Ingest Guide


Step #1: Create a Local Ontology

Data Ingest Menu

Step #2: Create Workspace Models

Step #3: Pull External Data File into RDF

Step #4: Map Tabular Data onto Ontology

Step #5: Construct the Ingested Entities

Step #6: Load to Webapp
Step #1: Create a Local Ontology

Data Ingest Menu

Step #2: Create Workspace Models

Step #3: Pull External Data File into RDF

Step #4: Map Tabular Data onto Ontology

Step #5: Construct the Ingested Entities

Step #6: Load to Webapp
VIVO Ontology

http://isi.edu/integration/karma

Pedro Szekely
VIVO Data Ingest Guide


Step #1: Create a Local Ontology
Data Ingest Menu
Step #2: Create Workspace Models
Step #3: Pull External Data File into RDF
Step #4: Map Tabular Data onto Ontology
Step #5: Construct the Ingested Entities
Step #6: Load to Webapp
Step#5: Construct the Ingested Entities

Write the following SPARQL query

Construct {
  ?person <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> 
      <http://vivoweb.org/ontology/core#FacultyMember> .
  ?person <http://vivoweb.org/ontology/core#middleName> ?middle .
  ?person <http://vivoweb.org/ontology/core#workPhone> ?phone .
  ?person <http://vivoweb.org/ontology/core#workFax> ?fax .
  ?person <http://vivoweb.org/ontology/core#workEmail> ?email .
}

Where {
  optional { ?person <http://localhost/vivo/ws_ppl_middle> ?middle . }
}
SPARQL Ingest Is Difficult

Construct {

?person <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> 
  <http://vivoweb.org/ontology/core#FacultyMember> .
?person <http://vivoweb.org/ontology/core#middleName> ?middle .
?person <http://vivoweb.org/ontology/core#workPhone> ?phone .
?person <http://vivoweb.org/ontology/core#workFax> ?fax .
?person <http://vivoweb.org/ontology/core#workEmail> ?email .
}

Where {

optional { ?person <http://localhost/vivo/ws_ppl_middle> ?middle . }
}

Construct {

?org <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> 
  <http://xmlns.com/foaf/0.1/Organization> .
}

Where {

}

Construct {

?position <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> 
  <http://vivoweb.org/ontology/core#FacultyPosition> .
?position <http://vivoweb.org/ontology/core#FacultyPosition>  
  <http://vivoweb.org/ontology/core#organizationForPosition> ?org .
?position <http://vivoweb.org/ontology/core#positionForPerson> ?person .

FILTER((?posthrid)=(?perhrid))

FILTER((?postOrgID)=(?orgID))
}

Pedro Szekely
Harvester Data Ingest

Program in XSLT

```xml
<core:positionInOrganization>
  <rdf:Description rdf:about="{$baseURI}org/org{$orgID}">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
    <xsl:if test="not( $this/db-CSV:DEPARTMENTID = '' or $this/db-CSV:DEPARTMENTID = 'null' )">
      <score:orgID><xsl:value-of select="$orgID"/></score:orgID>
    </xsl:if>
    <xsl:if test="not( $this/db-CSV:DEPARTMENTNAME = '' or $this/db-CSV:DEPARTMENTNAME = 'null' )">
      <rdfs:label><xsl:value-of select="$this/db-CSV:DEPARTMENTNAME"/></rdfs:label>
    </xsl:if>
    <core:organizationForPosition rdf:resource="" />
  </rdf:Description>
</core:positionInOrganization>
```
Karma Approach

Sources → KARMA → RDF → VIVO

http://isi.edu/integration/karma
Overall Karma Effort

http://isi.edu/integration/karma
Using Karma to Ingest Data into VIVO

KARMA

Pedro Szekely

http://isi.edu/integration/karma
Karma Benefits

Interactive

Easy

Fast

http://isi.edu/integration/karma
Karma Workspace

Model

Worksheets

Command History

http://isi.edu/integration/karma
Karma Models: Semantic Types

Semantic Types
Capture semantics of the values in each column in terms of classes and properties in the ontology

Karma learns to recognize semantic types each time the user assigns one manually

http://isi.edu/integration/karma
Karma Models: Relationships

Capture the relationships among columns in terms of classes and properties in the ontology.

The relationship between Position and FacultyMember is positionForPerson.

Karma automatically computes relationships based on the object properties defined in the ontology.

http://isi.edu/integration/karma
Using Karma to ingest data samples from the “Data Ingest Guide”

Karma Demo
Conclusions
Conclusions

- Generic data-to-ontology-to-RDF mapping tool
- **Easy to use**: interactive, no programming
- Used Karma to populate **USC VIVO instance**
- **Open source**: you can use it too

http://isi.edu/integration/karma
From Simon Gaeremynck, Sakai Foundation
More Information

• http://youtu.be/EQcMc4TrfuE
  • Using Karma to ingest VIVO data

• http://isi.edu/integration/karma
  • Publications and videos
  • Software download (open source)

• Contacts:
  • pszekely@isi.edu
  • knoblock@isi.edu