

Using the Karma data integration tool

The [Karma data integration tool](#) allows users to semantically model tabular data (i.e. spreadsheets) in a visual environment. This makes it easy to understand the best way to map data to ontologies and provides a visual comparison to the VIVO-ISF [Ontology Diagrams](#). This approach is especially useful for new VIVO adopters and those who prefer not to create and use their own scripts.

It's probably most common for data to be provided in spreadsheet format, which can be very simple to transform into RDF if each column of every row refers to attributes of the same entity, usually identified by a record identifier. The process becomes more complicated if different cells in the same row of the spreadsheet refer to different entities. This page includes example tabular datasets and screenshots of models created using Karma.

Modeling organizations, people, and positions

The following spreadsheet of organizations (one organization per row) is very easy to load into a VIVO:

org_name	org_ID
Accounting	8b9df681
Aerospace Engineering	e5c84f0a
Agriculture Leadership Education and Communication	2d953da5
Animal Science	2735b399
Anthropology	26ddd99c

You can readily imagine representing the information about each organization – id, name, contact information, and web site address – in additional columns. The Unique Resource Identifier (URI) used by VIVO to identify each organizational unit can be generated by using the org_ID and the institutional VIVO namespace. This is the starting point of creating the basic structure of your VIVO data.

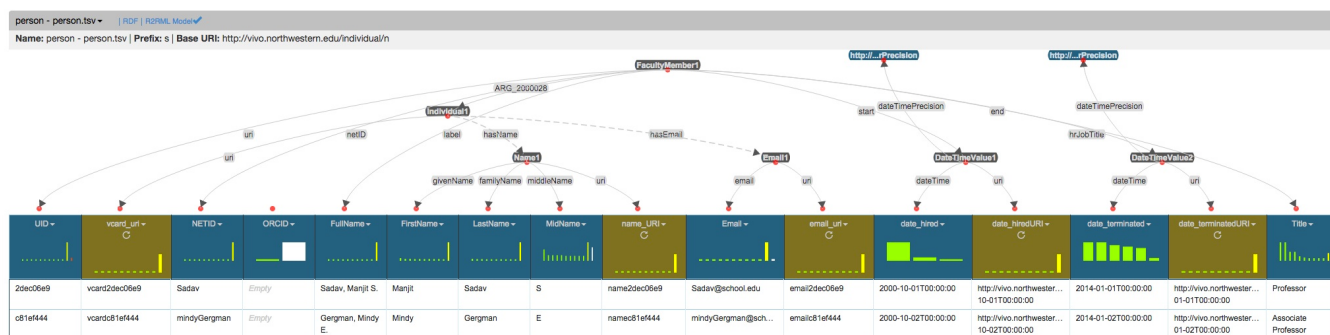
The next step is to store the information about people affiliated with those organizational units, and finally, their positions within those units.

A spreadsheet of people data typically looks like this:

UID	NETID	FullName	FirstName	LastName	MidName	Email	date_hired	date_terminated	Title
2dec06e9	Sadav	Sadav, Manjit S.	Manjit	Sadav	S	Sadav@school.edu	2000-10-01T00:00:00	2014-01-01T00:00:00	Professor
c81ef444	mindyGergan	Gergan, Mindy E.	Mindy	Gergan	E	mindyGergan@school.edu	2000-10-02T00:00:00	2014-01-02T00:00:00	Associate Professor
5c2ce0e	L_Vay	Vay, Marilda W.	Marilda	Vay	W	L_Vay@school.edu	2000-10-03T00:00:00	2014-01-03T00:00:00	Lecturer
b6f25071	Susan	Rach, Susan A.	Susan	Rach	A	sach@school.edu	2000-10-04T00:00:00	2014-01-04T00:00:00	Clinical Associate Professor
613870d1	Robert	Roberts, Kathryn J.	Kathryn	Roberts	J	kRoberts@school.edu	2000-10-05T00:00:00	2014-01-05T00:00:00	Assistant Professor
980e771d	JasenMaldonado	Maldonado, Jasen J.	Jasen	Maldonado	J	jasenMaldon@school.edu	2000-10-06T00:00:00	2014-01-06T00:00:00	Assistant Professor
e532a51f	Sahana20	Sahana, Krishan C.	Krishan	Sahana	C	Sahana20@school.edu	2000-10-07T00:00:00	2014-01-07T00:00:00	Visiting Associate Professor

In this spreadsheet the person identifier is called UID (your institution will have a different name for this identifier) and is the unique identifier for a person at your institution, by which that person is uniquely identified in databases at the your institution. NETID (your institution will have a different name for this identifier) is another identifier for a person, often used as a username for logging into university systems. The other columns are self explanatory.

In Karma, the model of this data set containing people information is shown in the image below:

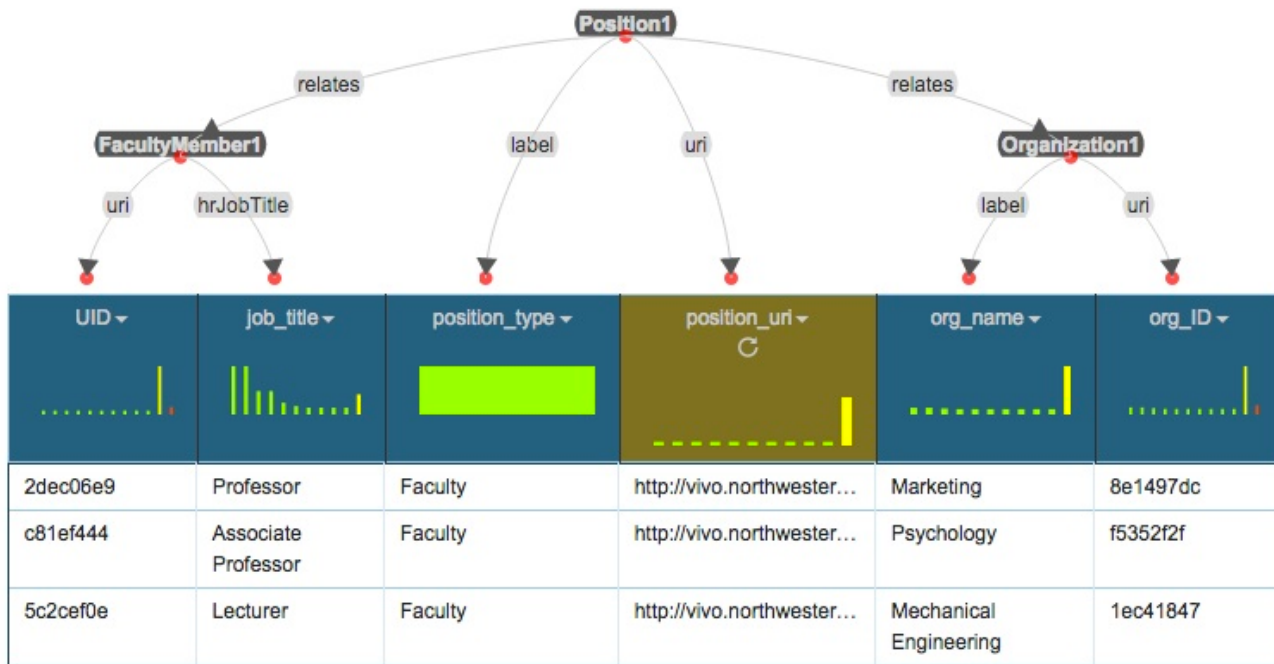


A spreadsheet of **people's positions** in the organization will look like this:

UID	job_title	position_type	org_name	org_ID
2dec06e9	Professor	Faculty	Marketing	8e1497dc
c81ef444	Associate Professor	Faculty	Psychology	f5352f2f
5c2ce0e	Lecturer	Faculty	Mechanical Engineering	1ec41847
b6f25071	Clinical Associate Professor	Faculty	Health and Kinesiology	ef68c136
613870d1	Assistant Professor	Faculty	Biology	533de40b

As one can notice in this spreadsheet we use the UID of the person and the org_ID of the organizational unit to connect the person with that unit and create the RDF statement containing information about that person's position.

The model of this data set containing information about **people's positions** within an institution is shown in the image below:

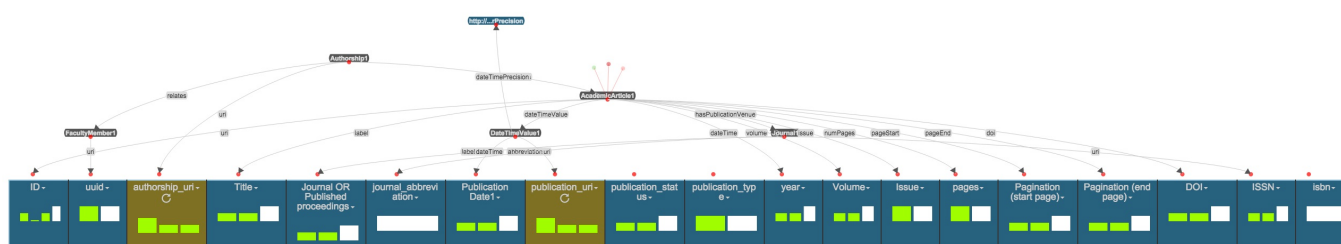


Modeling Academic Articles

A spreadsheet of academic articles is more complicated:

articleID	UID	Title	Journal OR Published proceedings	Publication Date1	year	Volume	Issue	pages	Pagination (start page)	Pagination (end page)	DOI	ISSN	pubmed	pubmedcentral
12345	90e6f6c0	Disruption O	Toxicology And Applied Pharmacology	2003-08-11T00:00:00	2003	23	1	8	263	271	10.1016/j.tai.0041-008x	0041-008x	34567	PMC12345
23456	90e6f6c0	Distinct Imm	Toxicology Letters	2006-02-01T00:00:00	2006	30	1	8	58	66	10.1016/j.to.0378-4274	0378-4274	45678	PMC23456

In this spreadsheet we have two important identifiers to connect the person with the article he/she authored: the articleID and the UID. Article ID can be any unique identifier assigned to the article. A model of this data set containing information about academic articles is shown in the image below:

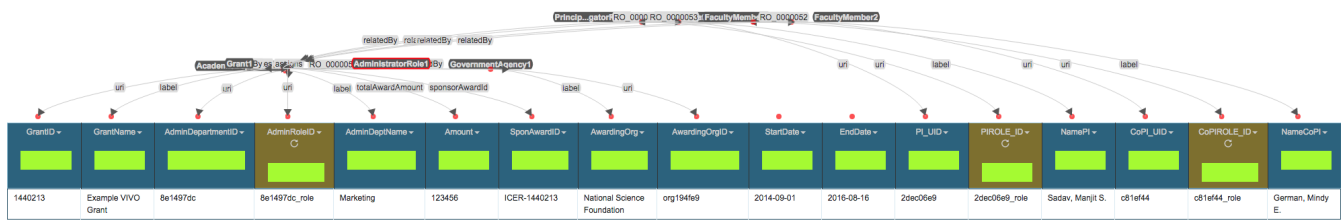


Modeling Grants

Grants also involve making a number of connections, plus adding two classes that are likely not in your spreadsheet, vivo:AdministratorRole and vivo:PrincipalInvestigatorRole (and maybe vivo:CoPrincipalInvestigatorRole).

GrantID	GrantName	AdminDepartmentID	AdminDeptName	Amount	SponAwardID	AwardingOrg	AwardingOrgID	StartDate	EndDate	PI_UID	NamePI	CoPI_UID	NameCoPI
1440213	Example VIVO Grant	8e1497dc	Marketing	123456	ICER-1440213	National Science	org194fe9	2014-09-01	2016-08-16	2dec06e9	Sadav, Manjit S.	c81ef44	German, Mindy E.

You must also assign a unique URI to these extra classes, shown below as URIs with the '_role' suffix in the gold columns. Below, the National Science Foundation has been modeled as vivo:GovernmentAgency. If you have a wide variety of funding organization types on a single spreadsheet source, you may want to create a separate spreadsheet and model for your funding organizations, or generalize the type to vivo:FundingOrganization. Likewise with the people modeled as vivo:FacultyMember below, you may generalize to foaf:Person.



Using PyTransform to create URI

Karma's PyTransform capability allows you to transform your source data using Python. A common use of PyTransform is to create additional unique URIs based off an existing one. The URIs with the `'_role'` suffix above were created using this Python code:

```
return getValue("AdminDepartmentID")+'_role'
```

More information on PyTransform is available [here](#) within Karma's documentation.

One example when you will need to use the PyTransform option is to create the **position URI** in order to create the needed triples for representing each person position within his/her institution. As you can see above in the **people's position** data example there are few columns that you need the values from to create the correct triples for the position. In the drop down menu found on each column you select the PyTransform option and in the window you type:

```
return "http://vivo.northwestern/position/n"+getValue("UID")+"_"+getValue("org_ID")+"_"+getValue("position_type").replace(" ", "_")
```

In this case the first part of the URI is the Northwestern University namespace which you need to change to your own namespace, such as `http://vivo.school.edu/individual/n` and then select values from three columns as shown above. Selecting values from three columns is necessary to identify positions uniquely, since most likely you have people at your institution that have multiple appointments and this allows you to create separate triples for each of their multiple positions within your institution.

Another example when you will need to use the PyTransform option is when you create the **authorship URI** for modeling the publications data as shown above. To do that you would want to select the PyTransform window found on the drop down menu on each column. Once you open the PyTransform window you type in the following:

```
return "http://vivo.northwestern.edu/authorship/n"+getValue("ID")+getValue("uuid").replace(" ", "_")
```

The first part is your namespace and you would want to change that with your own namespace. The `"ID"` and the `"uuid"` are the names of the columns from which we have decided to create the authorship URI and they represent the article unique ID and the person unique ID respectively.

External links

- [Karma Documentation](#)
- [Videos on using Karma for VIVO by Violeta Ilik](#)
- [Video on using Karma for VIVO by Pedro Szekely](#)