

VIVO ontology design principles

VIVO Ontology design principles

Reuse other ontologies

There are a large number of existing ontologies and new ontologies regularly appearing. By reusing existing ontologies, data in VIVO is immediately more interoperable with data from other sources, and we reduce the overhead of understanding what VIVO data represents.

[Ontologies Integrated into the VIVO Ontology](#)

Remain independent of specific domains

The VIVO ontology tries to remain independent of any specific discipline, since its primary use to date spans whole institutions, whether full universities or biomedical research institutions or government agencies.

A scientific resource ontology largely drawn from the top-level classes of the [eagle-i](#) ontology is included with VIVO, but research resources also span multiple disciplines.

VIVO also does not maintain an internal controlled vocabulary of subject terms, in the belief that it is more useful to reference terms in existing vocabularies selected by each institution adopting VIVO. Recent work at Stony Brook University funded through the VIVO mini-grant program provides a web service for resolution of a user's entry term to a UMLS concept identifier URI and label, and we anticipate supporting use of this service with VIVO 1.4.

Represent temporal relationships

The basic unit of data on the semantic web is a simple subject, predicate, object triple representing a single assertion of data. That simple assertion cannot itself contain additional information about who made the assertion or when it is true.

Supporting the capture of full provenance information about data in VIVO will take further development staged over several releases of the software; however, the VIVO ontology has adopted a partial solution through what we called context nodes.

Context nodes elevate the relationship between subject and object individuals to be an individual or node that can itself have properties such as a label, a time span, or a description. This pattern – essentially a reified relationship – has become an important way that the VIVO ontology can represent information about people's activities in their appropriate context.

For example, a person may hold two appointments in different parts of a university, each with its own title, start date, and end date. Rather than simply stating that the person *works in* each department, we create position individuals linked to both a person and a department and add appropriate statements about title and dates to the position.

This pattern is also used for roles in grants or other activities, and to represent authorship on publications, where author order may be significant.

Restrain the overall number of classes

The VIVO core ontology has fewer classes than many ontologies used primarily for classification and differentiation. The emphasis with VIVO is on finding commonality within and across institutions.

For example, while there are typically several ranks of professorships in the tenure process, VIVO does not differentiate these ranks in its core ontology since the activities and responsibilities normally apply to faculty members in general and not to specific ranks. We do differentiate faculty from other non-professorial researchers, however, because faculty often serve quite different functional roles in the institution.

Allow for local extensions

While the VIVO core ontology seeks to remain relatively small, emphasizing commonality across institutions, there are many local needs to express greater internal diversification. We recommend that specialization be done in a local ontology namespace and that classes be created as sub-classes of an existing core VIVO ontology class so that data will be inferred to be a member of the broader core class when harvested for indexing on <http://vivosearch.org> or for other purposes.

Some properties such as the institution-specific identifier property used for single sign-on are also better maintained in a local ontology.

Represent what we know

With the [open world assumption](#) of the semantic web, it is preferable to indicate what you do know about a person, organization, or any other type of individual in VIVO than to make statements that could well become false when additional information is discovered.

The most common example of this arises when representing unknown publication authors. Rather than typing these authors as an unknown person, we suggest identifying them simply as a foaf:Person and waiting until further information may be discovered about the person, as for example when the same name is found on another publication that might have an email address identifying the person's institution.

The general principle is that it's easier to state only what you know than to hunt down and remove statements which may have become false as more data becomes available.

Provide a way to distinguish what is internal to an institution

VIVO's menu pages are set up out of the box to show people, organizations, research, and events. These menu pages are usually assumed to highlight people, organizations, research, and events at the institution hosting VIVO, as a way for users on and off campus to browse researchers and their activities.

In practice, the data captured in VIVO instances very often also references people, organizations, research, and events from elsewhere – co-authors on publications, the universities people list as part of their educational training, the funding agencies for grants, and conferences people may have presented at around the world.

To avoid cluttering up the VIVO browse interface, the VIVO 1.3 application includes support for a locally-defined institutional internal class that may optionally be used as a filter on data showing up on menu pages. This provides a way to highlight local content without restricting the data VIVO can store.

Other principles

Comments and suggestions