Concepts, Keywords, Controlled Vocabularies, and Identifiers

Introduction

VIVO has an obvious need to represent subjects, keywords, terminology from controlled vocabularies, and identifiers – not just for people, but in relation to a variety of different types of entities.

Goals

- To clarify how VIVO uses and distinguishes keywords, subject areas, research areas, research interests, and expertise as different kinds of associations between entities in VIVO and entries for terminology, whether local to one VIVO instance, shared across many, or linked directly from another published vocabulary or any other location on the Web
- To review different options for annotating the associations between entities and terminology, perhaps using the Open Annotation Data Model (see Open Annotation in this wiki)
- To consider options for controlled vocabularies to reference from VIVO, including those already available for search and selection via web services

Free text vs. freestanding concepts

A number of data sources include keywords, descriptors, or other terminology that may be very formally structured or entirely free-form text strings. How they appear in data sources will depend on whether users select terms from a list or enter them by hand.

- A **free text keyword** is any word or phrase used to describe a VIVO entity of almost any type, including people, publications, events, processes, and organizations. The word or phrase is stored as the object of a vivo:freetextKeyword data property; the individual it describes is the subject of the statement.
- A **subject area** is an independent entity in VIVO typed either as a skos:Concept (if internal to VIVO) or an owl:Thing (if the URI of the subject area uses a namespace external to VIVO, such as an external vocabulary). External URIs are typed only as owl:Thing since they may be either classes or instances in the remote namespace.
- A **research area** is a subtype or subclass of subject area serving as a topic for research, not just describing what the associated resource is about.
- A **research interest** has not been defined in VIVO; in practice there is some precedent for use as a field of research where a person intends to work but has not yet established direct knowledge or experience.
- When a person's association with a research area is labeled as **expertise**, this normally implies that the person can demonstrate direct knowledge and/or experience with that research area. The VIVO ontology does not directly model expertise because of the subjective nature of the label and relative nature of the assessment. Expertise is a matter of interpretation and therefore challenging to represent in an ontology focused on objective facts and relationships.

Because of the variety of input typically made available to VIVO, the VIVO-ISF ontology supports both unstructured keywords (via the vivo:freetextKeyword data property) and freestanding SKOS concepts related to the subject of an RDF statement by the vivo:hasAssociatedConcept object property or its sub-properties, hasSubjectArea or hasResearch Area.

The advantages of using a SKOS concept:

- as a freestanding individual entity in RDF, the same concept can be related to multiple other entities – and in VIVO, users can go to the concept and see who and/or what else is linked to the same concept
- SKOS is a well-established W3C standard in very common use, and the SKOS ontology is used for representing many controlled vocabularies that are made available as RDF, including Agrovoc and the U.S. National Agriculture Library Thesaurus
- SKOS includes the relationships broader, narrower, exact match, and close match to link one concept to another. These relationships can be very helpful in expanding search results or finding linkages among what would otherwise not be recognized as close connections between any two other entities (people, organizations, publications) in VIVO

When does it make sense to use free text keywords?

- when the data simply don't warrant creating concept entities – if people are asked to type in keywords, they often enter compound terms or entire phrases that may not be easy to interpret as one or multiple concepts. "Sustainable international agriculture" is an example of a difficult phrase to translate into concepts.
- when it's important to show data exactly as it was originally gathered

Representing concepts in VIVO

If you have a relatively small number of concepts to represent, or they have a unique local origin or meaning, it may well make sense to create or import the list of concepts using the SKOS RDF ontology as implemented in VIVO.

Here's an example:

subject	predicate	object
http://vivo.cornell.edu/individual/bw324	http://vivoweb.org/ontology/core#hasResearchArea	http://vivo.cornell.edu/individual/n237736
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
http://vivo.cornell.edu/individual/n237736	http://www.w3.org/2000/01/rdf-schema#label	water chemistry

| http://vivo.cornell.edu/individual/n237736 | http://vivoweb.org/ontology/core#researchAreaOf | http://vivo.cornell.edu/individual/bw324 |

Identifiers in VIVO

The ISF/VIVO ontology supports common identifier types as data properties, and many institutions adopting VIVO also create a local extension for the local institutional id.