Representing Scholars and Scholarship

Mike Conlon, PhD
Professor Emeritus, University of Florida
VIVO Project Director, Duraspace
Outline

• What we do in VIVO
• Ontologies we use and create
• Data we use and create
• Issues in representing scholars and scholarship
What is VIVO?

- VIVO is ontologies, software, data, and a community for representing scholars and their works.

- Ontologies — 1 core ontology (VIVO-ISF), 3 additional VIVO ontologies, 22 related ontologies, including OBO and BFO.

- Software — a triple store for subject-predicate-object assertions, a web-based ontology editor, generic assertion management (Vitro) and scholarship assertion management and display (VIVO). See [http://github.com/vivo-project](http://github.com/vivo-project)

- Data — assertions about scholars and their scholarly works. At UF, 23M assertions for 63K people, 62K papers, 24K grants, 87K courses taught. 1.7M assertions of the form “uri isa BFO_0000001” See [http://vivo.ufl.edu/all-uf-triples.tar.gz](http://vivo.ufl.edu/all-uf-triples.tar.gz)

- Community — 130+ sites in 24 countries, 200K scholars, 1M+ works. See [http://vivoweb.org](http://vivoweb.org), Twitter: @VIVOCollab
Domains of Representation

The activities and accomplishments of your people

- Positions
- Educational Background
- Grants
- Courses
- Mentoring
- Awards and Honors
- Publications, Presentations, Books, Datasets
- Service

The people in your VIVO -- your faculty, staff and trainees, and their collaborators at other organizations

- People
  - Faculty
  - Staff
  - Trainees
  - Collaborators

Entities referred to in your scholarly work. Your internal organization, organizations related to the scholarly work of your people

- Dates
- Locations
- Journals

Organizations
- Internal
- Funders
- Publishers
- Employers
- Degree Granting
- Award Confering

Concepts
- Fields of Study
- Research Areas
- Keywords

Resources
- Equipment
- Facilities
- Services
Ontologies integrated into VIVO-ISF

VIVO – http://vivoweb.org/ontology/core
Basic Formal Ontology (BFO) – http://www.ifomis.org/bfo
Bibliographic Ontology (BIBO) – http://code.google.com/p/bibotools
Cell Ontology (CL) – http://cellontology.org/?q=download
Event Ontology – http://motools.sourceforge.net/event/event.html
Friend of a Friend (FOAF) – http://www.foaf-project.org/
Gene Ontology (GO) – http://geneontology.sourceforge.net/#code
Geopolitical.owl, from the U.N. Food and Agriculture Organization
Information Artifact Ontology (IAO) – http://code.google.com/p/information-artifact-ontology/
Ontology for Biomedical Investigations (OBI) – http://obi.sourceforge.net/ontology/Information/
Reagent Ontology (ReO) – http://code.google.com/p/reagent-ontology/
Relations Ontology (RO) – http://obofoundry.org/ro/
Software Ontology (SWO) – http://theswo.sourceforge.net/
Sequence Ontology (SO) – http://www.sequenceontology.org/
SKOS (Simple Knowledge Organization System) – http://www.w3.org/2004/02/skos/
Uberon (Uber anatomy ontology) – http://obo.svn.sourceforge.net/viewvc/obo/uberon/releases/
vCard – http://www.w3.org/TR/vcard-rdf/
26 Ontologies used in VIVO

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX swrl: <http://www.w3.org/2003/11/swrl#>
PREFIX swrlb: <http://www.w3.org/2003/11/swrlb#>
PREFIX vitro: <http://vitro.mannlib.cornell.edu/ns/vitro/0.7#>
PREFIX bibo: <http://purl.org/ontology/bibo/> 
PREFIX c4o: <http://purl.org/spar/c4o/> 
PREFIX cito: <http://purl.org/spar/cito/> 
PREFIX event: <http://purl.org/NET/c4dm/event.owl#>
PREFIX fabio: <http://purl.org/spar/fabio/> 
PREFIX foaf: <http://xmlns.com/foaf/0.1/> 
PREFIX geo: <http://aims.fao.org/aos/geopolitical.owl#>
PREFIX p1: <http://purl.org/dc/elements/1.1/> 
PREFIX p2: <http://purl.org/dc/terms/> 
PREFIX p3: <http://vivoweb.org/ontology/provenance-support#>
PREFIX obo: <http://purl.obolibrary.org/obo/> 
PREFIX ocrer: <http://purl.org/net/OCRe/research.owl#>
PREFIX ocresd: <http://purl.org/net/OCRe/study_design.owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX ufVivo: <http://vivo.ufl.edu/ontology/vivo-ufl/> 
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
PREFIX vitrop: <http://vitro.mannlib.cornell.edu/ns/vitro/public#>
PREFIX vivo: <http://vivoweb.org/ontology/core#>
PREFIX scires: <http://vivoweb.org/ontology/scientific-research#>
VIVO-ISF in VIVO

<table>
<thead>
<tr>
<th>element</th>
<th>VIVO-ISF</th>
<th>UF</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>169</td>
<td>29</td>
<td>633</td>
</tr>
<tr>
<td>ObjectProperty</td>
<td>103</td>
<td>14</td>
<td>267</td>
</tr>
</tbody>
</table>
Conlon, Mike  |  VIVO Project Director

Positions

- Emeritus Faculty, Clinical and Translational Science Institute, College of Medicine 2015 –
- Project Director, VIVO Project, Duraspace 2015 –
- Faculty, Clinical and Translational Science Institute, College of Medicine 2009 – 2015
- Co-director, Clinical and Translational Science Institute, College of Medicine 2013 – 2015
- Director, Biomedical Informatics, CTSI Transformational Programs 2009 – 2014
- Associate Director and Chief Operating Officer, Clinical and Translational Science Institute, College of Medicine 2010 – 2013
- Director of Data Infrastructure, Information Technology, Office of the Vice President and Chief Information Officer 2002 – 2010
- Interim Director, Clinical and Translational Research Informatics Program 2008 – 2010
A Person

• Has type(s) — faculty, staff, graduate student, etc

• Has a vcard for contact information

• Has various identifiers — UFID, SCOPUS, ResearcherID, ORCiD, eRACommons

• Photo, home department, research areas (concepts)

• Is related to positions, education, and roles in grants, papers, courses, organizations, and other activities

• The ontologies are sufficient to represent the relationships between the entities and properties of scholars and their works
VIVO Authorship: Connecting an Author with an Academic Article

- foaf: Person
  - rdfs:label
  - vcard: Individual
    - vcard:name
    - vcard:email
    - vcard:url
  - vcard:hasName
  - vcard:hasEmail
  - vcard:hasURL

- vivo: Authorship
  - rdfs:label
  - vcard: Individual
    - vcard:name
    - vcard:email
    - vcard:url
  - vcard:hasName
  - vcard:hasEmail
  - vcard:hasURL

- bibo: Academic Article
  - rdfs:label
  - vivo: relates
  - bibo: Publication Venue
    - hasPublicationVenue
  - bibo: Journal
    - rdfs:label
  - bibo: dateTimeValue
    - vivo: asString
      - http://www.w3.org/2001/XMLSchema#dateTime
    - vivo: testCase
      - http://vivoweb.org/ontology/core#yearMonthPrecision
      - http://vivoweb.org/ontology/core#yearPrecision
    - viva: datePrecision
      - http://www.jcs.org/jucs_19_12/article27
      - http://www.w3.org/2001/XMLSchema#anyURI

Data in VIVO

Entering data in VIVO involves “mapping” data values to appropriate places in the graph of VIVO data, either replacing literal values in the graph (eg, an updated phone number), or adding new nodes to the graph (eg, a new published paper)
Representing a mentoring relationship

• Example: Gyda is Erik’s doctoral thesis advisor since May 2014

• The data values:
  
  • Advisor (identified unambiguously through use of a property): Gyda
  
  • Advisee (identified unambiguously through use of a property): Erik
  
  • Type of relationship, such as “faculty mentoring,” “undergraduate advising,” “graduate advising”: graduate advising
  
  • The degree being pursued by the advisee: Ph.D.
  
  • Start date: May 2014
  
  • End date: No end date
Assertions

Inverse assertions and inferences omitted

n7924448848 a vivo:AdvisingRelationship

n7924448848 vivo:degreeCandidacy vivo:academicDegree4 (Ph.D.)

n7924448848 vivo:relates n2503 (Gyda)

n7924448848 vivo:relates n3209541243

n7924448848 vivo:relates n3739541303

n7924448848 vivo:relates n8066 (Erik)

n7924448848 vivo:dateTimeInterval n9616523741

n9616523741 vivo:start n2871342684 (May 2014)
Issue: The Middle Ground

- VIVO occupies the “middle ground” in ontologies

- Above VIVO-ISF (more abstract) is the BFO and OBO

- Below VIVO-ISF (less abstract) are domain specific ontologies and vocabularies (clinical trials, agriculture, cultural heritage, …)

- Most ontologies “below” VIVO introduce their own “middle concepts” — person, paper, project — these must be “reconciled” with VIVO.
Issue: Missing Ontologies

To represent scholarship, VIVO must represent detail that is best defined by authoritative sources

• Example: Academic Degrees
  • VIVO supports its own vocabulary of 173 academic degrees (because the world doesn’t have one)

• Example: Research Areas
  • Many vocabularies exist for expressing research areas of faculty members. What might be “useful” for faculty to find people working in related areas?

• Example: Contribution Ontology
  • Many people contribute to academic works, not just the “authors.” What are the contributions such as data analyst, study designer, copy editor, etc?
Issue: Boundaries of sub-graphs

To represent a scholar and the scholar’s work, we include the scholar (a graph) and the graphs of the works, but these lead to other scholars, and to graphs representing journals, or publishers of journals. What is the boundary for a graph representing a scholar?
Issue: Assertions about Assertions

We want to be able to make assertions about a graph of assertions for a scholar:

• “these assertions were made by the University of Florida”
• “these assertions are provided under license x”
• “these assertions were verified as of date y”