The Design and Development of an Integrated Researcher Profile System at Texas A&M to Enrich Scholarly Identity of Faculty

Dr. Bruce Herbert | Michael Bolton | Doug Hahn | Dong Joon Lee
Workshop Overview

Introductions

VIVO Basics (20 mins)
   Implementation at Texas A&M

Group Discussion: Use Cases & Stakeholders (15 mins)

VIVO Technical Aspects (30 mins)
   Ontologies
   Data Workflow
   Next steps

Group Discussion: Questions (15 mins)

10-minute break: 2:20 PM
“The value of scientific knowledge dispersed across the world can increasingly be captured by those who build networks to take the local to global scale and bring the global back for local impact.”

A semantic-web-based research and researcher discovery tool

People plus information on the research they do

Publicly-visible information, across disciplines

Serves both external and internal audiences

An open, shared platform for connecting scholars, research communities, campuses, and the world using Linked Open Data (LOD)
A Brief History of VIVO

2003-2005  First realization as a relational database for the life sciences at Cornell

2006-2008  Expansion to all disciplines at Cornell; conversion to Semantic Web

2009-2012  National Institutes of Health (NIH) grant VIVO: Enabling the National Networking of Scientists transforms VIVO to a multi-institutional open source platform

2013-2015  VIVO now a community-supported project under DuraSpace with open community development
Key VIVO Principles

- Open software
- Open data
- Open ontology
- Open community
- Local control
Integrates Institutional Data

- Organizations
- Scholars
- Grants/Projects
- Publications/Scholarly Works
- Teaching/Engagement
- Websites
- Portfolio/Vitae
- Expert Finding
- Network Analysis
- Ad hoc Queries
- Reports

VIVO
People

Riha, Susan Jean | Charles L. Pack Professor in the Department of Earth and Atmospheric Sciences

Positions

- Director New York State Water, Earth and Atmospheric Sciences (EAS), College of Agriculture and Life Sciences (CALS)

I am a professor in the department of Earth and Atmospheric Sciences, and joined the Cornell faculty in 1980. At that time, I was appointed the Charles L. Pack Research Professor of Forest Soils. My research interests are in the area of the interaction of plants with their physical environment and in dynamic simulation modeling. I work on both environmental and plant production problems on the state, national and international levels. I am a member of the graduate fields of Soil and Crop (... more)

Research Areas

biocomplexity, biogeochemistry, climatology, computational biology, crop management or crop science, earth science, ecosystem biology, environmental sciences, forest management, hydrology, information science, integrated crop management, integrated pest management, international agriculture, land use, soil and crop science, surface processes, sedimentary basins, & paleontology, sustainable development, weed science

Geographic Focus
And How They Connect
Structured Data for Visualizations

Co-Author Network  [GraphML File]

Profile

Riha, Susan Jean
Charles L. Pack Professor in I...
VIVO profile | Co-author network
132 Publication(s)
33 Co-author(s)
1980 First Publication
2010 Last Publication

Note: This information is based solely on publications which have been loaded into the VIVO system. This may only be a small sample of the person's total work.
Enter Data Once, Use Many Times
VIVO Basics: Implementation at Texas A&M

Dr. Bruce Herbert & Michael Bolton
## VIVO Implementation

<table>
<thead>
<tr>
<th>VIVO Implementation Interest Group</th>
<th><a href="mailto:vivo-tech@googlegroups.com">vivo-tech@googlegroups.com</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Outreach and Engagement Interest Group</td>
<td><a href="mailto:vivo-community@googlegroups.com">vivo-community@googlegroups.com</a> Information</td>
</tr>
<tr>
<td>VIVO Development Interest Group</td>
<td><a href="mailto:vivo-tech@googlegroups.com">vivo-tech@googlegroups.com</a></td>
</tr>
<tr>
<td>Apps &amp; Tools Interest Group</td>
<td><a href="mailto:vivo-tech@googlegroups.com">vivo-tech@googlegroups.com</a></td>
</tr>
</tbody>
</table>
TAMU Use Cases

- Discovery of expertise when building collaborative teams
- Organizational practices for faculty, departments and colleges
- VPR and TEES Proposal development
- Research funding compliance
- Informing Society
TAMU Data Sources

HR: people and their positions

Symplectic Elements Harvester: Faculty Publications

Registrar: courses

Faculty reporting: awards, professional service, education, research areas, research blurb

Institutional Repository: ETDs, publications, grey literature

Events calendar

MARCOMM: internal and external news

Extension: outreach, technology transfer

Research administration: grants & contracts
Data Stewardship

Manage data at its appropriate source with appropriate privacy

HR, grants management, registrar, graduate school, colleges and schools, research centers, extension

Department/agency/division/geographic location/research unit

Consciously derive public data for exchange

Engage stakeholders and build relationships

Recruit power users for training and local knowledge

Data that are visible get corrected!
Policy Issues

Representing faculty reputations
Dirty data
Lack of common definitions of organizational structure or who’s “faculty”
Data ownership
Many dimensions of privacy beyond simple “opt-in vs. opt-out”
Short-term “go it alone” vs. common good
Institutional risk
Researcher Profile Information Ecosystem

TAMU Data

Profile Editor
Faculty Data

Manual Input

Curator QC

SQL Database

Crosswalk

Reports

Harvest

DSpace

Elements

ORCID

Connecting Research and Researchers

Harvest

Widgets

Harvest

Widgets

PLUM Analytics

Connect share discover

Harvest

VIVO

Connect share discover

Harvest

ATM | LIBRARIES

TEXAS A&M UNIVERSITY
Scholarly Impact Metrics
Faculty Profiles in PlumX
Identity
Control
Evaluations
Strategic Decision Making
Best Practices for the Use of Scholarly Impact Metrics

Citation analysis and other bibliometric methods help justify your scholarly impact narrative by providing evidence of three characteristics of scholarship: scholarly output, scholarly impact, and the nature and development of scholarship over time or discipline. This guide provides research-based best practices for the use of scholarly impact metrics and recommendations on strategies to enhance the scholarly identity of researchers.

URI
http://hdl.handle.net/1969.1/156054

Subject
scholarly impact metrics, citations, bibliometrics

Department
University Libraries

Collections
Papers, Presentations & Training Materials [8]

Citation

The following license files are associated with this item:

Creative Commons

Usage
DSpace - Abstract Views: 222
DSpace - Downloads: 39

Social Media
Twitter - oaktrust.library.tamu.edu: 1

Except where otherwise noted, this item's license is described as CC0 1.0 Universal
Discussion: Use Cases & Stakeholders
VIVO Technical Aspects
VIVO Technical Aspects: Ontologies

Dong Joon Lee
What is Ontology?

- Ontology
  - A framework that defines a particular set of concepts, the relationships among them, and the nature of those relationships (Stewart, 2008)
  - An explicit and formal specification of a shared conceptualization (Gruber, 1993)
Ontologies

Concepts & Relationships

Authorship - Documents

Faculty - Position - Role

Affiliation - Course
Ontologies

Semantically-typed relationships

Authorship relates Article

Faculty has Position

Faculty bearerOf Instructor

Faculty unitOf Affiliation

Faculty teaches Course

Faculty relates Course
Ontologies

As a data map

Dr. Bruce Herbert

Director of Schol. Comm.

TAMU Libraries

GEOL 101

Instructor

has

servedBy

unitOf

relates

bearsOf

teaches

Pyrene sorption by water-soluble organic..
VIVO-ISF Ontology Classes
Why VIVO?

- Machine-readable Web?
- Defined concepts and relationships?
- Semantically-typed links?
- Data map?
- Searchable?
Semantically typed links
Title: Mineralogy and incipient pedogenesis…
Author: Bruce Herbert,
Date: 1988
Institution: University of California

Ontologies provide data

Source: Bizer, Chris. The Emerging Web of Linked Data
VIVO-ISF Ontology Classes
Question: Can I modify ontologies for my institution?

- Answer: YES
Scenario: Faculty wants to have a list of their chaired theses and dissertations with the students’ names.

- Relationship between Faculty and Theses
- Not a relationship between Faculty and Students
- Faculty’s Role on students’ theses
“ETDChairRole” in your Ontology

- Thesis URI 213
- Faculty URI 546
- ETDChairRole

Diagram showing relationships:
- Thesis URI 213 → ETDChairOf → Faculty URI 546
- Faculty URI 546 → ETDChairedBy → ETDChairRole
“ETDChairRole” in your Ontology

Resource Description Framework (RDF)

- One simple data model for publishing structured data on the Web
- A simple graph-based data model (RDF Triples)
“ETDChairRole” in your Ontology

Resource Description Framework (RDF)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETDChairRole</td>
<td>label</td>
<td>ETD Chair Role</td>
</tr>
<tr>
<td>ETDChairRole</td>
<td>supClassOf</td>
<td>Role (BFO_0000023)</td>
</tr>
<tr>
<td>ETDChairRole</td>
<td>type</td>
<td>Class</td>
</tr>
</tbody>
</table>

S <http://vivo.library.tamu.edu/ontology/TAMU#ETDChairRole>
P <http://www.w3.org/2000/01/rdf-schema#subClassOf>
O <http://purl.obolibrary.org/obo/BFO_0000023>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI_etdChairOf</td>
<td>type</td>
<td>Object Property</td>
</tr>
<tr>
<td>URI_etdChairOf</td>
<td>inverseOf</td>
<td>etdChairedBy</td>
</tr>
<tr>
<td>URI_etdChairOf</td>
<td>label</td>
<td>Chaired ETD</td>
</tr>
</tbody>
</table>
Data Modelling

- Using Karma Data Integration Tool
- A semantic modeling tool to visualize data with ontologies

* Data only for Karma visualization, not a real data
Data Modeling

- Using Karma Data Integration Tool
- A semantic modeling tool to visualize data with ontologies

Karma sources: DuraSpace, USC, How-to-use guide
VIVO Technical Aspects: Data Workflow
Douglas Hahn
Manage the unit tasked with helping figure out how to do...
A Brief History: Initial Implementation

Currently, we are on our second ramp up of a VIVO installation.

We were introduced to VIVO a few years ago and immediately saw the potential for its use.

So like anyone with a new toy. We started digging up any dataset we could find and load them into VIVO.
Initial Datasets

Scholars / People
Organizations
Publications/Scholarly Works
Everything else we could find...

People and Organizations
Human Resources Data
Feeds
Academic Analytics

VIVO

Scholars

Organizations

Publications/Scholarly Works

Everything else we could find...
Within a short amount of time we quickly had a VIVO instance running. We had loaded ~ 4K people. 37K publications.

We had no idea how to manage it, or what data was actually loaded in it.
An Opportunity to Start Over

Shortly after our initial launch of VIVO we experienced quite a lot of staff turnover.

We took this opportunity to re-evaluate how we were using VIVO and what we wanted out of it.

We wanted quality data, and the ability produce intelligent reports. To ensure this it was determined additional staff would be required.
Target Specific Data

For the second go round we intend to
• Target specific data.
• Develop tools, and workflows to manage the data.
• Stay focused and not chase new and “exciting” data at the expense of the quality of the existing data.
Specific Tools: Clean the Data

Clean Up Normalize Data with a variety of tools

Create various databases with authoritative lists

Use OpenRefine, basic programmatic scripts, and other text manipulating tools for Dirty Data.

The goal is a well documented, and reproducible workflow for dealing with raw data. Producing standardized data sets for working with VIVO.
Specific Tools: Model the Data

Clean Data

Model the data

Produce N-Triple

Test N-Triple by loading into VIVO

Repeat the Process till you have it right.
Semi-automatic Repeatable Process

1. Clean Up Normalize Data
2. Produce N-Triple
3. Load N-Triple into VIVO via the SPARQL Endpoint
Taking our Process to the Next Step

Empower the Faculty a simple, way to modify professional information. Allow some delegation of rights for departments to modify data.

Allow VIVO administrators to load additional data.
VIVO Default Permissions

By default VIVO allows for 6 levels of permissions and these are based on broad groups.

We needed a simple interface that would allow the modification of user data and the delegation of editing authority without sharing user accounts.
VIVO Profile Editor

We decided to leverage the existing infrastructure that was developed for some of the ingest process and allow faculty to edit the information there.

Place an edit icon on the profile that takes faculty member to the profile editor.
VIVO Profile Editor

A web front end allows for the editing of various elements in a simple clean user interface.
VIVO Profile Editor

We only expose very specific data elements through this interface. It allows us for delegated permissions. So a department can assist in updating faculty information. It also allows for curation of certain information before it goes into VIVO. Information can be automatically injested into VIVO via the SPARQL endpoint.
VIVO Other Uses

We also use the tool for curation / approval of other datasets.

Example: When our Dissertations are processed with VIREO and then published into DSPACE, we track the ETDChair. This allows us to harvest from DSPACE and then load it into our VIVO.
VIVO Other Uses
<table>
<thead>
<tr>
<th>Status</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unapp</td>
<td>A Study of Motivation Types and Behavior of Graduate Students in Future Faculty Preparation Programs</td>
</tr>
<tr>
<td>App</td>
<td>Updating Thesis / Advisor</td>
</tr>
<tr>
<td>App</td>
<td><a href="http://localhost:8080/vivo/individual/n152464">http://localhost:8080/vivo/individual/n152464</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#label">http://www.w3.org/2000/01/rdf-schema#label</a></td>
</tr>
<tr>
<td>Unapp</td>
<td>Published</td>
</tr>
<tr>
<td></td>
<td>A Study of Motivation Types and Behavior of Graduate Students in Future Faculty Preparation Programs</td>
</tr>
<tr>
<td>Status</td>
<td>Title</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Published</td>
<td>Artan, Sinem (2011–05). Aeolian Delivery of Organic Matter to a Middle Permian Deepwater Ramp Thesis</td>
</tr>
<tr>
<td>Unapp</td>
<td>Ray-Blakely, Charita Dionne (2011–05). A Study of Motivation Types and Behavior of Graduate Students in Future Faculty Preparation Programs Thesis</td>
</tr>
</tbody>
</table>
Data Workflow Diagram

Profile Editor
Faculty Data

SQL Database

Crosswalk
SQARQL

Data Editor

Manual Input

VIVO

Curator QC

Reports

Data Editor

Data

Data

Curator QC

Reports
Discussion: Questions for the Panel