Thursday April 6, 2017

Welcome, Intros, Orientation, Goals

Course goals:

1. Understand the purpose of VIVO
2. Understand basic VIVO concepts, including
    a. Relationship of VIVO to the organization and its members
    b. VIVO technology
    c. VIVO ontology
3. Understand VIVO implementation, including
    a. Purpose
    b. Scope
    c. Project planning
    d. Data acquisition
4. Understand how VIVO grows and prospers

VIVO Camp Curriculum (itinerary) here: https://goo.gl/bAO7vm

Wi-fi password: @cni-17s

Attendees (voluntary listing)

- Jay Brodeur, McMaster University Library brodeujj@mcmaster.ca
- Rachel Cascio, Fred Hutchinson Cancer Research Center (Seattle, WA) rcascio@fredhutch.org
- Armand Doucette, MIT Libraries doucette@mit.edu
- Brian Flaherty, University of Auckland Libraries & Learning Services b.flaherty@auckland.ac.nz
- Peter Mangiafico, Stanford University Libraries pmangiafico@stanford.edu
- Harry Snow, UNM-HSC-CTSC hsnow@salud.unm.edu ; UNM Health Sciences Center VIVO at https://vivo.health.unm.edu/
- Michaeleen Trimarchi, National Center for Atmospheric Research, trimarch@ucar.edu ; VIVO instance from NCAR's Earth Observing Laboratory: Arctic Data Connects, http://vivo.eol.ucar.edu/vivo/

Attendees (involuntary)

Introduction to VIVO

- Not only is VIVO code open source, completely free to download, but the VIVO community wants to help you make your VIVO site successful. Hundreds of members

can provide assistance, advice, and experience. We have a lot of resources--wikis, listservs, open calls, and community events-- to network with other VIVO institutions.

- Flexible and customizable: VIVO has been used in organizations of all sizes, from a single research group to a consortium of universities. VIVO adapts completely to the unique situation and requirements of your institution. Because it's open source, your organization decides what information to display on VIVO profiles, as well as which features to implement and how they are presented.
- Enterprise application: VIVO is a full enterprise application, capable of showcasing the work of institutions of all sizes. More than 150 organizations in 26 countries use VIVO to provide consistent, open, public information about research and scholarship.
- Authoritative, verified data: Using information from local databases to manage human resources, grants, teaching, and other systems, VIVO leverages existing institutional systems and populates profiles with confirmed information. Repurposing this data saves researchers time and takes advantage of investments in existing systems.
- Duke's search interface: https://scholars.duke.edu/scholars_search/?allWords=johnson
- U. of Wollongong: https://scholars.uow.edu.au/
- Q: Griffith: made a lot of customizations. How do you ensure that those are forward-compatible? A: We've tried to make the code as modular as possible.
- Graham: we're trying to improve the ability of implementation sites to make changes to the theme without running into problems with future upgrades
- Weill Cornell Medicine VIVO Dashboard - https://vivodashboard.weill.cornell.edu/ Code here: https://github.com/wcmc-its/vivodashboard
- Julia: at Duke, we use a Tableau instance to report on our data. We have SPARQL queries that load data into Tableau.
- Scholars@Cornell - will go live this summer. Focus will be on visualizations.
- Q: can you share some of these projects? A: The biggest barrier is whether the code is modular enough for someone else to use it.
- Mike: Duke's disclaimer about how they got the data is super useful.
    - "Some information on this profile has been compiled automatically from Duke databases and external sources. (Our About page explains how this works.) If you see a problem with the information, please write to Scholars@Duke and let us know. We will reply promptly."
- Duke's widgets
    - Open source code - https://github.com/OIT-ADS-Web/vivo_widgets
    - If questions, follow up with Richard Outten.
    - Example: This website pulls data from Duke Scholars. https://dibs.duke.edu/scholars/alejandro-aballay
- Griffith University
    - https://experts.griffith.edu.au/
    - Sample profile: https://experts.griffith.edu.au/academic/r.stewart
- Q: When we're looking at the above, is it VIVO, or a custom stack? A: the former.
- UNAVCO: http://connect.unavco.org
    - Thousands of public GPS datasets: http://connect.unavco.org/datasets

    - Research tab (publications): can sort by Altmetric Score
- University of Wollongong Australia: https://scholars.uow.edu.au
    - Sample profile: https://scholars.uow.edu.au/display/michael_adams
- The appeal of using the VIVO ontology is some sort of worldwide federated search product.
- 150 sites in 26 countries
- VCU's VIVO - http://communitynetwork.vcu.edu/
- Q: what has made VIVO useful? Julia: sharing data with others sites has had the biggest benefit to the institution (Duke).
- Mike: depends on the site. U. of Melbourne is focused on expert finding. U. of Florida is really focused on network analysis, University of California San Francisco focused on search engine optimization, providing high Google visibility for faculty profiles…. When you're pitching a VIVO to executives, it's good to solve a problem they know they have.
- Q: how do requirements change over the life of the project? Violeta: certain roles need to stay. Project manager, data wrangler, outreach person.
- Mike: my suggestion is to have a budget. If you don't have one, that's a red flag.
- Mike: you may have heard VIVO is difficult to install. That is a relic of the past especially with the implementation of Maven in 1.9.
- I see VIVO as being more of a stealth way to introduce semantic concepts.

Lunch

- Here's where the instructors are going to have lunch: https://www.yelp.com/biz/tucanos-brazilian-grill-albuquerque
- You are welcome to join us!

OpenVIVO

- http://OpenVIVO.org/
- A hosted that VIVO that anyone can join.
- Persistent identifiers for all entities - people (ORCiD), works (DOI or PMID), organizations (Grid), journals (ISSN), concepts (FAST)
- If something does not have an identifier, it is not in OpenVIVO
- You can login with ORCiD
- Every time you login, publications associated with your ORCID are imported
- You can select different roles for your works
- GRID
    - Has assembled 64k research organizations of the world. Includes geolocation, official names, alternate names; constantly putting out new updates
    - Available at http://grid.ac
- http://OpenVIVO.org/data/ published every hour
- Conforms with the FAIR Data Principles: findable, accessible, interoperable, reusable

- Q: Is there any dictionary that can we helps us to unified criteria about any element of the ontology? (ej. College, Principal Professor, researcher... etc)
- Graham: sometimes CrossRef will return arbitrary info when there's no DOI in CrossRef.
- VIVO Domains
    - Entities to be referred to in your VIVO
        - Journals
        - Locations
        - Organizations (UF had 17k organizations)
        - Dates
        - Degrees
        - Concepts
    - Your orgs, people, and locations in your VIVO
        - Local orgs
        - Local people
        - Local locations
    - Activities/accomplishments of your people
        - Positions
        - Mentoring
        - Awards and honors
        - Grants
        - Pubs
        - Service
        - Courses
        - Educational background

Introduction to ontology

- Source ontologies for VIVO: https://wiki.duraspace.org/display/VIVODOC19x/Source+ontologies+for+VIVO
- Vivo classes diagram: https://wiki.duraspace.org/display/VIVODOC19x/VIVO+Classes
- Ontology diagrams: https://wiki.duraspace.org/display/VIVODOC19x/Ontology+Diagrams
- Q: Why didn't you use another ontology for date times? A: Melissa Haendel said she couldn't find something better.

Data Sources

- Which data do you care about?
    - Citations
    - Patents
    - Books

    - Conference proceedings
    - Academic output
    - Positions (academic appointments, employment)
    - Name
    - Identifier
    - Courses
    - Educational degrees
    - Licensure
    - Invited presentations
    - Conflicts of interest (often a part of grants)
    - Intellectual property
    - Grants
    - Non-grant funding (contracts, company investment)
        - Mike: we have a concept at UF called outside employment
    - Advisory roles
        - Postdocs and grad students
        - External - corporate and govt bodies,
        - Availability for supervision
    - Editor at journal, journal reviewer
    - Social media presence and/or lab website; ORCID link
    - Expertise (could be used for identifying people who are media-friendly)
        - Fred Hutch: diseases and disciplines
    - External collaborators
    - Affiliations / memberships (e.g., Royal Society)
        - Centers, institute membership
    - Profile pictures
    - News mentions
    - Embedded videos (invited presentation, etc.)
    - Contact information
        - Email
        - Office and building location
        - Office hours
    - Sabbatical / on leave
- Create a tab here: http://bit.ly/vivodatasources
- What is an ideal data source?
    - Machine-readable
    - Clear data ownership
    - Authoritative
    - Accurate
    - Available
    - Maintained
    - Free
    - Accessible

- Complete
- Identified
- Lacks duplicates
- Ethical
- Channel for accepting feedback
- Provenance / clear metadata
- Transparent
- Ability to interact with data

Dinner

6:15 Hotel restaurant, Forque, is expecting us as a party of up to 16 people. Instructors are going and you're welcome to join us if you'd like.

Karma: First download the Karma data integration tool here:

https://github.com/usc-isi-i2/Web-Karma/wiki/Installation%3A-One-Click-Install

https://github.com/vioil/VIVO-Karma

on this github repo you can find a set of files that you need to model data with the Karma tool. Folder "TSV" contains files in tabular format that you model by applying the R2RML models for each of the TSV files with corresponding R2RML files from the folder titled "R2RML."
In the folder titled: "RDF" you have the product of the data modelling in RDF [N-Triples] format ready for upload in your VIVO Instance.

Note: you will need to preload the following ontologies in your Karma tool prior to attempting to model the data [ontologies are located in the shared google folder:bit.ly/vivocamp2017]
BIBOexport.rdf
Citationexport.rdf
CiTOexport.rdf
Eventexport.rdf
FABIOexport.rdf
FOAFindex.owl
OBOexport.rdf
rdf-schema.rdf
Vcardexport.rdf
VIVOCoreexport.rdf

VlocalVI.rdf - this is a local ontology that you are free to modify according to your own institution's namespace. You will need to replace the namespace from http://vivo.northwestern.edu/ontology/vlocal# to your own.

Examples of TSV files modeled are:

R2RML-Karma

R2RML models for modeling tabular data into semantic web VIVO compliant data. There are R2RML models for modeling person, organization, position and publication data sets. Some examples are shown below:

person - person.tsv can be used with a CSV or TSV file that has the following columns

UID || NETID || FullName || FirstName || LastName || MidName || Email || date_hired || date_terminated || Title ||

Where:

1. UID is the ID of the person

positions - positions.tsv can be used with a CSV or TSV file that has the following columns

UID || job_title || position_type|| org_name || org_ID ||

articles.csv can be used with a CSV or TSV file that has the following columns

ID || uuid|| Title || Journal OR Published proceedings || journal_abbreviation || Publication Date1 || year || Volume || Issue || pages || DOI || ISSN || isbn || pubmed || pubmedcentral || abstract

Where:

1. ID is the ID of the article
2. uuid is the ID of the person
3. Title is the title of the article
4. Journal OR Published proceedings is the title of the journal
5. Journal_abbreviation is the journal abbreviated title
6. Publication Date1 is the publication date of the article

7. year is the year of publication
8. Volume is the volume number of the journal
9. Issue is the issue number of the journal
10. pages is the number of the pages
11. Pagination (start page)
12. Pagination (end page)
13. DOI is the article unique identifier
14. ISSN is the journal unique identifier
15. isbn
16. Pubmed is the Pubmed ID of the journal article
17. Pubmedcentral is the PubMedCentral ID of the journal article
18. abstract is the abstract of the article

Friday April 7, 2017

Intro to SPARQL Part I

VIVO represents data in “triples” – subject, predicate, object
The ontology is a semantic model that describes the world of scholarship
For example (not actual VIVO-ISF):

- mike isA faculty-member
- mike wrote paper
- paper hasTitle “The Long and Winding Road”
- paper publishedIn “Journal of Irreproducible Results”

All things in VIVO have a Uniform Resource Locator (URI)

- In VIVO, URI usually look like: http://yourplace/individual/nxxxxxx
- Mike Conlon’s URI at UF VIVO is: http://vivo.ufl.edu/individual/n25562

SPARQL Queries - Examples

Query 1

```
# Returns all faculty members
# “Ask for the x, where the x looks like ____.”
SELECT ?x
WHERE {
      ?x a vivo:FacultyMember .
}
```

Query 2

```
# Returns faculty members and their names (hasName predicate)
SELECT ?x ?label
WHERE {
      ?x a vivo:FacultyMember .
      ?x rdfs:label ?label .
}
```

Query 3

```
# Returns an error
# When using GROUP BY, we need to say what we want to have
happen
# to the potentially multiple values of labels
SELECT ?x ?label
WHERE {
      ?x a vivo:FacultyMember .
      ?x rdfs:label ?label .
}
GROUP BY ?x
```

Query 4

```
# Cannot use ?label twice on the same line
# Change the second occurrence to another word,
# e.g., labelSorted
SELECT ?x (MIN(?label) AS ?labelSorted)
WHERE {
      ?x a vivo:FacultyMember .
      ?x rdfs:label ?label .
}
GROUP BY ?x
ORDER BY ?labelSorted
```

Query 5

```
# Returns only current faculty that have at least one label
# “a” is equivalent to “rdf:type”
# Red things are SPARQL elements, orange are ontology prefixes
SELECT ?x (MIN(?label) AS ?labelSorted)
WHERE {
      ?x a vivo:FacultyMember .
      ?x rdfs:label ?label .
      ?x a ufVivo:UFCurrentEntity .
}
```

```
GROUP BY ?x
ORDER BY ?labelSorted
```

Query 6

```
# COUNT the URI to count the UF Faculty (distinct values of x)
# Use DISTINCT to be sure you are counting values of "?x" not
# rows in a query result
SELECT (COUNT(DISTINCT ?x) AS ?nfaculty)
WHERE {
      ?x a vivo:FacultyMember .
      ?x rdfs:label ?label .
      ?x a ufVivo:UFCurrentEntity .
}
```

Query 7

```
# Get all the triples for a specified subject URI
# (any predicate, any object, specific subject)
SELECT ?p ?o
WHERE {
      <http://vivo.ufl.edu/individual/n8281411678> ?p ?o .
}
```

Tips:

- Write the most restrictive clause *first*.
- To view the returned data as a spreadsheet, return as a CSV file and open it in Excel.
- Use sample SPARQL queries as a jumping-off point: http://mconlon17.github.io/sparql/
- To experiment with UF data, download http://vivo.ufl.edu/all-uf-triples.tar.gz

Planning a VIVO Implementation

- Will you change data before putting it in VIVO? This may be called "data cleaning" -- but every change results in data in VIVO that is different from data in the system that provided data to you (i.e., VIVO data will differ from original source data). Some changes are not controversial (e.g., converting uppercase to mixed case). Other changes may be more difficult to explain.
- Link to paper showing network analysis done using VIVO: http://dx.doi.org/10.1111/cts.12267
- Link to Scholars@Duke code about generating CVs: https://github.com/OIT-ADS-Web/scholars_cv_generator
- Link to Scholars@Duke repo: https://github.com/OIT-ADS-Web

Putting Data into VIVO

- Link to VIVO Pump, a tool for putting data into VIVO. http://bit.ly/vivo-pump

- Graphs are collections of triples. You can put triples into separate graphs, if you like. You can then remove an entire graph, and reload it. VIVO SPARQL combines all the graphs when you are running a query. VIVO puts its "inferred" triples in an inference graph. The main VIVO data graph is called http://vitro.mannlib.cornell.edu/default/vitro-kb-2 The inference graph is called http://vitro.mannlib.cornell.edu/default/vitro-kb-inf You typically do not need to know the names of the VIVO graphs.
- VIVO Pump "get" from VIVO returns a spreadsheet.
- Karma is a tool for making RDF from spreadsheet data. R2XML files are available on GitHub for standard VIVO scenarios.

Engagement

- VIVO project team should cultivate relationships with data owners because common requests for support are to update data fed into VIVO from other sources
- Duke's support model relies on power users, often administrative assistants, who refer questions from their units / departments to the VIVO help desk
- Duke hosts monthly meetings with coffee for power users to showcase new features and encourage discussion and networking
- Crisis management
    - University of Florida: rare instances of complying with court orders to remove contact information from VIVO within one hour
- Weill Cornell: always thank the person who takes the time to give feedback
- Grants: issue of awards vs. sub-awards and label of "Principal Investigator"
- Stanford: keep in mind the purpose of the system, to better the lives of faculty & students
- University of Florida: presented VIVO for approval to UF Faculty Senate; pitched first to the Senate Steering Committee to get on the agenda of full Senate; try to get a sponsor or champion. A "gift" to the faculty -- automated profiles, high SEO, democratized data on grants, teaching, publications. Well-received by the faculty.
- Focus groups: offer options to encourage people to express their preferences and jump-start conversations
- User experience (UX) expertise can be very helpful
- Know your audience: grant-funded researchers (especially in medicine and engineering) usually have up-to-date CVs/resumes so pitching VIVO as a source for profiles or CV generation doesn't work for these people; however, many faculty in other disciplines don't keep their CVs/resumes up-to-date or maintain non-VIVO profiles so they are likely to welcome the help that VIVO can provide
- Graphics-focused marketing materials are more effective, with links to more information that can be more text-intensive
- Faculty don't read email announcements unless the sender is their department chair, their students, or fellow faculty
- Duke's Elements system sends monthly notices of pending publications for faculty approval, and response has been good
- Consider the "opportunity cost" - time spent on administrative tasks (e.g., requests for feedback on VIVO) takes times away from research, teaching, patient care, family time

- Duke's Twitter feed: @scholarsatduke

Some references

Learning SPARQL by Bob Ducharme

Building Ontologies with Basic Formal Ontology by Robert Arp, Barry Smith, and Andrew D. Spear

Semantic Web for the Working Ontologist, Effective Modeling in RDFS and OWL by Dean Allemang and Jim Hendler

Saturday April 8, 2017

VIVO Trajectory and Roadmap

- VIVO Charter: https://goo.gl/YEOxh9
- VIVO Strategic Plan: https://goo.gl/y47Vtc
- VIVO Roadmap: [need address] https://goo.gl/

- Leadership Group meets quarterly: in person at Duraspace Summit and VIVO conferences plus two conference calls
- Steering Group meets weekly
- Task Forces: two to seven, at any given time
- Committers Group: eight or nine trusted technical people, invited by Graham (potential members are proposed then reviewed by group), core code
- Community Projects - https://wiki.duraspace.org/display/VIVO/Community+Projects
- Interest Groups hold calls every four weeks on Thursdays
    - Apps & Tools
    - Development
    - Engagement
    - Implementation
- VIVO 1.10 Release Planning: https://goo.gl/ivLzGI
- Federated search: big step forward with forthcoming triple pattern fragment capability

VIVO Community and Events

- Interest Group calls every Thursday
- Email lists / Google Groups
- Annual conference
    - Community-focused, not driven by vendor presentations

    - Inclusive: posters & talks open to in-process projects, not limited to finished products
    - 8th Annual Conference, August 2-4, 2017, NYC at Weill Cornell
- VIVO Wiki: https://wiki.duraspace.org/display/VIVO/VIVO
    - Technical documentation & community resources
    - VIVO 1.9.x Documentation (current release)
- Harvard Catalyst WebAPI for identifying PMID for people http://profiles.catalyst.harvard.edu/docs/ProfilesRNS_DisambiguationEngine.pdf
- Duraspace membership
- Join us!


EXI