***Note: the group notes can be found here: http://bit.ly/vivocamp17 **
Link to Itinerary here

Thursday - Day 1

0830 - Introductions (Mike Conlon, Julia Trimmer)

Instructors:
Mike Conlon
Julia Trimmer
Graham Triggs - Duraspace
- Interesting note: He began working with

Violetta Ilk - chair of conference committee
Paul Albert - Cornell Medicine - leader of implementation group

Who's here:
- Stanford - Lots of Stanford people here! Looking to expand in-house Profiles platform to include other information (grants, research data, IR) --looking at VIVO as a starting point.
- NCAR
- Universidad del Rosario in Columbia
- UNM (interested in a new harvesting process for VIVO)
- MIT - no VIVO, but about to implement Symplectic Elements.
    - Office of institutional research has been building their own sort of VIVO, so trying to find a solution that works for both.
- University of Auckland
    - SE subscriber for a number of years--looking to replace a bespoke profile system

0900 - Intro to VIVO (Julia Trimmer)

Open VIVO - open instance of VIVO hosted by Duraspace.

VIVO is a way to find researchers/experts and is centred on a standard data model
- VIVO Allows people to find experts and expertise
- A way to make research output discoverable and accessible

VIVO is open
- Open source, open community, open data
- Flexible and customizable

- Enterprise application
- Authoritative, verified data

How VIVO data can be shown:

- Network map of science
- Map of research (geographical)

VIVO data is linked

- The links allow users to browse the connection between researchers and their output

VIVO can:

- Include researchers in all disciplines
- Can be internationalized
- Can consume data from ORCiD and Figshare
- Be used on mobile devices

Customizations

Griffiths University - Capability Map

- Question: Is there a worry about doing too much 'in-house' customization, since it might not be able to be moved to a newer VIVO version?
    - A: It really does depend on how the customization is done (and what is done). The VIVO development team is working toward a modular approach that will minimize these kinds of issues in the future.
    - A2: If your institution is a member of Duraspace, you can sit on the Governance team and have your say in the community.

Mountain West Research Consortium

- Search works across a number of VIVO sites — good example for a country or a university system in a state. ***This is probably what RDC is (or should be) working toward***

Weill Cornell Medicine — Citation index tool

- Displays publications
- Filter by date, type and journal ranking
- Search by author, journal
- Pulled from their upstream data (but uses VIVO data model). Pull into VIVO dashboard, updated daily. Displays information in a Dashboard
- Can filter the dashboard view (e.g. only certain department) by querying against the profile information in VIVO
- CItation Index Tool - standardized citations by typical # citations for pub type (article vs. review) and Journal — shows results in percentiles of comparable articles.
    - Can do for individual departments or authors
- Not released yet—hoping to go live in a couple of weeks.
- VIVO extension — not part of the VIVO software
    - Q: Why separate extension?
        - A: Some information is private — consider VIVO public data, and it may be complex to adapt VIVO to keep some data hidden
        - A: Developers more familiar working with php
    - UNM — interested in learning how to export data from VIVO into other systems
        - May run into performance issues?

    - Scholars@Duke: Use SPARQL queries pull from VIVO and into Tableau for visualization

Scholars@Cornell

- Revision of VIVO — focused on visualizations
- Collaboration Wheel, CIrcular diagrams, keyword clouds

Demos

S@D

- Faceted search — displays search results by type
    - E.g. Publications, Artistic works
    - Note: Artistic works are manually entered into VIVO by faculty members or delegates.
- How do they populate “In the News” for each researcher — they worked with Web Services and News Office to set up a Drupal site that allows the news folks to tag the articles with Duke IDs.
- Disclaimer on every page:
    - “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.”
- All sorts of widgets that allow users to import and export information
    - E.g. his information in VIVO is used to populate his personal research webpage via the widget.

Griffiths University

- Very stylized front-end —> see this example.
- Q: Are we looking at something that is built within VIVO, or are we looking at a separate platform that pulls information from VIVO and styles it.
    - A: All of this (for Griffiths and S@D) are done in VIVO
- Comment: Showing the new publications on top of the cumulative publications is a nice way to show data — you don’t end up with ‘precipitous fall’ in the current year.
    - The standard graph of #pubs per year is called the “sparkline”
- Q: How does the community keep up on what’s going on:
    - Google group
    - Conference calls
    - Wiki page — apps and tools
    - Conference

Q: How are people changing/modifying their VIVO instance

- A: Forking VIVO has not happened --everyone works with the mainline code but makes their own custom modification with it.
- Typically, this means that older sites are a bit behind in version and the newer ones are using newer versions
- What’s more concerning is when the data model changes — this happened once in a past and it caused incredible difficulties (especially for the big-time implementers).
    - Task force dedicated to

University of Wollongong

- Example of a profile
- Have a button indicating availability as a research supervisor.
- Q: Where does information for “available for supervision” come from?
    - A: At least at UFlorida, School of Graduate Studies has information on advisor qualification and availability for supervision.

1015 - Intro to VIVO 2 (J. Trimmer)

Why not integrate data?

- Decentralization is easier!
- Independent missions / data silos

All Universities track:

- Employees and faculty records
- Organizations
- Grants/projects
- Students and courses

Units manage:

- Faculty reporting
- Websites for labs, faculty, depts
- Academic programs

Integrating institutional data
Input:
Output: Websites, expert finding, portfolio / CV, Network analyses, ad hoc queries, reports

Integrating data privately

- Annual report, data for websites
- Dashboards, reporting, analyses
- Institutional reporting, strategy development

… vs. Integrating data publicly

- Research discovery for local and global communities
- Re-purposing, correcting public data
- Promotes faculty, programs, institutions

Tools and systems

- Profiles RNS
- PURE
- Elements
- Activity Insights
- Converis
- Homegrown
- VIVO

Why use VIVO?

- Open source: license allows VIVO to be modified and shared.
    - BUT needs resources
- Ontology--a shared data model that includes relationships

- Linked open data:
    - Machine readable format
    - Public
    - Easier to combine with other data sets
- Cross-disciplinary, extendable, internationalized
- Build on community resources
- Participate in the ecosystem of open scholarship

Q: Why not just use front-end of PURE or Elements (when they implement it)?

- PURE does a pretty good job with disambiguation, but not for everyone (and the modifications seem to be overwritten)
- PURE is not very extensible
- Elements front-end —> will probably be a front-end profile.
    - This is only one part of what VIVO can do-- VIVO also serves as a data source for other systems.
- VIVO provides the option for a federated search at some point in the future.

1030 - Getting your VIVO Project started

- Currently 150 institutions in 26 countries

Evaluating Tools:

- Assess needs
    - Duke did needs assessment with 25 faculty members
- Clarify goals
- Define success parameters
- Choose tools that align best

VIVO success indicators:

- High-level sponsorship
- Commitment to ongoing support
- Alignment with an institutional plan

Q: How to get to sponsorship and support

- Step 1: Create a successful demo
- Step 2: Keep track of complaints from key stakeholders and come back to them to let them know how VIVO now meets their needs.
- Step 3: Figure out how VIVO can be used to satisfy other compliance measures / regulations
    - E.g. meeting open access policy requirements.

Example Virginia Commonwealth University - http://communitynetwork.vcu.edu/

- Provide a map of where community-based research is taking place.

Sustainability

- Resources available?
    - Programming / open source experience
    - Project mgmt
    - communications/support
- Capacity for the long haul?

It's all about goals:

- What's most important to your institution?
    - Public-facing web profiles?

    - Research discovery?
    - Platform for Open Science
    - Others
- Important to review (and revise) goals
- Changing goals midstream is a recipe for trouble!
- It's important to record requests (and who asked) and send them through the Project Manager and the Governance team.
    - And perhaps, have some way to assess the difficulty and potential benefit from the request.

Goals for S@D:

1. Create public-facing faculty profiles
2. Provide scholars data to schools and departments
3. Help populate the Open Access Repository with full-text publications

Rollout strategies:

- Broad and shallow
    - Lots of people, less data
    - Add data over time
- Narrow and deep
    - Fewer people, more data
    - Add people over time

Ideas for examples

- Make uncleaned data available to users in non-public area, use VIVO to encourage them to clean up their data
- Display message history

Assembling (and maintaining) a Project Team

Julia's recommendation for an ongoing team:

Technical:

- Programming lead (0.5-1 FTE)
- Programmers/developers (1-3 FTEs / 2 bodies)
- Database administrators
- Systems administrators (networks, servers)

Project / Business:

- Project or product manager (0.5 - 1 FTE)
- Data analyst or curator, plus ontologist (0.5 - 1 FTE)
- Communicator or trainer or outreach person (0.25 - 1 FTE)
    - Provides classes, individual meetings with users

Mike Conlon's thoughts:

- You NEED a person who is willing to get down and dirty with the data.

Paul Albert's thoughts:

- Ideally, you're aligning VIVO with the larger data ecosystem in the institution
- Who cares about data at the institution? How can we get everyone on the same page?
- VIVO's strict ontology exposes data ambiguity in your systems — it's an ongoing issue.

Other allies

- Librarians, scholarly communications officers
- Chief academic officers
- Grant offices, other data stewards
- News office, communications
- Web designers, corporate identity
- Anyone who cares

Budgeting and Funding

- Costs involved: IT infrastructure and HRs
- Always estimate pessimistically
- Funding model can evolve (starter funding now, ongoing funding later)
    - Provost, library split for S@D
- Contributions by % of faculty organizaiton
- Budget for Duraspace membership

MC's tho
ughts: HAVE A BUDGET and have a commitment!

Intro to VIVO 3 - An introduction to the archetecture

Application server is a Java application server

- Delivers information to the web server
- Vitro within the application server
    - is a generic semantic web
- VIVO within application server
    - Vitro + ontology

At Duke:

- Many different virtual machines for all of the pieces in VIVO

<table>
<tr><td>Your Vivo site</td><td></td></tr>
<tr><td>Web Server</td><td></td></tr>
<tr><td colspan="2">Ontology Layer
<table><tr><td>VIVO</td><td>Vitro</td></tr></table>Application server</td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td>login</td><td>database</td></tr>
</table>

VIVO Technology Stack:

Shibboleth +
MySQL Databases (SDB/TDB)

- Content is in SDB and configurations are in TDB
- Jena used to translate the SQL database into the application server.

Apache Tomcat (Java Servlet)

- VIVO (specialized application)
- VItro (general purpose application) - a set of extensions included in VIVO
    - You can put any sort of ontology that you want in there.
        - Some Univs have used it for library catalogue
        - Spaceships, keeping track of clinical trial info
    - But Vitro is general, so it has no information about the data beyond the ontology
    - What VIVO does is it puts structure (scripts, templates, collections, etc.) on top of Vitro to present the data the way that we want/expect it.

Apache HTTP (Web server)
http://….. (your web site)

MySQL is easier to use for most folks to implement, but a SPARQL server may be needed if we're pushing in or out a lot of information to/from VIVO

Maven is the build automation tool to install entire stack or to build and deploy specific parts into the application server

Data Ingest

Harvester

- Standalone application
- Set of tools for taking data from a data source and converting to RDF & loading into content triplestore

Who is the authoritative source of the data (what is the system of record)?

- E.g. VIVO may be system of record for faculty overviews, but probably not for other data (e.g. grant information)
- @UF the culture has changed so that everyone is a data steward.

Important Terms

- Semantic web
    - Collection of unique identifiers and relationship descriptors between entities
- Resource Description Framework (RDF)
    -
- Triples
    - Everything in VIVO is represented as a triple
    - Triples work in both directions
    - In relational database — things in rows and columns. Connect with a common identifier between tables

        - In semantic web, relationships don't need to rely on tables.
    - It is possible to get use to the use of RDF and triples.
- Triplestore

Q: How to deal with administrators that are RDF-averse (because they don't understand it):
- Any competent IT shop can handle information in this format
- Identities and relationships (focus less on triples)

1315: OpenVIVO and an introduction to data

How to log onto OpenVIVO
- Log in with your ORCiD

OpenVIVO goals:

1. Provide a VIVO experience for everyone. A demonstration, a platform for experimentation, ownership experience for the VIVO team
2. Use persistent IDs for all entities
    - People (ORCiD)
    - Wokds (DOI / PMID)
    - Orgs (GRID)
        - https://grid.ac/
        - Official name, identifiers, location information for nearly all research organizations in the world (~64000)
        - Used the VIVO RDF model to express the GRID data.
        -
    - Journals (ISSN)
    - Concepts (FAST)
3. Automatic, real-time ingest of metadata from identifiers via public APIs
4. Attribution of works by scholars to indicate roles in works
    - Want to be able to express a role in a publication other than author
        - E.g. I wrangled data; I drew the graphics
5. Publication of data
6. Consumption and reuse of data

What they did

- http://openvivo.org/
- Not much styling
- Some editable fields removed, so that people couldn't add information not linked to data with persistent IDs.
- As soon as you log in with ORCiD, it pulls in publication data from Crossref, using information from ORCiD. Regularly checks ORCiD profile against existing publications in VIVO Profile
- Developed an extended ontology with an expanded set of roles for involvement in publication/scholarly works.
- Data is published frmo OpenVIVO every hour
    - http://openvivo.org/data/
- For research

- 44000 Journals are ingested from ISSN —> available through their data page
- Research Areas
    - Uses Faceted Application of Subject Terminology (FAST) to select research areas.
- Generated URIs for individuals using the ORCiD

Lessons Learned

- Std data for journal was challenging CrossRef and NLM provide public data on journal titles and IDs
- Std data for orgs available (GRID)
- Need improved taxonomy for organizations (VIVO's is primitive with 28 types and little hierarchy)
- Importing data from Figshare was straightforward
- The Ecosystem has issues:
    - Wide variety of name formats, hence ORCiD
    - Crossref API sometimes returns arbitrary data for a DOI lookup — need to check the DOI returned by CrossRef with the one sent to it.
        - E.g. if DOI does not exist, Crossref returns arbitrary information.
    - DOI lookup returns books with ISSNs that have nothing to do with the book??
    - Depending on who mints the DOI, the publisher information may be the DOI provider rather than the publisher
- Creating a controlled vocabulary for attribution and contribution is ongoing

VIVO Domains

1. Activities and accomplishments of your people
    a. Positions *
    b. Educational background
    c. Grants
    d. Courses
    e. Mentoring
    f. Awards & Honours
    g. Pubs, books, presentations, data, other *
    h. Service
2. The people to be included in OpenVIVO
    a. Attribution
    b. People
3. Your org's people and locations to be included in your VIVO
    a. Local organizations
    b. Local people
    c. Location locations
4. Entities to be referred to in your VIVO
    a. Journals
    b. Organizations
    c. Concepts
    d. Locations

e. Dates
f. Degrees

Identify, update, confirm

- Get this figure from slide deck
- Green = data sources that are open to use
- Blue = your institution's data
- Yellow = items you might license or pay for
    - Microsoft Academic
        - Might want to explore it
- Red = Can't use it; can't even pay to use it
    - Academia.edu, Google Scholar, Research Gate

Ingest method:

- Find the (identified) works for an (identified) person (automated, curated, or self-identify)
- Given the works, get the metadata, transform to VIVO-ISF and update VIVO
- Confirm with person (claim)

1415 - Intro to Ontology

What is linked open data?

- Structured info
- Common, simple format

What is an ontology?

- Set of representational promitives with which to model a domain
    - Attributes, types, relationships
- VIVO is ontologies, SW, data,

Ontologies

- 1 core ontology - VIVO-ISF
- 3 additional VIVO-specific ontologies
- 22 related ontologies
    - OBO, BFO, CRO (contribution ontology)

Triples

1. Subject (an entity)
2. Predicate (an object property or data property)
3. Object (any individual in VIVO)

Domains of Representation

- Activities
- People (faculty, staff, trainees, and collaborators)
- Entities referred to in your scholarly work

VIVO Ontologies (refer to this list)

VIVO Classes (433)

- Contact info
- Types of works (documents)
- Concepts

- Generic information concepts
- Processes
- Roles
- Locations
- Relationships
- People
- Organizations
- Higher-level

VIVO Object Properties between Classes

A Person

- Has types (faculty, staff, grad student, etc.)
- Has a vcard for contact information
- Has various ID - SCOPUS, ResearcherID, ORCiD, etc.
- Photo, home dept. Research areas (concepts)
- Is related to positions, education, and roles in grants, papers, courses, orgs, and other activities
- The ontologies are sufficient to represent the relationships between entities and properties of scholars and their work.

Person model (Get figure)

- Dark blue is what we're trying to model
- Light blue are dependent on the dark blue (they only exist to support the dark blue)
- Green are other entitities that exist elsewhere on their own (e.g. organization)
- E.g. foaf:Person —> vivo:Position

Organization Model (Get figure)

Journal Model (Get figure)

Publication

- The publication model

Issues:

- 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, LD4L, agriculture, cultural hertiage, etc.
    - Most ontologies "below" VIVO introduce their own 'middle concepts' — person, paper, project, — these must be 'reconciled' with VIVO-ISF

VIVO vs Vitro

- See table from slides

Vitro editing

- Basic orientation

    - Site admin menu
    - Front end vs. JSP ‘blue screen’ editing
- Adding

Vitro editing: Basic orientation (See slides)

1545 Breakout Session - Data Sources

See the group notes for full information on this part

- Our assessment of our data sources

Which data we care about:
(Hint: go to “index” on the site to see the complete collection of data represented

- All of it!
- Demographic/profile information
- Positions
- Scholarly activity - publications, books, articles, performances
- Grant information
- Areas of expertise and research overview
- Educational history / training
- Collaborators
- Geographical areas of interest
- Teaching
- Supervision (+available for supervision)
- Committee participation (internal and external)
- Awards
- Membership and affiliations
- Membership
- Identity
- **Media**
- **Profile pictures**
- Digital presence - social media / lab website

How to ensure you’ve identified all the person types (and each person in each type)?

- Information on multiple sources vs. single-source

What if you don’t have
Look at Paul’s ID mgmt “Person Types”

- Sat down with people that know the data best
- Ask them who they think is a “faculty” member
- Enterprise directory is an aggregation point
- A lot of ID mgmt person types came from work with VIVO

What is a perfect data source

Other Notes from the Day

Contacts for potential contract work to build connectors.

As I mentioned during our discussion at 2017 VIVO Camp in Albuquerque, Brian Lowe does contract programming for institutions with VIVO instances.

His email is brian@ontocale.com
His LinkedIn profile is at https://www.linkedin.com/in/brian-lowe-b7835719/

Brian was one of the first few hires made by Jon Corson-Rikert at Cornell when Jon was developing Vitro & VIVO. Brian has in-depth knowledge of the structure of VIVO and Vitro, in additional to strong ontology and programming skills. After 8 years at Cornell, around the time when external grant support for VIVO at Cornell ran its course, Brian decided to launch his own shop based in Bucarest, Romania, a hub for tech start-ups. He travels to the US for major VIVO meetings, including the annual conferences.

I worked with Brian for a few years when I was the VIVO project lead at The Scripps Research Institute and he did excellent work. We had already developed a good working relationship from 2009-2013 on the then-weekly VIVO Implementation & Development team conference calls as well as from seeing each other at conferences. When he started working on projects for Scripps (around 2013-2014), we found that email was fine for most of our communications, supplemented by occasional Skype sessions.

Feel free to mention my name if you decide to contact Brian. He would also be an excellent resource for advice on the type (professional skills) and extent (FTE) of staffing that McMaster might require to develop a VIVO instance of a given size and functionality.

Friday - Day 2

0900 - Intro to SPARQL

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

- In vivo URIs usually look like: http://yourplace/individual/nxxxx
    - E.g. Mike Conlon's URL at UF is: http://vivo.ufl.edu/individual/n25562
    -

A real triple

- Mike is a faculty members
- We have seen that Mike has the URI
    - http://vivo.ufl.edu/individual/n25562
    - We will put this in brackets to tell the computer "stuff in bracket is a URI"

Predicate

- We want to say "is a" — in Semantic Web we say "Mike ***hasType*** Faculty Member"
- The way we say it precisely is to use the rdf ontology
- VIVO uses rdf:type to say "is a"

More than one ontology

- Rdf:type
    - Ontology is rdf
    - Predicate is type
- So we say <http://vivo.ufl.edu/individual/n25562> rdf:type …...

The object

- VIVO-ISF defines a class Faculty Member
- We write
    - vivo:FacultyMember to specify that class
- So, <http://vivo.ufl.edu/individual/n25562> rdf:type vivo:FacultyMember

Two Addenda

- Rdf:type can be abbreviated "a"
    - So we can write: <http://vivo.ufl.edu/individual/n25562> a vivo:FacultyMember
- When we write triples we always end them with a period
    - <http://vivo.ufl.edu/individual/n25562> rdf:type vivo:FacultyMember .

We want to use VIVO Data

- "Use"
    - Make lists
    - Count things
    - Make reports
    - Get data out of VIVO for use in Excel, SPSS, reporting, vis software

We will write SPARQL Queries

- SPARQL is a query language for asking for data from a set of triples
- Mike says it's fun and easy!
- It is precise--you will get what you ask for.

Using UF data

- Good mid- to large-scale example of data in the “real world”

Step 1 - Get signed in

- Vivo.ufl.edu
- Click login > Login Using Gatorlink
- Login
- Click on SPARQL Query
- List of Prefixes — these are a list of shortcuts

1015 Breakout Session - Use Cases and User Stories

See this document.

How to build collect information on keywords / expertise?

- Get it from the faculty member?
    - Overview statements
    - Best approach is probably to sit with faculty members to help them improve their Google ranking — and get the research expertise through this.
- Infer it from publication titles/keywords
    - E.g. Pubmed - Title, abstract, mesh major,
- Certifications
- Procedural credentials

UF does not include all information on animal-related research, for security reasons

- Everyone in animal research are mapped to office of research (one level up)
- Everyone in police services are mapped to the higher level

Publication reporting in the Institution

- Paul’s thoughts: It’s surprising how many units across campus have an interest in compiling publication counts (dept. Admin units, external relations, etc.). Good idea to drum up call up variou
- Getting books is difficult! Affiliation isn’t always included, and it’s difficult to get this.
    - Some publishers will send a quarterly list of authors they think are yours
    - Faculty record their books in a word doc (or similar) for reporting purposes, but how to harvest this (often incomplete) data?
    - Can use students to enter books into Elements?

1115 Introduction to SPARQL, Part 2

What we’ve done so far:

- Triples (subject, predicate, object)
- SPARQL Queries
    - Select, where, group by, order by
- Thinking about data
    - Unique URIs
    - Labels

How do we know what’s in the ontologies?

1. Look at pictures (relationship diagrams can be found in the VIVO documentation wiki)

2. Read the files
3. Search the web
4. Ask friends
5. Reverse engineering

What diagram tells us:

- What articles are called (bibo:AcademicArticle)
- What journals are called and how articles and journals are connected

Example 1 output tells us:

- The publication has 9 assertions
- Has a datetime, a doi, publication venue, start and end pages, pmid, pmcid, nihmsid, title, volume
- UF adds data regarding date of harvest and harvester name, grants cited
- Has 2 subject area assertions
- Has 5 'related by' assertions
- Has obo:ARG_20000028
    - "Has contact info"
    - In this case, the article has a vcard because the article has a webpage
    - Obo library is a collection of ontologies maintained as Stanford

We can inspect entities using "reverse engineering" to inspect the

31 VIVO SPARQL Queries — NICE!

- 31 tested queries on VIVO 1.9
- Each is marked with level of difficulty
- Reports, Data Mgmt, People, Papers, Grants, Orgs
- Real-world examples. Most are the result of questions asked by stakeholders

Alphabetical list of people who have opted into VIVO

- UF has extension to tag people who have opted into VIVO.
    - Opted-in people are those that are not Faculty, but have asked to be added (e.g. CIO of the hospital)
- Query looks like the list of Faculty Members.

How to start understanding the ontology

- Inspect entities one at a time

Notes on Running Reports

- Can export RDF into another application and run the queries there
    - E.g. Stardog — triple store query engine
    - For example, can take the OpenVIVO data
- How to actually export the triples?
    - Turns out exporting the triples isn't quite so easy
    - Can't just export all subjects, predicates, objects if VIVO is too big
        - We can probably get away with this with ours right now.
    - VIVO 1.10 will have a utility to do this.
- Duke has a Tableau license, which uses that to generate reports

    - SPARQL experts to pull publications from VIVO

1330 - How to Implement VIVO

Implementation (and ongoing team) includes:

- Project management
- Outreach and community engagement
- Data Mgmt
- Technical Development

Analysis

PM

- Establish Governance
    - Get project sponsor
    - Build governance model
- Resource identification
    - Know how many people you'll need / have a sense of who you have available and what their skills are

Outreach & Comm. Engagement

- Id stakeholders
- Gather use cases (See this document)

Data Management

- Identify potential data sources
- Licensed / free / accessible / reliable / complete?
- See notes from yesterday's exercise.

Technical Development

- Learn system architecture
- Identify Customizations

Design & Implementation

PM

- Branding
- Further define scope
- Request data feeds

Outreach & Comm. Engagement

- Share prototypes and/or existing VIVOs (VIVO registry)
- We can add our instance to the registry at any time (doesn't necessarily need to be live)!
- Identify Power users (data contributors, champions, etc.)
- Develop training materials

Data Management

- Map data to ontologies
    - No blank nodes (entity without defined URIs)
    - Blank nodes could be created by some ingest processed
    - Blank nodes will come back to haunt you.

- Document data cleanup strategy
- Prepare data loads
- Document data provenance

Technical Development

- Build customized system
- Test performance

Launch

PM

- Oversee publicity campaign
    - Consider your goals
    - Legacy statements
    - Data in your VIVO
    - Support mechanisms
    - Collect testimonials and USE them!
        - Pre-write the testimonial and ask the faculty member/provost if they'd allow us to use the testimonial
    - Look at Mike's Paper: "Because we had VIVO" - in the other notes.
- Implement assessment plan

Technical Development

- Provide support and system update

Maintenance

PM

- Maintenance
- Contribute to the VIVO Community

Q: Allow information to be manually-entered into VIVO?

- There was no clear answer for this
- In many cases the solution depends on what data you need, where it exists and what shape it's in.
- It seems that most sites allow their users to enter some information (that can't be found anywhere else) and isn't HR data
    - E.g.

Q: How to load data?

- Dept by dept or load by stripes (teaching then load department, etc.--diff domains)
- Violetta & Mike: They both use the 'stripes' approach and go in this order:
    - Organizations
    - People
    - Positions (the position data makes the connection between the orgs and the people)
    - Publications

- Look at the domain diagram
-

Case Study: Weill at Cornell

Two main uses

- Place to see who the expert is at _____
- Providing a polished profile to researchers

Other notes

- Only piece of information that needs to be updated is Research Overview
- Changes flow upstream
- Few extensions to the ontology — only really use the person type
- Wants to focus on improving search functionality

Case Study: Scholars @ Duke

- Customized the home page
- Upgraded the search
    - Default search returned too much information
    - Customization of Solr search
    - Results shown in tabs
- CV-generation: use VIVO as a replacement for faculty reporting for Arts & Science
    - Had to load data (hard and dirty work)
    - Had to extend the functionality
- Delegate access
    - It's possible for a person to search the directory and establish delegates for people

How to engage faculty members to update their information

- Departments load Scholars@Duke information into their departmental sites — so, if they don't add information, that's their loss
- There will be people who just won't want to participate (or will update ORCiD, Google Scholar, etc., as an individual)
- UF: go over heads of faculty members to answer issues from dept. Chairs — when they ask
- Edit profiles for people manually, if required
- Sunshine cleaning--getting more buy-in by having more groups using the data within VIVO, or that which is upstream of VIVO.

1430 - Getting data into VIVO

Questions from the get go

- What's your team's technical skill set?
- Will the data in your VIVO be read only?
- How important is it that data will be current and reliable?
- What other systems will depend on your data?
- One graph vs. many graphs…
    - A graph is a collection of triples with a name
    - The collection of graphs are available to SPARQL
    - WC: Each source system has a graph — all linked with common identifiers

Problems

- Lack of familiarity with technology, including RDF, SPARQL
- Introduction of duplicates
- Illogical data (e.g. one object has multiple labels)

Tools (see table from slides)

- Manual Entry
- Simple VIVO
- Karma
- Harvester
- SPARQL API

The VIVO Pump (link)

A new tool for managing VIVO data using spreadsheets

- The “Pump” translates
    - from spreadsheet to RDF (update)
    - From RDF to spreadsheet (get)
- Rows represent the objects to manage
    - E.g. one row per person
    - If uri column is blank in the spreadsheet, you’re asking the Pump to add a uri
    - This works well with get and updates
- Data values in VIVO are either literal values or Object references
    - Person has name (literal)
    - Person has a research area (object reference)
    - The Pump handles both!
- When you run it
    - Definitiion file
        - Describes mapping between rows&columns and graph
        - Pump comes with 14 definition files
        - Can edit the SPARQL in the definition file to customize the actions of the ingest
        -
    - Config file
        - URL
        - username/password to API
    - Data file

1545 Engagement Breakout Session

Supporting and Marketing VIVO

Crafting a support model

- Large implementations
    - Support like other institutional systems
        - Training, helpdesk, knowledge base, website with help and about
        - Who will hear about you; how will they try and find information
    - Engage multiple layers of support
- Smaller implementations

        - At least two support people
        - Engage others in your network

Use your network

- Project team members
- Units: Business managers, dept assistants, faculty assistants,
- Web content editors, communicators
- Librarians, repository outreach and support
- Faculty assistants, delegates
- Help desk analysts

Provide local support

- Super or ‘power users’ close by to answer questions
    - Faculty /
- Designate through leadership
    - Establish that relationship with chair
    - Takes a bit of time to train; but not a substantial burden
- Train them and provide materials
    - S@D does training once or twice a month
- Communicate through the power/super users
- Show appreciation

S@D Support model

- faculty/delegates/admin staff directed to the “Power User” or “Help Desk”
    - Power user or help desk —> Faculty data project team
        - Faculty data project team —> OIT developers, Library, Data Owners

How to solve issues with poor data in your system, which you don’t have control over

- S@D: Self-serve to PeopleSoft to correct some of the personal data

Best practices for support

- Online documentation, videos
- Help ticket submission
- Ongoing training
- Events and user meetings

Messaging

- Starts with strategy and goals
    - Communication plan should include a timeline
- Big implementation
    - Use institutional resources
    - Keep message simple
- Smaller group
    - Start with simple message
    - Training events, demos
    - More complex information

You need a plan!

- Timeline, goals, tactics, events
- Communications
- Training sessions
- Give people lunch — opportunity for focus groups with Faculty
- Focus groups —

    - “which do you prefer?”
    - “What problems are you dealing with?”

Best practices for marketing

- Focus on specific benefits to each stakeholder
- Repetition is important
- ***Keep emails short*** with links to more info
    - Catchy, engaging email --graphical, catchy and have a link to more information
    - Assume that faculty members won’t read emails unless it comes from their chair
    - BUT people do seem to respond to Elements monthly reminder that they have unclaimed papers.
- Be positive and patient
- Facilitate community engagement, conversations

Scholars@Duke Support Site: http://about.scholars.duke.edu/support-duke-faculty-delegates

Scholars@Duke editing manager - “Profile Manager”

- Editable fields
    - Photos, Research Overview,
    - Awards and honours
    - Service
    - Outreach & engaged scholarship
    - Complete CV information for health sciences
        - Previous academic positions
        - Clinical activities
- Controlled vocabs?
    - For subject areas

Subject area headings

- FAST vocab from OCLC is the best
    - Hierarchical, 500000 entries
    - But still not good enough
- S@D: Let people search MESH / LOC
    - Allow people to submit ticket for missing headings and they add it to their local ontology
- Not editable - HR info; pubs; grants; teaching; news
    - All of these are auto-ingested
    - Allow users to hide any information that is auto-ingested
- DO NOT allow them to add their subject heading as free text!!!
    - Maybe limit to 3 or 4 words

Things to think about / Discussion:

- We need to ask ourselves: “Who are we serving?”
    - We’re serving faculty and students — if they think the system is useless, then it is.

Selling points:

- Democratizing data
- Allowing for data to be reused / repurposed

- Compiling data and making it accessible — you can see who taught what and who did what
- Knowledge mobilization

One issue at Duke was that they tried to do too many things at once — too many requests

- Ended up starting over and trying to make things as consistent as possible

Q: If you were doing this on a shoestring, what would you focus on:

- Insert what data you have and what has the most value
- Modify the interface to remove attention to data that you don't have
- If you're going to run a VIVO, you need someone with SPARQL knowledge

Suggested Resources:

- O'Reilly book for SPARQL
- Semantic web for the working ontologist
- Building ontologies with basic formal ontology

VIVO-related

- VIVO Ontology modules list https://mjaved495.github.io/VIVOISFOntologyModuleClient/
-

Saturday - Day 3

0900 - How the VIVO Project works

(All following documents here: https://wiki.duraspace.org/display/VIVO/VIVO)

- The VIVO Charter
  - https://goo.gl/YEOxh9
  - Describes how the project work, constitution, governance, membership processes
- VIVO Strategic Plan (https://goo.gl/y47Vtc)
  - Directions, community and governance-developed
  -
- Roadmap Process
  - People vote with their effort

VIVO Groups

- Leadership group
  - Reps at each membership levels
  - Determined annually based on membership
  - Strategy, accountability, meet quarterly
- Steering group
  - Reps of leadership orgs + at-large members
  - Elected for three-year terms
  - Tactics
  - Meets weekly
- Task Forces

    - Self-organized working groups
    - Documented on the VIVO Wiki
    - Create documentation, code, ontology processes, organize events
- Committers group
    - Makes decisions regarding code inclusion, testing, readiness for release
    - Trusted technical resources
    - Meets as needed

Roadmap Taskforce

- Four themes
    - Improve the interface
    - Improve performance
    - Improve processes for getting data into / out of VIVO

Release 1.10 (release planning page in the wiki)

- 1.10 will have a triple-pattern fragment endpoint
    - Makes it easier to serve the information out to other sites (not a SPARQL query)
    - Enables federated search
    - But potentially issues with making all information available to everyone
    - Some institutions have used flags that control what information is available to internal (logged in users) only — e.g. phone numbers

Interest group calls - every thursday by WebEx (1300 ET)

- 1st thursday - outreach and engagement (Julia)
- 2nd thursday - implementation (Paul)
- 3rd thursday - Apps and tool — Chris Barnes & Ted Lawless
- 4th thursday - Development (Graham)

1000 - VIVO Community and Involvement

VIVO Community groups to join:

- vivo-community@googlegroups.com
- vivo-tech@googlegroups.com
- vivo-all@googlegroups.com

VIVO Wikis

- VIVO Community Wiki
- VIVO Technical documentation Wiki for 1.9

1130 - Wrap-up

Q: Selling VIVO vs. other products

What are the main selling points of VIVO?

- vs. PURE
    - Commercial product of Elsevier — not a friend to the academic community — predatory pricing problems — history of not playing well.
    - Expensive and focused on their data sources and their publication

    - A closed-world system; not friendly to VIVO
        - Open source gives you control
    - Scopus database is very useful with VIVO — load into Elements into VIVO
    - The worst ability to customize from any tool
- Vs. Harvard Profiles RNS
    - Very clever data loading approach
        - Give it a spreadsheet of people and it's populated
    - Open source, but not much has been changed; not very active community
    - Typically focused on biomedicine
- Symplectic Elements
    - Corporate sponsor of VIVO
    - The new front-end for symplectic is a theme for VIVO
- Clarivate
    - Built from Converis — grant management product
    - a CRISS / Grant management perspective
    - Works well for biomedical institutions where all researchers are grant-funded
-

General notes

- Open source gives you control — don't let a vendor lock you in
- Salespeople will say things to sell you stuff
- Honest answer
    -
- Scripps hired Brian Lowe to do specific customizations for them
    - Good solution if you have short-term money
- Handling duplication
    - You have a problem — you need all publications from your faculty
    - You have sources — but how to identify your faculty members' work?
        - Heuristics
    - BUT, sources have errors
    - How Harvard Profile system works:
        - Give it a bundle of information about a person — will return a collection of PubMed IDs (only PubMed) that it thinks is your person