Results from the 2014 survey of VIVO implementation sites

Paul Albert, Alex Viggio, Jon Corson-Rikert, & Kristi Holmes

August 8, 2014
Who participated?

Section 1
Seeking clarity from the community

• How do sites solve problems?
• How important is improving performance to the community?
• Is VIVO-ISF too complicated?
• How do VIVO sites really feel about documentation?
• To what extent are we meeting the needs of the international community?
• Which code improvements are most important to the community?
What types of implementation sites participated?

- University
- Research center or institute
- Governmental organization
- Multiple descriptions
- Medical school
- Scholarly society
19 sites* participated

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Existing/Expected Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Psychological Association</td>
<td>Washington DC, USA</td>
<td>200,000</td>
</tr>
<tr>
<td>Brown University</td>
<td>Providence, Rhode Island, USA</td>
<td>3,500</td>
</tr>
<tr>
<td>Cornell University</td>
<td>Ithaca, New York, USA</td>
<td>12,273</td>
</tr>
<tr>
<td>Duke University</td>
<td>Durham, North Carolina, USA</td>
<td>8,000</td>
</tr>
<tr>
<td>Eindhoven University of Technology</td>
<td>The Netherlands</td>
<td>4,157</td>
</tr>
<tr>
<td>Fundação Getulio Vargas</td>
<td>Rio de Janeiro, Brazil</td>
<td>1,000</td>
</tr>
<tr>
<td>Laboratory for Atmospheric and Space Physics, CU Boulder</td>
<td>Boulder, Colorado, USA</td>
<td></td>
</tr>
<tr>
<td>Smithsonian Institution</td>
<td>Washington, DC, USA</td>
<td>1,000</td>
</tr>
<tr>
<td>Tetherless World Constellation at Rensselaer Polytechnic Institute</td>
<td>Troy, New York, USA</td>
<td>2,000</td>
</tr>
<tr>
<td>The Scripps Research Institute</td>
<td>La Jolla, California, USA</td>
<td>350</td>
</tr>
<tr>
<td>University of Colorado Boulder</td>
<td>Boulder, Colorado, USA</td>
<td>1,792</td>
</tr>
<tr>
<td>University of Florida</td>
<td>Gainesville, Florida, USA</td>
<td>&gt;20,000</td>
</tr>
<tr>
<td>University of Melbourne</td>
<td>Melbourne, Australia</td>
<td>7,000</td>
</tr>
<tr>
<td>US Environmental Protection Agency</td>
<td>US wide</td>
<td>1,900</td>
</tr>
<tr>
<td>Virginia Tech</td>
<td>Blacksburg, Virginia, USA</td>
<td>3,000</td>
</tr>
<tr>
<td>Weill Cornell Medical College</td>
<td>New York, New York, USA</td>
<td>5,900</td>
</tr>
</tbody>
</table>

* as well as three anonymous sites
What is your stage of development?

- In production
- In a public beta
- In beta but behind institutional firewall
- Have an institutional commitment. Now developing.
- Still deciding on software. Exploring VIVO.
What do the results mean?

See raw results here: bit.ly/vivosurvey
Personnel, hardware, ontology, and performance

Section 2
Personnel

• If in production, how many full-time equivalent (FTE)-years were needed to get VIVO ready for production?

<table>
<thead>
<tr>
<th>Average</th>
<th>Min</th>
<th>Max</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.44</td>
<td>0.5</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

• How many FTEs in the past year?

<table>
<thead>
<tr>
<th>Average</th>
<th>Min</th>
<th>Max</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.66</td>
<td>0.1</td>
<td>4.5</td>
<td>18</td>
</tr>
</tbody>
</table>

• If in production, how many FTEs are required to maintain it?

<table>
<thead>
<tr>
<th>Average</th>
<th>Min</th>
<th>Max</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.57</td>
<td>0.1</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>
Experience

• Before starting on VIVO, did any of your developers have experience with semantic web technologies?
  – Yes: 2
  – No: 16
# Role distribution

<table>
<thead>
<tr>
<th>Role (expressed as FTE)</th>
<th>Avg</th>
<th>Min</th>
<th>Max</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmer/ETL specialist</td>
<td>0.52</td>
<td>0</td>
<td>2.1 – 3.0</td>
<td>Writing scripts for ingest and update</td>
</tr>
<tr>
<td>Project manager or coordinator</td>
<td>0.30</td>
<td>0</td>
<td>1.1 – 2.0</td>
<td></td>
</tr>
<tr>
<td>IT/systems engineer</td>
<td>0.19</td>
<td>0</td>
<td>0.6 – 1.0</td>
<td>Installation, configuration, tuning</td>
</tr>
<tr>
<td>Content coordinator</td>
<td>0.16</td>
<td>0</td>
<td>0.6 – 1.0</td>
<td></td>
</tr>
<tr>
<td>Ontologist</td>
<td>0.15</td>
<td>0</td>
<td>0.6 – 1.0</td>
<td>Local ontology extensions</td>
</tr>
<tr>
<td>Project executive or champion</td>
<td>0.13</td>
<td>0</td>
<td>0.1 – 0.5</td>
<td></td>
</tr>
<tr>
<td>Web developer</td>
<td>0.13</td>
<td>0</td>
<td>0.6 – 1.0</td>
<td>Branding and customization</td>
</tr>
<tr>
<td>Marketing/outreach person</td>
<td>0.12</td>
<td>0</td>
<td>0.6 – 1.0</td>
<td></td>
</tr>
<tr>
<td>Database administrator</td>
<td>0.09</td>
<td>0</td>
<td>0.1 – 0.5</td>
<td></td>
</tr>
<tr>
<td>Business analyst</td>
<td>0.09</td>
<td>0</td>
<td>0.1 – 0.5</td>
<td></td>
</tr>
<tr>
<td>Manual entry clerk</td>
<td>0.07</td>
<td>0</td>
<td>0.1 – 0.5</td>
<td></td>
</tr>
</tbody>
</table>
VIVO version

1.4.x

1.5.x

1.6.x

1.7.x
Hardware

- virtual server(s)
- dedicated physical server(s)

Memory

- don't know
- 4 GB
- 7-10 GB
- 16 GB
- 32 GB
Environment

• MySQL database (1 AllegroGraph)
• Tomcat (1 Glassfish)
• Apache web server
  – 2 sites using Apache mod_cache, 1 Squid
• 1 site using Amazon EWS
• 1 site doing load balancing
• 2 use other technologies (Ruby, .Net) in their stack to support editing
Stakeholder satisfaction: performance and uptime/reliability

- no answer
- very unhappy
- somewhat unhappy
- moderately happy
- mostly happy
- extremely happy

0 2 4 6 8 10 12

- page load speed
- uptime and reliability
SPARQL endpoints

- public
- non-public
- no endpoint
- no response

0 2 4 6 8 10
How is your SPARQL endpoint used?

- custom visualizations
- Deep Carbon Science Community
- repurposing data for institutional sites, i.e. displaying skills and datasets
- quality assurance
- offload queries from main VIVO app
- internal applications/business intelligence
- impressing Chris Barnes
- no answer
If not, why not?

- N/A or no response
- haven't gotten around to it yet/not enough data
- worried about performance
- concerned about sharing institutional data/security
Desired features

- Data integrity tools
- Multi-institutional search
- HTML embeds of SPARQL queries
- Bidirectional data exchange with ORCID
- VIVO Drupal dashboard
- Export to CV/biosketch

Legend:
- Highly
- Moderately
- Neutral
- Not at all
- No response

#VIVO14
Ontology

• Have you included other established ontologies?

- No
- Yes

- No
- Yes

- No
- Yes

• Have you made any local customizations or extensions? (range: 5-57 classes, 5-100+ properties)
Rating the VIVO-ISF 1.6 ontology

- Uncertain - we have yet to try it
- More detail than we need but we can ignore what we don't use
- About right
- Lacking in areas we need
Desired areas for ontology extensions and/or inclusion of other existing ontologies

- gene location
- global interests
- patents
- research projects
- track of results/evaluation
- additional publication types
- placeholders for local identifiers
- humanities/artistic works
- data catalog/datasets
- provenance
External vocabularies linked to VIVO

APA ontology
VSTO (Virtual Solar-Terrestrial Observatory), DCAT (data catalog)
SPIN research interests
Library of Congress Subject Headings
UMLS (includes MeSH)
all provided with VIVO (UMLS, LCSH, Gemet)
no answer/none used
What content do sites care about?

Section 3
Where can end users manually add data?

- Preferred name
- Preferred title
- Positions
- Educational background
- Journal articles
- Books and book chapters
- Grants, agreements, and contracts
- Honors and awards
- Published datasets or databases
- Patents
- Certifications
- Professional service
- Research projects
- Courses taught
- Artistic works or performances
- Events
- Consulting or professional practice
- Community service or outreach
- Advisees
- Labs, facilities and equipment

Options: Upstream source, Directly within VIVO
Which data are currently in your VIVO and what’s the source type?

- Internal
- External
- Both

- Internal people
- Positions
- Preferred name
- Preferred title
- Journal articles
- Educational background
- Departments
- Books and book chapters
- Courses taught
- Grants, agreements, and contracts
- Sub-units or programs
- Labs, facilities and equipment
- Research centers
- Honors and awards
- Artistic works or performances
- Professional service
- Published datasets or databases
- Patents
- Certifications
- Research projects

0 5 10 15 20
Which sources do VIVO sites use for publications and for what purpose?

- **Internal authoritative source**
- **Personal CVs/biosketches**
- **Web of Science / Web of Knowledge**
- **Symplectic Elements**
- **ORCiD**
- **PubMed/MEDLINE**
- **CrossRef**
- **Thomson Reuters' ResearcherID**
- **Scopus**
- **Web-based profiles such as lab pages**
- **Google Scholar**
- **SciVal Experts**
- **Scopus Author ID**
- **WorldCat**
- **Microsoft Academic**
- **Mendeley**
- **VIAF.org**

**As a source for ingest**

**To disambiguate authors**

**Both purposes**
Institutions’ unfulfilled desires for data
Institutions’ unfulfilled desires for data

- Community service or outreach
- Events
- Advisees
- Patents
- Certifications
- Published datasets or databases
- Third party research resources
- Labs, facilities and equipment
- Research services
- Professional service
- Consulting or professional practice
- Research projects
- Media coverage
- Books and book chapters
- Honors and awards
- Grants, agreements, and contracts
- Institutional news releases

No, but we're working on it
No, but we're interested
What data do institutions have and want, combined?

<table>
<thead>
<tr>
<th>Category</th>
<th>Have it</th>
<th>Want it</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal people</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred name</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal articles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Books and book chapters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labs, facilities and equipment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Departments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grants, agreements, and contracts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational background</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courses taught</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published datasets or databases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-units or programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honors and awards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community service or outreach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advisees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research projects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research centers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consulting or professional practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artistic works or performances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institutional news releases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other research resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact statements or extension success stories</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Have it and Want it combined?
State of the VIVO community

Section 4
How was developer time spent during the first year of VIVO implementation?

Learning Semantic Web
37%

Implementation Tasks
48%
Implementation Challenges (cont’d)

Political: getting access to data from institutional sources?  Not at all / To a limited extent

Addressing poor performance of the application?  Not at all / To a limited extent

Getting adequate resources from administrative leadership to proceed?  To a limited extent
Implementation Challenges (cont’d)

Theming and other changes to the templates?
To a large extent

Understanding VIVO's semantic data model?
To a large extent / Extremely

Cleaning up dirty data from institutional sources?
Extremely

Ingesting data using Harvester or other similar tools?
Extremely
How effective have each of the following methods been for solving implementation roadblocks?

- Trial and Error? **Highly**
- Email to the mailing lists? **Moderately**
- Searching mailing list archives? **Moderately**
- Contacting developers off-list? **Moderately**
- Implementation/development call? **Moderately**
- Using the wiki? **Moderately**
- Posting to Stack Overflow? **Not at all**
Q: Should VIVO docs describe a tool, an application, or a platform?
Documentation Priorities

Self-guided tutorials for data ingest?  Very important

Standard data set to work with?  Very important

White paper highlighting use cases for how VIVO can benefit an institution?  Very important / Important

Detailed list of standard application and server configurations?  Important
Internationalization

- Yes: 53%
- Working on it: 16%
- Considering it: 21%
- Not Sure: 5%
- No: 5%

Recommendations:
- Attend international conferences
- Language-specific docs

Languages: Spanish, German, French, Chinese, Japanese, Italian
Contributing to VIVO

90% of responders have contributed to open source projects

>50% using GitHub
Contributing to VIVO

- Contributed to VIVO listserv traffic
- Participated in Implementation Fest(s)
- Contributed to conference(s)
- Contributed to demos or webinars
- Participated in hackathon(s)
- Shared scripts/code for VIVO-related apps and tools
- Contributed to wiki
- Submitted or commented in JIRA
- Made contributions to VIVO core
- Shared education, support, or outreach materials
Why aren’t we contributing to VIVO code base?

- Competing projects
- Failure to understand how the application works
- Don't know where to start
- Decision makers don't see the value
- Lack of perceived permission or encouragement from current developers/contributors
- Don't have any particular needs that could be addressed
- Other
Outreach / Competing visions for VIVO

Section 4
How frequently do you receive the following feedback about your VIVO?

- My data is wrong
- Key data of mine is missing
- I don't want VIVO to replace my institutional website
- I'm not allowed to edit certain data from external source
- I don’t want my data public
- I wish I could easily repurpose my VIVO data (e.g. populate a LinkedIn or ORCID profile, or lab site)
For your institution, what is the importance of the following visions of VIVO?

1. A vehicle for building intra-institutional networks of people
2. A vehicle for building inter-institutional networks of people
3. VIVO as a tool for awareness of current activities
4. A loose federation of many lightweight, extensible, and creative apps on a common semantic framework
5. A flagship product that needs to be easier to adopt, populate, and grow
6. A discovery front end for repositories of documents & data
7. VIVO as a tool for archiving scholarly activities over time
8. A vehicle for building networks of research data
9. A tool to help universities and government agencies meet open government mandates
10. A home and integration point for a virtual organization
11. A reference implementation for the VIVO-ISF ontology as a standard for international data exchange
A vehicle for building intra-institutional networks of people

- The metaphor of VIVO as vehicle may be attractive because:
  - Vehicles are generally simple to use.
  - Vehicles do something useful rather than serve its own end.
- VIVO is not an archive or repository.
- VIVO’s most popular purpose – current awareness about internal researchers’ activities.
- Intra-institutional network > Inter-institutional network
- Researchers > Researcher Data
For your institution, what is the importance of the following visions of VIVO? (internal)

- A vehicle for building intra-institutional networks of people
- A vehicle for building inter-institutional networks of people
- A discovery front end for repositories of documents & data
- VIVO as a tool for awareness of current activities
- VIVO as a tool for archiving scholarly activities over time
- A home and integration point for a virtual organization

Indispensable | Valuable | Moderately useful | Not at all
---|---|---|---
A vehicle for building intra-institutional networks of people
A vehicle for building inter-institutional networks of people
A discovery front end for repositories of documents & data
VIVO as a tool for awareness of current activities
VIVO as a tool for archiving scholarly activities over time
A home and integration point for a virtual organization
For your institution, what is the importance of the following visions of VIVO? (external)

- A flagship product that needs to be easier to adopt, populate, and grow
- A vehicle for building networks of research data
- A loose federation of many lightweight, extensible, and creative apps on a common semantic framework
- A tool to help universities and government agencies meet mandates for open government
- A reference implementation for the VIVO-ISF ontology as a standard for international data exchange

For your institution, what is the importance of the following visions of VIVO? (external)
Themes in the answers to the question, “What is your site’s priority?”

- Tools for read/write access to back end
- Symplectic Elements implementation
- Other 3rd party integrations (ORCID, Fedora)
- Sharing data across VIVOes
- Tools and apps
- Project viability
- Search across VIVOes
- Visualizations
- Marketing
- Adoption
- 1.6 upgrade
Would the following messages be successful in marketing VIVO to your profiled users?

• VIVO promotes your work within your discipline
• VIVO promotes your work to granting agencies
• VIVO promotes your work to prospective students and post-docs
• VIVO saves staff and researcher time
• VIVO meets key administrative needs such as business intelligence
• VIVO helps researchers find collaborators
• VIVO promotes your institution's work and makes you more competitive
Thank you to everyone who took time to share your site’s perspective.