

# VIVO

connect • share • discover

## VIVO Technical Roadmap Discussion

Jon Corson-Rikert

August 6, 2014

# Goals for discussion

- Articulating strategy and high-level priorities
- Identifying opportunities and needs
- Growing the contributor base
  - Encompassing much more than the core Java code
- Crafting a participatory process for developing a roadmap with timetable and milestones tied to releases
- Accommodating changing technologies and personnel

# What goes into the roadmap?

- Goals – where (all) are we going with VIVO?
- Opportunities – e.g., SciENCv exchange
- Strategy – how do we get there?
- Priorities – what do we do first?
- Players – who will be involved?
  - How can we leverage other efforts?
- Resources – how will it happen?

# Roadmap process

- Goals
  - Transparency
  - Diversity of input
  - Greater voice for sponsors
- Methodology
  - Writing user stories and estimating resource requirements
  - Reviewing dependencies
  - Defining appropriate chunks of work
  - Voting
  - Prioritizing and staging
  - Committing funds and/or FTE

# Scope

- A VIVO roadmap needs to address communications, documentation, training, recruitment and support for new committers, collaboration tools, changing technologies and standards, and many other factors beyond new features
- We still fall short on many of the above, most urgently on **documentation** and **training**
- A clear roadmap is only one criterion among several for moving beyond incubator status with DuraSpace
- Other criteria include moving from a BSD to Apache 2 license to clarify the status of contributed code

# Alternative visions of VIVO

- A flagship product that needs to be easier to adopt, populate, and grow
- A vehicle for building networks of research data
- A reference implementation for the VIVO-ISF ontology as a standard for international data exchange
- A loose federation of many lightweight, creative apps
- A home and integration point for a virtual organization
- A tool to help universities and government agencies meet mandates for open government
- A discovery front end for repositories of documents & data
- These all exist and are not mutually exclusive!

# Caveats

- We are not yet working on most of the following ideas
- There will be many more good ideas we haven't thought of (or better ones)
- Any project has technical debt – parts of the code that need refactoring or rewriting to enable new development but don't themselves deliver new features
- We don't currently have the resources required to work on more than a very few
- There are many ways to participate and contribute
- We need strategy as well as tactics

# Architecture | Installation

- Modularity
  - Reorganize access to data models
  - Continued improvements to model-view-controller separation
- Distribution & installation
  - MySQL-free distribution (e.g., Jena TDB for 1.8)
  - Binary distribution to simplify deployment
  - Self-contained “instant” VIVO for pilot projects
  - Interactive configuration with embedded smoke tests



# Performance

- Evaluate other triplestores using existing RDF API
  - Jena TDB, Virtuoso, AllegroGraph, others
- Redesign & reimplement simple reasoner to speed re-inferencing
- Restructure grouped property list
  - 60% of time required for profile page generation

# Search within VIVO

- Unit tests with body of sample data to serve as verifiable baseline
- Support for additional facets
  - Contextual facets by type
- Configuration and tuning of relevance ranking
- Decoupling Solr to allow alternative search engine(s) like Elasticsearch or Funnelback (in 1.7)

# Search engine optimization

- Dynamic sitemaps
- Better titles on pages
- Snippets with metatags
- Schema.org tags embedded in HTML to expose more of VIVO's internal semantic structure

# Internationalization

- Further understanding our international users' needs
  - Occasional need, as for book titles in other languages
  - Interface and all content in a language other than English
  - Full bi-lingual or tri-lingual interface & content
- UI design to support multiple languages during content creation & editing
- Support for multiple languages in ontology editor and other admin functions
- Downloadable language bundles
  - Spanish in final preparation for use in Mexico, Costa Rica, Peru, and Spain
  - Active interest in Mandarin, German, Dutch, French

# Integrity maintenance

- Better support for interactive and batch deletion
  - Removing all strictly dependent RDF
  - Leaving related authors, organizations, journals
- Site management data integrity tools
  - Scanning for orphaned data
  - Dead links
  - Being implemented as python/SPARQL tools at the University of Florida

# Archival VIVO

- Maintaining historical record of former researchers
- Could provide authoritative information for external links
- How to separate out from current information in display, analysis, search
- Significant interface issues
- Should we run separate archival VIVO?

# External URIs

- Extend VIVO support to new controlled vocabularies
  - E.g., Getty vocabularies being published as linked data
- ORCID integration through A&I grant
- Linking to people/organizations in another VIVO or via multi-site search index
- Linking to external organizational identifiers
- Coordination with library authority efforts, VIAF, & OCLC – in part through Linked Data for Libraries grant
- Modifications to better support persistence of VIVO identifiers and data

# Ontology

- Tool to produce VIVO-ISF modules from broader ISF repository
  - Being implemented for eagle-i now
- Coordinated governance, evolution, and extensibility reaching well beyond the current VIVO community
- VIVO application ontology
  - Making 1.6 changes configurable via admin interface
  - Keeping ontology simpler while improving display & editing interfaces
- Ontology editor improvements
  - Unions and intersections for property domains and ranges



# Ontology extensions

- Datasets and their relationships to publications, grants, projects, and contributors
  - Including information on downstream usage
- More detail on grants, contracts, & projects
- Impact, altmetrics, and usage
- “Knowledge Mobilization” & open government
- Facilities and equipment
- Library resources

# Multi-site search

- Finish re-implementing linked data harvester from 2011 prototype
  - Harvesting in parallel, with interrupt/resume
  - Problem reports on harvested data
  - Mapping other ontologies to VIVO on harvest
- Refresh & extend front end of vivosearch.org
- Develop business model for hosting and participation
- Create and market disambiguation/resolution services using harvested data

# VIVO tools

- VIVO Vagrant virtual machine, test datasets, and other lightweight, fast ways to get started
- SPARQL endpoint documentation & training (e.g., database replication)
- Ingest tools filling gaps and incrementally improving the Harvester framework, Karma, UF VIVO tools, and other local solutions (Python, Ruby) well worth sharing
- New and more lightweight visualizations
- Analysis and reporting

# VIVO apps

- Work in progress by Mike Conlon at UF
  - CV & biosketch generation
  - Visualizing organizational structure
- Duke widgets for consuming VIVO data on other campus websites
  - Copy and paste, like embedding YouTube
- VIVO Drupal dashboard at Weill Cornell
- Core facility search ([cores.ucsf.edu](http://cores.ucsf.edu)) built on ISF ontology using the eagle-i modules
- Other reporting opportunities

# VIVO linkages

- Feeding VIVO from Elements
  - Symplectic open source connector
  - Implementations by Cornell and other Elements users
- VIVO to CKAN data repository (RPI)
- VIVO to Drupal two-way interaction
  - RPI, Memorial University, Weill Cornell
- Karma semantic ingest tool (USC)
- Opus Project at UCLA (using VIVO within a comprehensive faculty information system)

# What's missing?

- Do you have other needs or expectations?
- Do you want to connect VIVO to a larger ecosystem of data and/or tools?
- How do we help people past barriers and make developing clean institutional data easier?
- How important is networking multiple VIVO's to you?

# Ways to participate

- VIVO Hackathon October 13-15
- Regular working group calls
- Spring VIVO Implementation Fest
- Annual VIVO conference
- Meetups at other events like CNI, Force 11
- Advocacy for open standards and persistent identifiers
- Related DuraSpace and Hydra initiatives