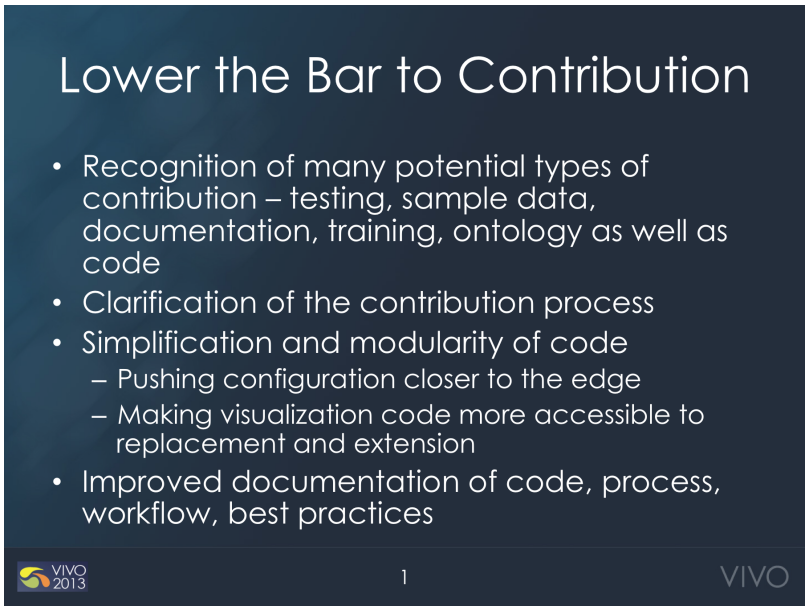
VIVO Roadmap Goals and Process

The following presentation was given on the January 30, 2015 VIVO implementation and development call to help stimulate and scope a roadmap process for VIVO.

Slide	Notes
	To be successful, a roadmap process should lower the bar to contribution, not only in the process but for any aspect of the VIVO community. The VIVO community is a big tent welcoming participation by anyone interested in the VIVO software, the VIVO-ISF ontology, or more generally in the building networks of research and researchers in any domain.
	A roadmap process must start broad enough to encompass many activities and goals, and must be closely aligned with overall project and community strategy and goals.

Roadmap process

- Writing user stories and estimating resource requirements
- Reviewing dependencies
- Defining appropriate chunks of work
- Voting
- Prioritizing and staging
- Committing funds and/or FTE

Here are some of the tasks that developing a roadmap will likely include

Likely implementation goals

- Streamlining each stage of starting a VIVO and ramping up to full scale production
  - E.g., reviving a virtual machine via Vagrant configuration and/or single-board computers
- Developing more robust ingest patterns
  - Making better use of common source data such as NIH RePORTER and PubMed
- Stronger liaisons with other open source communities to learn best practices and generate new ideas and approaches

A roadmap should respond to feedback from the field in order to address pain points in practice, not just new features.

Likely development goals

- Performance
- Modularity
- Configuration
- Search – SEO, relevance, local configuration
- Interoperability
  - Expanded terminology and authority services
  - Data flow from and to other tools – e.g., for visualization
- Responsive design & true mobile UI
- Provenance as in granular data review, error reporting, and logging/history

Each of the main headings here is broken out in more detail that follows.

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

See also the VIVO/Vitro Architecture page and Software Architecture Overview.

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

Performance and scalability are frequently recurring themes as VIVO projects assemble more data and support additional web traffic.

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)

VIVO's search indexer has improved performance and additional configuration options for VIVO v1.8.

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

The UCSF Profiles team has put major efforts into SEO and significantly increased traffic to their site. VIVO can benefit from well-understood, fairly simple and straightforward improvements.

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

VIVO internationalization helps build adoption outside the U.S. and can even support subtle regional differences in a single language.

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

Data cleanup is a headache at almost all VIVO sites – and can be helped both by improvements to the VIVO software and the development/adoptions of other apps & tools.

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?

The library and linked data worlds are taking steps to assure the stability of identifiers for researchers, organizations, places, journals, subject areas, events, and other common shared types. Many VIVO instances focus on current awareness and remove researchers who have left the organization. What should our approach as a community be?

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

These issues are closely related to each other.

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
  - Keeping ontology simpler while improving display & editing interfaces
- Ontology editor improvements
  - Unions and intersections for property domains and ranges

The VIVO-ISF ontology has a broader constituency by virtue of supporting the eagle-i application and Plumage as well as VIVO – and many concerns and common interests closely related to the VIVO application, versioning, and the process of making new releases of the software.

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

Working together on ontology extensions will help to prevent diverging ontologies and data.

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

The vivosearch.org prototype was developed in 2011 but has never evolved into the anticipated production-level service. How do we move this forward, both to connect all VIVOs but to address the needs of consortia such as the Southeastern Universities Research Association, the Mountain West Consortium, the Knowledge Mobilization community of Canadian universities, or 30 agricultural universities in China?

Key Business Questions for Search

- Costs are reasonable but real, and have natural triggers like size & harvest frequency
- Operational challenges bringing new institutions and new platforms on board
- Versioning issues with ISF and changes going forward
- Other players in the market
- Great potential for building additional value
  - Lookup/disambiguation services on top of the search index
  - RDF mashups and distributed SPARQL queries

A coordinated VIVO multi-site search effort faces many challenges but offers a key missing piece to connect individual VIVO efforts and connect researchers across software platforms.

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!

Finally, we all need to understand the diversity of the VIVO community – now and in the future.