

# *SciENCv - Putting the Pieces Together*

## VIVO

Jon Corson-Rikert

August 27, 2012

# What is VIVO?

- An open community with strong national and international participation
  - Focusing primarily on research information discovery and exchange
- An ontology with a community-driven process for extension
  - Grounded in partnerships with CASRAI and euroCRIS
- An open suite of software tools
  - VIVO
  - VIVO Harvester
  - Datastar research data registry
  - Tools for using semantic data in common web platforms

# VIVO community activity



# VIVO as an interchange format

- Resource Description Format (RDF) data defined in terms of an Web Ontology Language (OWL) ontology
  - Simple structure and defined meaning
  - Directly interoperable through standard HTTP protocol
  - Data retrievable as linked open data from original namespace
- Embedded identifiers
- RDF and the VIVO ontology adopted as a recommendation of the 60 NIH Clinical and Translational Science Awards (October, 2011)

# A shared ontology enables sharing data

- Choice of software
  - Harvard Profiles
  - Iowa's Loki
  - Elsevier's SciVal Experts
  - Digital Vita and Stanford CAP reputedly in progress
- Common indexing for relevance-ranked search
  - Prototype at <http://vivosearch.org>
- Independent development of applications on top of the data
  - Northwestern University recommender system

# Goals for VIVO in the SciENCv context

- We hope to make VIVO an effective vehicle for institutions to provide research identity, employment, and activity information
- Minimize human data entry, but essential to provide a way for researchers to review
- Institutional provenance of data is important
  - We are working with ORCID to ensure VIVO can submit affiliated researcher information in batch for claiming
  - Affiliations will change over time
  - Being able to link back to institution linked data will help with disambiguation

# Important principles

- Researcher engagement and control
  - More than just authors – contributors to all kinds of data and other research outputs
- Institutional control
  - For sustainability, SciENCv should be an extension of university functions meeting local needs as well
- Promotion of common identifiers
  - Not just on people – orgs, journals, events
  - Allows linking data from multiple sources
  - Where possible, reference the same URIs or UUIDs to minimized sameAs assertions and mapping

# It's all about the data

- Model data at a level that permits re-use
  - Articles, journals, and people, not formatted citations
- Openness
  - Clear separation of public and private
  - Make appropriate data available back to data providers, to peers, and to 3rd parties, not just government
- Demonstrated sharing
  - Learn where misalignments and mis-codings occur
  - Move the horizon beyond a single institution or initiative
- Think globally
  - Standards that only work in the U.S. are not standards
  - Does not mean there can't be U.S. extensions



# Cautions

- Building enterprise scale systems of any kind is hard
- Researchers, administrators, and IT staff are all overburdened
- Nothing is fully automatic
- Using data beyond its original context will bring surprises
- Hidden data probably has problems
  - Making it visible helps motivate corrections
- Set realistic short-term milestones that contribute to more ambitious long-term goals

# Going further with standards

- VIVO is working with CASRAI to build out the CASRAI dictionary
  - Additional data elements
  - More focus on relationships
  - Extending to encompass research resources
  - Will reuse existing ontologies where possible
- This is a community process in the U.S. as well as internationally
  - Strong engagement from the research resource and clinical communities through the CTSAconnect project (<http://ctsaconnect.org>)

# VIVO sustainability

- VIVO and DuraSpace have announced a partnership to create a VIVO incubator project
  - Provides a sustainability path beyond NIH funding
- DuraSpace is a good fit
  - Established 501(c)(3)
  - Community driven organization
  - Supports the leading repository platforms DSpace and Fedora
  - Will allow VIVO independent direction and branding while mentoring on community building, communication, and open source software development
  - Very strong response from the community already

# For more information ...

- Ask me here
- Email me ([jc55@cornell.edu](mailto:jc55@cornell.edu))
- Visit <http://vivoweb.org>
- <http://www.vivoweb.org/conference2012-materials> for workshop materials and presentations from last week's conference

Thank you