AgriVIVO
A Global Ontology-Driven RDF Store Based on a Distributed Architecture

Valeria Pesce*, John Fereira^, Jon Corson-Rikert^, Johannes Keizer~

*Global Forum on Agricultural Research
^Cornell University
~Food and Agriculture Organization of the UN
Contents

• What we wanted to do
• Why we chose VIVO*
• How we adapted VIVO and built AgriVIVO
  – Ontology
  – Importers
  – Search interface
• Future plans

* VIVO is a research discovery tool based on semantic technologies initially developed at Cornell University and now an incubator project under DuraSpace.org
What we wanted to do
What is “we”

The Global Forum on Agricultural Research (GFAR)

“Agricultural Knowledge for All” program:
a set of activities to improve information and communications management in agricultural research for development (ARD)

Cornell University
Initiator of VIVO

Food and Agriculture Organization (FAO)
of the UN
In particular, the Agricultural Information Management Standards team
The scope of GFAR’s data projects

- Data source scope: global, cross-disciplinary (within ARD)
- Use and application: local, regional, global, thematic
What we wanted to achieve

• More effective collaborative research and networking across countries and regions
• Facilitating capacity strengthening and networking of skills

• Fostering collaboration and synergy through greater awareness of ongoing research
• Reducing duplication of research
• Determining strategic trends based on strengths and weaknesses of the network
• Identifying missing expertise
Whom we wanted to support

We wanted to help researchers, research managers, practitioners as well as decision makers to identify / discover:

- their potential best collaborators all over the world for a project
- a person with an answer to their question
- an organization running a project on a specific area of research
- an organization funding projects in a specific area of research
- all the publications written by a potential collaborator
- numbers or geographic distribution of available competencies or ongoing projects
How?

The **connections** between you and your potential collaborators can take many forms. They usually follow the well-understood **patterns of affiliation**, **publication**, **participation** and **funding**.

Jon Corson-Rikert, VIVO team

---

We wanted to give access to:

- Profiles of experts
- Profiles of organizations
- Research outputs
- Projects
- Grants
- Events
- ... worldwide ... geographically
CRIS models cover such aspects

CERIF main classes
Model for “Current Research Information Systems” (CRIS)

VIVO main classes
VIVO is defined as a “research discovery tool”
What is a CRIS

A Current Research Information System

- Normally, managed at an institutional level
- Normally, managed in research institutions: universities, research centers
- Some data entered manually, some imported from other institutional databases, some aggregated from external sources

How? Some CRIS tools

- Pure (Atira > SciVal)
  http://info.scival.com/pure

- Converis (Avedas)
  http://www.converis5.com/

- Symplectic Elements (Symplectic)
  http://www.symplectic.co.uk/product-tour/

- VIVO (now a DuraSpace Incubator)
  https://wiki.duraspace.org/display/VIVO/
Why we chose VIVO
Our special requirements

• Data already collected in institutional, national or thematic databases / platforms
• Principle: data have to be entered once, as close to source as possible, and reused
  → No data entry in the global system
  → Aggregation from relevant data sources
  → Distributed architecture

• Global, cross-institutional, expertise-based
  → The model needs to be less tied with institutional structures (university, research institute)
  → Need to adjust the CRIS model to our needs

• Semantic technologies, Linked Data!
• Open source
What is VIVO

• VIVO is an open-source semantic publishing platform for making **data about research activities** visible and accessible.
  — *based on semantic technologies initially developed at Cornell University and now an incubator project under [DuraSpace.org](http://DuraSpace.org)*

• Organization of data is based on a **bundle of ontologies** and data are stored in a **triple store**.

• When installed and populated with researcher interests, activities, and accomplishments, it enables the **discovery of research** across disciplines at that institution and beyond.
Why VIVO 1: Distributed aspects

• Besides its CRIS model, VIVO can enable the discovery of researchers across institutions

• See VIVOweb (http://www.vivoweb.org/):
  – Participants in the network include institutions with
    • local installations of VIVO
    • other profiling applications
  – *The information accessible through VIVO's search and browse capability will reside and be controlled locally*

• See VIVOSearch (http://beta.vivosearch.org/):
  – A demonstration of multi-institutional search over several VIVO installations
Why VIVO 1: Distributed aspects

Linked Open Data
Distributed architecture: how

• **Aggregated Solr index**
  – If data providers are able to produce custom indexes based on similar metadata models

• **Harvesters**
  – Allow to parse different types of sources, map their elements to VIVO metadata and ingest them

• In our project, foreseen data providers manage data with very basic tools and provide them in very basic formats

→ **We chose the harvesters approach**
Why VIVO 2: Adaptable model

VIVO has an extensible ontology

- You can extend the ontology without modifying the tool
  - Tradeoffs of generality vs. optimal interface*

- The VIVO model can be **customized to fit agricultural research** e.g. by
  - extending it to **include non-academic actors that are relevant to the agricultural domain** (revising the Organization and Person sub-classes)
  - integrating properties for **annotation with external concepts** from Agrovoc**

* From the VIVOweb presentation by McIntosh, Cramer, Corson-Rikert: “VIVO Researcher Networking Update”, 2011
** Widely used agricultural thesaurus: [http://aims.fao.org/standards/agrovoc/about](http://aims.fao.org/standards/agrovoc/about)
Why VIVO 3: standards

• Uses and links to standard vocabularies
• Uses RDF
• Exposes Linked Data
• Is being mapped to other standards (CERIF)
• Has been connected to SPARQL endpoints and Linked Data APIs
• Is open source
• Is widely used and supported
How we adapted VIVO and built AgriVIVO
What is AgriVIVO

• AgriVIVO is an RDF-based and ontology-driven global aggregated database harvesting from distributed directories of experts, organizations and events in the field of agriculture.

• AgriVIVO is also a search portal giving access to the AgriVIVO database

• AgriVIVO will broaden its scope to cover the relationships between people, institutions, projects, publications and datasets
AgriVIVO data flow

- e-Agriculture
  - People
  - Institutions
- CIARD / RING
  - Institutions
- IAALD
  - People
  - Institutions
  - Events
- EGFAR
  - People
  - Institutions
- AIMS
  - People
  - Institutions
- AgriFeeds
  - Events

AgriVIVO importers / mappers

Map to (Agri)VIVO RDF classes and properties

CMS for manual submission and curation

AgriVIVO discovery portal

Search engine using VIVO RDF through SPARQL and API

New sources:
- YPARD
- CABI
- SIDALC
The search portal

AgriVIVO is a search portal built to facilitate connections between all actors in the agricultural field, bridging across separately hosted directories and online communities. This is a prototype.

You can search for people, organizations and events. Read more on how to have data included in AgriVIVO. Read our new FAQ and our terms of use.

Last import: 12/07/2013 - Next import: beginning of August 2013

10 DATA PROVIDERS
526 PEOPLE
4,474 ORGANIZATIONS
150 EVENTS

AgriVIVO Tools

Searchlight
AgriVIVO Searchlight is a small application that shows profiles related to the webpage you are reading.

Read more +

VIVO
The VIVO software and ontology are the foundation upon which AgriVIVO is built.

Read more +

More tools to come...
AgriVIVO is designed to make data open and accessible, opening new possibilities to leverage content.

Read more +

Current search
- information management (remove)

Narrow your results
- Enter keywords...
Search

Filter by expertise:
- CIARD - Coherence in Information For Agricultural Research For Development
- Interoperability
- Knowledge Management
- Linked Open Data - LOD

Filter by location:
- Italy
- The United States Of America
- France
How we adapted VIVO and built AgriVIVO

1. Extension of the ontology
VIVO basic entities and relations
The whole ontology – just an overview
Extension of the ontology

Examples of needed extensions
Extension of the ontology

Examples of needed extensions

Organization

Person and education

Sub-sub-class

Agricultural research Institute

Sub-sub-class

Agricultural research center

Organization

Academy

NGO

Farmers Organization

International Organization

Extension Unit

Funding Organization

Government Agency

Hospital

Institute

Library

Museum

Private Company

Program

Publisher

Research Organization

School

Student Organization

University

Position

[Positions] Revise?

Agricultural researcher

Farmer

Extension / communication agent

Policy maker

Senior Officer

Administrative staff

Information manager

Agriculture researcher

Farmer

Extension / communication agent

Policy maker

Senior Officer

Administrative staff

Information manager

Examples of needed extensions
Extension of the ontology: where?

• VIVO ontology editor?
  → Issues of future compatibility with new versions of the VIVO ontology

• Ontology extension published independently?
  • If published independently, “domain-specific” or “scope-specific” ontology extensions (e.g. for libraries) can be re-used by VIVO instances with the same needs
  • Extensions that are general enough could be considered for inclusion in the core or as a general-use extension package
  • Extensions can be imported into the VIVO instance

→ We created an ontology extension called “agrivivo” and published it
Extension of the ontology so far

http://www.agrivivo.net/ontology

We used an RDF vocabulary editing tool called Neologism (a Drupal distribution)
Besides extending the ontology with necessary new classes, we decided not to use some of the existing VIVO classes.

This is sort of an “Application Profile” with selected VIVO classes and AgriVIVO classes that are suitable for the domain of agriculture.
Extension of the ontology

- Adding AGROVOC as domain-specific reference vocabulary

- For annotations
  - For research areas
How we adapted VIVO and built AgriVIVO

2. Importers
VIVO importers

• VIVO allows for different types of “importers” to ingest contents from heterogeneous sources

• Some basic RDF and CSV are available in the core and can be used via the GUI to ingest data

• New custom importers can be written
  – To allow to parse different types of sources, map their elements to VIVO metadata and ingest them
VIVO custom importers
Importers: core and extensions

Our approach: One VIVO core with different extensions

• The same VIVO core with a combination of different extensions (ontology, importers, languages) instead of local hard-coded customizations
• Some importers can be packaged as “domain-specific” extensions and be re-used in the domain-specific community
• Some importers can be packaged as “scope-specific” extensions (e.g. importers from HR databases, importers from library catalogs)
• Importers that are general enough could be considered for inclusion in the core or as a general-use extension package
How we adapted VIVO and built AgriVIVO

3. Search interface
Search interface

Search portal (Drupal)  
This is NOT the VIVO tool

www.agrivivo.net
VIVO data > search interface
Search interface: local model

AgriVIVO search portal content model

Legend:
- Entity and its properties
- Reference list, taxonomy
- Published ontology
- Dashed lines and borders indicate future additions

Person
- ID: Numeric
- URI: URI
- First name: Text
- Last name: Text
- Email: Text
- Bio: Text
- Country: Reference
- Affiliation: Reference
- Expertise: Reference
- --- In the future
- Participated in:
- Role
- Author

Reference list
- Expertise keywords

Organization
- ID: Numeric
- URI: URI
- Name: Text
- Website: Text
- Email: Text
- Description: Text
- Country: Reference
- Type of organization: Reference
- Roles / job titles

Event
- ID
- URI
- Name
- Country
- City
- Start date
- End date

Publication

Geographic location
- ID
- URI
- Name
- + ontological relations

Ontology AGROVOC

Ontologies in the cloud / Integrated or linked to AgriVIVO ontology

Ontology VIVO

Ontology Geopolitical ontology
Search interface: importing the data

- **Drupal Linked Data Import module:**
  https://github.com/milesw/ldimport
  plugins for the Feeds module that let you turn remote linked data resources into Drupal entities

- **Customized for VIVO:**
  https://github.com/milesw/ldimport_vivo

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>TARGET</th>
<th>TARGET CONFIGURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
<td>URL</td>
<td>Used as unique.</td>
</tr>
<tr>
<td>Label</td>
<td>Title</td>
<td>Not used as unique.</td>
</tr>
<tr>
<td><a href="http://vivoweb.org/ontology/core#primaryEmail">http://vivoweb.org/ontology/core#primaryEmail</a></td>
<td>Email</td>
<td></td>
</tr>
<tr>
<td><a href="http://xmlns.com/foaf/0.1/firstName">http://xmlns.com/foaf/0.1/firstName</a></td>
<td>First name</td>
<td></td>
</tr>
<tr>
<td><a href="http://xmlns.com/foaf/0.1/lastName">http://xmlns.com/foaf/0.1/lastName</a></td>
<td>Last name</td>
<td></td>
</tr>
<tr>
<td><a href="http://vivoweb.org/ontology/core#webpage">http://vivoweb.org/ontology/core#webpage</a> []</td>
<td>Webpage: URL</td>
<td></td>
</tr>
<tr>
<td><a href="http://vivoweb.org/ontology/core#webpage">http://vivoweb.org/ontology/core#webpage</a> []</td>
<td>Webpage: Title</td>
<td></td>
</tr>
<tr>
<td><a href="http://vivoweb.org/ontology/core#overview">http://vivoweb.org/ontology/core#overview</a></td>
<td>Overview</td>
<td></td>
</tr>
<tr>
<td><a href="http://vivoweb.org/ontology/core#phoneNumber">http://vivoweb.org/ontology/core#phoneNumber</a></td>
<td>Phone number</td>
<td></td>
</tr>
<tr>
<td><a href="http://vivoweb.org/ontology/core#positionInOrganization">http://vivoweb.org/ontology/core#positionInOrganization</a> []</td>
<td>Organization</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.w3.org/2000/01/rdf-schema#label">http://www.w3.org/2000/01/rdf-schema#label</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Future plans
Integration of publications

- Integration of publications
  - linked to experts (authors)
  - retrieval from open systems (e.g. AGRIS for agriculture) using universal identifiers
  - possibly also manual curation by the experts themselves
  - Essential preliminary step: disambiguation of authors
VIVO ontology: author - publication

Integrates with BIBO
More complex model than just BIBO
VIVO introduces a new class for Authorships
VIVO ontology: author - publication

Another view of the author – publication model in VIVO
Disambiguation and identifiers

AgriVIVO as authority data for agricultural research actors

Disambiguating authors and researchers, sharing universal IDs
VIVO is collaborating with ORCID (http://orcid.org) and the Publish Trust Project (http://www.publishtrust.org/)

Disambiguating institutions
Using external naming authorities (VIAF?)
Becoming a subsidiary authority for agricultural institutions

Providing URIs and links between URIs for people’s and institutions’ profiles
E.g. link between a person’s AgriVIVO URI and the corresponding author URI in AGRIS or the corresponding ORCID
Coordinate with data providers

Work with data providers to **improve their data management environments** as a way to improve overall data quality at the source

- Study the changes that are necessary in order for information to merge coherently in the RDF store: e.g.:
  - map **competence/skill information** about experts with Agrovoc
  - map **Institutions’ names** with their URLs or other URIs (VIAF?)
  - Use identifiers for people; use email addresses to **identify people** and help merge duplicates and disambiguate records
Multi-language support

• Both for ontology labels and data
• Support for translations
• How to recognize translations when harvesting?
Interactive data curation?

AgriVIVO could also be used as a **community platform for interactive data curation**.

- Users can add/remove “relations” in which they are part of the relation: *person A “is author of” publication B, person A “participates in” project C*

AgriVIVO can also be used for **maintaining one profile** that can provide consistent information **across multiple websites**.

- the VIVO development team is exploring ways of propagating editing changes from VIVO back to the original source system
- Provide ability to edit VIVO profiles in a client environment?

**How to combine harvesting, manual curation and synchronization of data in sources?**
Getting data re-used

- VIVO’s search functionalities can be **integrated in other websites** through remote calls. In this way, specialized and targeted search engines can give access to and offer highly customized “views” of the data coming from AgriVIVO.
Better visualizations

AgriVIVO data → Semantic aggregation

→ Maps, charts, statistics

from http://impact.cals.cornell.edu/
• AgriVIVO portal: http://www.agrivivo.net
• AgriVIVO project: http://www.egfar.org/agrivivo
• VIVO portal at Cornell: http://vivo.cornell.edu/
• VIVOweb: http://vivoweb.org/
• VIVO search: http://beta.vivosearch.org/
• On VIVO: http://www.dlib.org/dlib/july07/devare/07devare.html
• VIVO going national: http://www.news.cornell.edu/stories/Oct09/VIVOweb.ws.html
• VIVO at USDA:

Contact: agrivivo@gmail.com