

<b>Project Name</b> VIVO	<b>Request Date</b> 4/24/13	<b>Project Number (AGIT Use Only)</b>
<b>Project Sponsor</b> Jeremy Kenyon, UI Library	<b>Desired Date</b> 7/1/13	<b>AGIT Owner (AGIT Use Only)</b>
<b>Short description of project / purchase (one sentence)</b> Implementing an open-source semantic web application which serves as a discovery tool for UI research activity and in doing so enables collaboration, transparency, and promotion of University research efforts.		

**ALL requested and/or recommended technology projects must complete the checklist below.**

*Definition: A project is a temporary endeavor undertaken to create a unique product, service, change, or result. A project is different from “business as usual” activities (operational work) in that it has a number of distinguishing features: it brings about change, has unknown elements therefore create risk that must be managed, and has a defined start and end (is temporary). In contrast, “business as usual” activities are characterized by having known policies, processes, procedures or precedents which may be followed, virtually no risk is present, the activities are not new but repeated (albeit not necessarily very frequently) and therefore does not offer change.*

**CHECKLIST**

YES	NO	CRITERIA
	X	1. Requires new ITS and / or other departmental information technology resources (including new or increased support needs)
		2. List amount of initial plus five-year on-going information technology costs for the project (i.e. first time hardware/software acquisition, professional services, annual maintenance, staffing, etc) \$ _____
	X	3. Requires new integration with existing information technology or electronic data systems such as Banner, FAMIS, R25, and ITS Identity/Access Management.
	X	4. Affects information technology aspects of multiple organizational components of the university.
X		5. Has information security aspects with respect to implementation of the information technology (privacy, confidentiality regulatory or compliance).

- **IF checking ‘NO’ to all criteria above, submit the completed checklist to AGIT for final review and approval.**
- **IF checking ‘YES’ to any of the criteria above, the Proposal form (below) must be completed and submitted to AGIT for further review and assessment.**

**AGIT Findings**

**1. Criteria utilized in review process**

As a tool to enhance the communication and connections related to university research activity, this proposal supports the priority and commitment for growing the university’s research enterprise. Based on the thorough details provided in the form including resource identification, value and functionality this is seen as a positive addition.

**2. Findings**

Approved to proceed with implementation and operation as noted in proposal

**3. Comments**

A couple questions were raised and addressed related to use and the involvement of ORED

**Proposal****OVERVIEW****1. University Responsibility** - Identify the name, email and phone number for the following:**a. Sponsoring Department**

University of Idaho Library

**b. Executive Sponsor** (highest ranking person responsible for acquiring all required resources necessary to successfully complete this project as well as removal of barriers that would impede its progress).

Lynn Baird, Dean of Library Services, [lbaird@uidaho.edu](mailto:lbaird@uidaho.edu), 885-6713

**c. Project Manager** (individual responsible for the day-to-day coordination and management of this project)

Jeremy Kenyon, Research Librarian, [jkenyon@uidaho.edu](mailto:jkenyon@uidaho.edu), 885-7955

**2. Description** -- Include in the description the following: the need, opportunity and/or the problem to be solved; the project scope including the definition of success; as applicable, describe the existing process that needs to be automated or enhanced; and for medium to large requests, provide a short description of each feature comprised in the project.

**General Description:** VIVO is a database that describes the network of people, publications, equipment, grants, and groups that make up a research organization. By describing UI's research activity, VIVO provides the users with the ability to search, analyze, and visualize an organization's research activities through a web browser. This includes viewing who is publishing what and where, what grants we (UI) have received and from what funding agencies, what subject areas our institution is strong in, and what facilities, equipment, and resources we use for research, as well as where the funding for those resources came from.

Our system is being developed alongside the growing network of other VIVO institutions. Eventually, networked with other VIVO implementations across the country, the linked databases will enable the discovery of similar research interests both within and across institutions.

**Technical Description:** VIVO contains "linked data". The data is modeled using the Resource Description Framework (RDF) [<http://www.w3.org/RDF/>], a semantic web standard that helps automate data-related tasks. Individual objects can be created and updated manually or automatically via data harvesting tools.

A VIVO database is a large graph, wherein the graph's nodes are objects, defined by a "class" depending on the content of the object. The class hierarchy is defined using a framework called an ontology. Ontologies define the context for the objects as well as the relationships between objects. Types of classes might include "Faculty Member", "Academic Article", "Grant", "Department", "Facility", "Equipment", etc. Each of these is an independent object which needs to be related to another in order for it to mean anything. Here are several examples of relationships in VIVO:

"Faculty Member" (object A) has authored "Academic Article" (object B)

“Academic Article” (object B) was produced under “Grant”(object C)

“Grant” (object C) is investigated by “Faculty Member” (object A)

In essence, the entire database is comprised of objects connected by these sorts of relationships.

**Software:** MySQL is the database server. The semantic web engine that presents these data is a Java-based web application that resides within a Tomcat servlet container. The data is indexed using Apache Solr. The system is currently being served on the Linux-based Northwest Knowledge Network architecture.

**Need/Opportunities:** We have identified several reasons for this system to be implemented at the University of Idaho:

- The library has been looking for a solution to provide UI with a “publications database” – a system of record for University of Idaho publications. The Register used to provide this service, but does no longer. VIVO presents such a solution.
- Library staff have been on the receiving end of questions such as: how can we gather publications that were generated under a grant and show them in an easy fashion? Not having a tool for doing this has resulted in dozens of ephemeral websites just to show the output of a grant. VIVO can help mitigate this.
- VIVO can be presented as an endpoint for web services, providing a layer of metadata about UI that will enable applications that improve discovery and visibility of UI researchers and research activities. Examples might be the UI website, which could query these data and populate faculty profiles directly, or the proposed UI mobile app, which could query this information and use it however they wish.
- No current system exists that allows the discovery of people, research activities or interests. The closest open system is the UI website, which provides simple HTML-based individual and unit profiles. These pages typically suffer from a lack of updating, standardization, and consistency, not to mention the inherent challenge in searching across the entire UI website just to find someone with similar research interests.
- VIVO is open source software. Time and labor spent on its initial implementation and ongoing maintenance is the primary cost, as the systems infrastructure to support this product currently is available with NKN.
- VIVO is standards-based (OWL, RDF/XML), which means the data ingested will be able to be exported and/or migrated for use in related projects or future systems.
- There are dozens of instances at both academic institutions and federal agencies, suggesting growing acceptance in parts of the research community. (see: <http://vivo.cornell.edu>; <http://vivo.ufl.edu>; <http://vivo.colorado.edu>)
- VIVO-based applications, such as [www.vivosearch.org](http://www.vivosearch.org) and <http://about.vivosearchlight.org>, require VIVO profiles and suggest the potential for this application.

**Scope/Criteria for Success:**

We are looking to expand this system to the rest of the University. If successful:

- VIVO will be a fully populated, useful means to identify what UI researchers are doing and have done.
- Librarians will function as outreach agents to help procure buy-in from Colleges/Centers/Institutes.
- Automated harvesting from local databases (such as EIPRS) will be set up to make updating the system easier

- Most labor cost of maintenance will be absorbed into the staff we currently have.

**Required Automation:** None is required. Two possible automation-related requests from ITS are: 1) providing NetID authentication via LDAPS protocols to provide a more seamless login experience for users of the system; 2) using the harvesting tool, which has a JDBC/ODBC function to retrieve data from ITS-managed RDBMS.

### 3. Required Costs and Resources

a. **Budget estimate** -- *Include both one-time and ongoing for hardware, software, staff hours (both ITS and non-ITS personnel -- including within the sponsoring department and others), consulting /professional services.*

- No hardware/software costs, built on existing hardware using open-source software. Project can scale without requiring new hardware.
- Negligible time commitment from NKN required. >0.1 FTE
- At least half-time commitment from Library (0.5 FTE). This may be distributed among several classified and student staff but will likely require new temporary staff during the “build” stage of the project. Time commitment is expected to decrease once the initial bulk data ingest is complete and we shift to “maintenance” phase.
- 1.0 FTE for outreach/training to be distributed across the 10 library liaison librarians and other staff

b. **Financial commitments /resources for above**

- NKN has committed the >0.1 FTE for system administration
- Library administration is investigating the acquisition of temporary support for the .5 FTE requirement. Beyond this position, existing staff can absorb the cost.

## PROJECT BACKGROUND

1. **Context** -- *Describe if it is a regulatory requirement, a mandate, supportive of a local tactical/operational plan, or supportive of a specific element of the University of Idaho strategic plan.*

VIVO was initially developed at Cornell University and then was expanded under an NIH grant with nine other universities. NIH has been encouraging use of this system to other NIH-funded bodies.

Carolyn Bohach, Director of Idaho INBRE, asked us to investigate this tool in Spring 2012, and with their cooperation – 19 volunteer faculty – we created the proof-of-concept version of VIVO for UI. In that time, we learned the system architecture, how to add/remove/modify data, manipulate the presentation of the data, customize the organizational structure for UI needs, change the style templates, and configure the harvesting tool to automate the gathering of publication data from Pubmed and Web of Science. We have also identified open-source tools, such as Google Refine, now OpenRefine, that allow us to do normalization and mapping of tabular data, such as spreadsheets, and batch load them into VIVO.

The VIVO project supports:

- UI Strategic Plan, Goal 2, Objective A, Strategy 6: Increase the application of and public access to the results of scholarly and creative activities.
- UI Strategic Plan, Goal 2, Objective B, Strategy 1: Expand opportunities for ongoing interactions among faculty, students and staff to identify areas of common interest.
- UI Strategic Plan, Goal 2, Objective B, Strategy 4: Increase the national and international visibility of the University’s contributions to interdisciplinary activities.

- UI Strategic Plan, Goal 3, Objective B, Strategy 3: Increase the external visibility of our outreach and engagement activities.
- UI Library Strategic Plan, Mission Statement: the University of Idaho Library will 1) facilitate access to library resources, experts, and collections in both traditional and innovative ways, 3) foster and support all stages of research and scholarly activity

**2. Value proposition** -- *Identify projected cost savings and/or efficiencies and/or profitability (if applicable) and the associated timeframe as well as related process or service improvements.*

- Any inefficiencies (time taken searching, emailing, setting up meetings) in identifying or categorizing research activities across campus can be greatly improved on with implementation of VIVO
- For project teams that use VIVO, reporting out of relationships between grants and publications or equipment can be improved on.
- Alternatives to this system are typically commercial software systems that are very expensive (SciVal, Research In View for research visualization; BePress for institutional publications). VIVO, being open-source, costs us labor only. There is a development community actively continuing to develop this project and develop new features, so there are opportunities for interacting with this community as well.

**3. Beneficiaries** – *Identify who on campus or which off-campus constituent group will directly or indirectly benefit.*

- Faculty – both those who need to present their research activity and those who are trying to find out about others' research activity
- Off-campus researchers who are disconnected from daily, in-person opportunities to discover potential collaborators, particularly in other disciplines
- UI administrators who need to review and identify what research is going on in the University
- External funding agencies - reviewers and evaluators who wish to see the relationships between grants and scholarly output, as well as review research activity of investigators
- Public and State of Idaho officials who wish to see what goes on at a research university

**4. Risks** -- *Describe all risks associated with doing as well as NOT doing this project.*

- **Risks with doing:** This is a low-risk project. The data included should only be data that is considered open to the public, perhaps already in the public sphere, but not yet aggregated in a useful way. There is little security risk. Back-end access to the system can be heavily restricted through the use of roles. We are permitting very few site administrator roles and limiting most new accounts to "self-editor", which allows you to edit your own account only. There may be some security questions that arise when making the database queriable as a SPARQL endpoint, but we have not yet explored that feature.
- **Risks with not doing:** There is a large opportunity cost. All of the opportunities listed above would be missed, and the University would continue to inefficiently use HTML web pages to as the only means of presenting research activity, fail to track any publication activity in aggregate, and continue to hunt for solutions to bring this information together. Also, the University of Idaho Library would likely need to pursue a commercial institutional repository system, instead of this open-source opportunity.

**5. Alternatives Considered** -- *Document alternatives considered and why rejected. Changes in business processes to accommodate existing technology solutions should be considered and documented.*

- There is nothing exactly like VIVO. For research activity analytics and visualization, the alternatives to this are research analytics tools from major publishers, namely "SciVal" from Elsevier and "Research In View" from

Thomson Reuters. Each of these products are prohibitively expensive for the Library. For presentation of research outputs, such as publications, repository software, such as BePress or Dspace, is limited in functionality. They are for preserving digital publications, but are dependent on the publications in the system, cannot automate harvesting, and fail to provide as holistic a view of research activity as VIVO does.

**6. Timeframe and Urgency** -- *Identify relevant timeframes or related timelines. Do we need to do this right now or can it wait? Define the level of urgency and priority compared to other work, both project and operational.*

- There is no urgency for this project as we have met the initial need expressed from Idaho INBRE. There is a desire to begin this process during the summer of 2013, as the moment seems ripe to move forward and it requires little effort on the part of the rest of the university community to support this project.
- Tentative short-term timeline (abstracted from a larger plan of tasks):

<b>Summer 2013</b>
Procure buy-in from key UI stakeholder groups (VPs, Deans, Directors, etc.)
Develop a controlled vocabulary for research interests
Develop demographic and structural outline of UI
Extend the VIVO ontology for UI unique needs (INBRE, EPSCoR, Campus Centers)
Upgrade current installation version from 1.4.1 to 1.5.2
<b>Fall 2013</b>
Begin acquiring publications data for ingest
Begin working on awards data acquisition from EIPRS
Begin acquiring photos from UI website
Begin communicating with WebComm about branding

- System development is open-ended and dependent on campus-wide cooperation and approval, but our aim is to finish the build phase by the end of FY14.

**7. Constraints** -- *Identify constraints such as budget, schedule, staff, equipment, facilities and infrastructure, etc.*

- Currently, NKN support is given free of charge. As a result, it is low priority when problems arise. This is not anticipated to be a major problem, but it is a constraint.
- Library staff, namely Jeremy Kenyon and Devin Becker, have other commitments and require support if this project scales up. Library administration is currently investigating this.
- Some advanced harvesting of data requires expertise that neither librarian has and may need assistance to efficiently solve some of the technical issues.

**CONNECTIONS**

1. **System impacts / dependencies** -- *Identify any prerequisite projects as well as any related or connected processes or projects that will need modification or changes as a result of this project. Describe how existing systems and processes need to be modified or extended to accomplish this project.*
  - There are no existing systems that should be negatively impacted nor modified by implementation of VIVO. It is not intended to replace any existing systems, but rather, to augment them. Systems may be positively impacted if data in VIVO is linked out to them, thus providing a discovery service leading to those systems.
  
2. **Enterprise Systems Integration** -- *Requirements for interfacing to all institutional IT/ ERP systems including but not limited to: ITS Authentication (LDAPS), Banner – HR, Banner – Finance, Banner – Student, VandalCard, Document Imaging, Web Services, and Help Desk.*
  - Account login does not require LDAPS use, however, the user experience of the system may be improved with integration with this system.
  - One of the benefits of the system is setting up automated harvesting tasks from local databases. This would require connections to these systems.
  
3. **Stakeholders -- impact and/or buy-in** -- *Identify various stakeholders (both from within ITS and outside) whose support is required to make this project successful. Has their buy-in been secured? If not, what conditions or constraints have they placed on their support?*
  - Key stakeholders (without which this project will not work): the UI Library, NKN, ORED, ITS
    - The library and NKN participation has been secured.
    - We are in the process of approaching ORED and ITS
  - Major stakeholders (without which this project is weakened): the UI faculty community, University Communications and Marketing
    - We plan to approach Colleges and Institutes from the administrative level first to procure buy-in, before moving to individual faculty. Idaho INBRE is already on board.
    - We plan to approach University Communications once all key stakeholders have been included.