VIVO-In-A-Box: A Proposal to Develop A Simplified Version of the VIVO RIM System

RIM/CRIS System Support of Research Organizations

The incentives that drive research organizations are focused on generating greater prestige and reputation at regional, national and international scales, along with the resources that often accompany prestigious reputations (Bryant et al., 2020). Research organizations are increasingly focused on strategies that can enhance their reputation and resources, especially in times of uncertain funding. In interviews of senior research officers (SRO) at US institutions, virtually all interviewees prioritize their organization's research competitiveness, as reflected in funded awards and faculty recruitment/retention and many were instrumental in organizing strategic research collaborations (Rieger and Schonfeld, 2020).

SROs also viewed research support as important to enhancing their organization's research competitiveness, though providing these services are not without their financial and management challenges (Rieger and Schonfeld, 2020). Research information management (RIM) and current research information systems (CRIS), including VIVO, are often coupled with other software to form an interoperable suite of tools that can meet emerging and significant academic organizations needs including enhancing scholarly reputation, supporting team science, and conduct research intelligence that transforms authoritative data in RIMs into actionable insights on organization research and its scholarly impact and societal relevance.

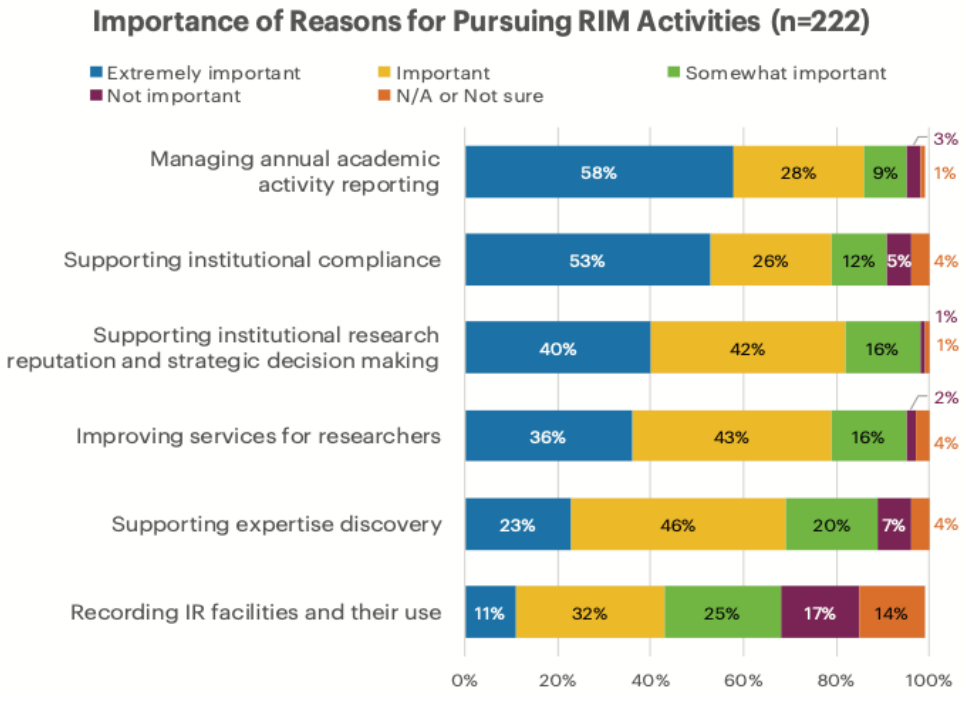


FIGURE 11. Importance of reasons for pursuing RIM activities for institutions with a live RIM system (n=222).

The development of a RIM/CRIS system as a research support service can face significant challenges. Unfortunately for most universities, research information is often fragmented across various systems within an institution. This requires individuals and units across the university, often including libraries, to work across internal silos (Bryant et al., 2020). Funding for research support, including RIM/CRIS systems, can also be at risk – especially in uncertain budgetary times (Radecki and Schonfeld, 2021).

This dual challenge for both financial resources and cross-institutional expertise is interesting to consider in light of the diversity of institutional units responsible for their university's RIM system (Bryant et al., 2018). While library and/or IT staff have important expertise that can support the design and implementation of a RIM or CRIS system, it is the administration that often has access to the financial resources.

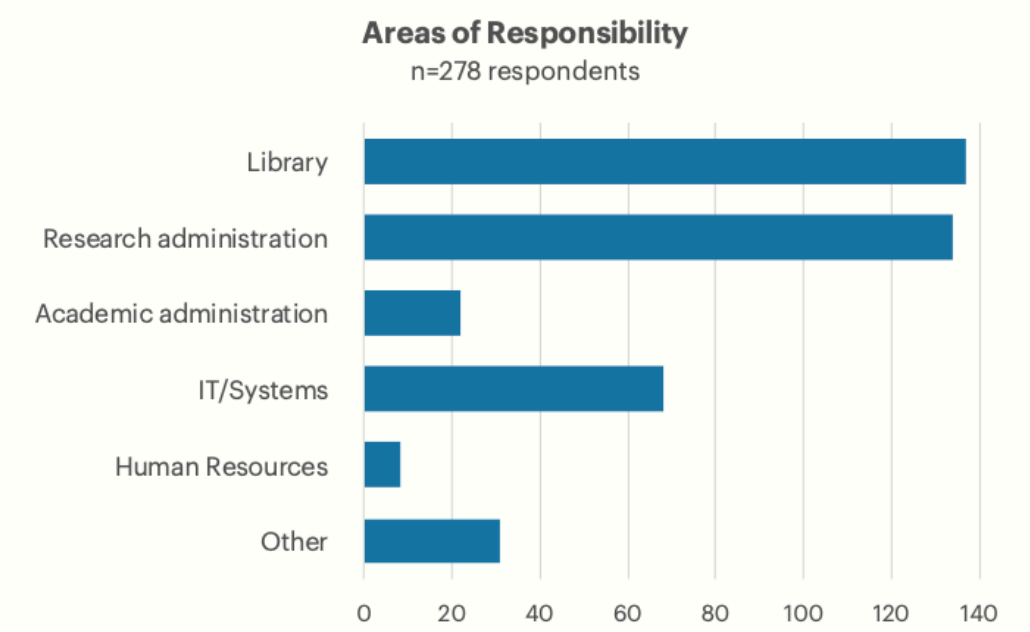


FIGURE 2. Areas of responsibility of survey respondents, where respondents could select more than one option.

Recent Inquiries on Implementing RIM/CRIS Systems

This dual challenge for both financial resources and cross-institutional expertise arose in recent discussions with a range of US institutions exploring the implementations of a RIM system for their university. Over the past couple of years, a large number (~25) of American Universities have contacted the VIVO team at Texas A&M to ask about our VIVO implementation, Scholars@TAMU. The conversations are often initiated by a librarian who then brings in administrators for subsequent meetings. The organizations are fairly consistent in their goals. They seek to develop a RIM that (i) enhances the institution's reputation, (ii) supports interdisciplinary research and team science, (iii) develops an authoritative database of organizational work, particularly of research where the data can be reused for business purposes, and (iv) supports research intelligence.

The organizations are also challenged in much the same way. The conversations often start out that they would like to implement the Texas A&M's VIVO system but when they hear the costs, both in

terms of money required to purchase commercial systems (Symplectic Elements, Altmetrics) and people (System Admin & programmer time, librarian and data analyst, and my efforts), they balk. In fact, of all the conversations only one institution (Oklahoma State) has moved forward to implement VIVO. The organizations often cite difficulties in finding money to buy commercial software or realign people to support system development.

The organizations are interested in a more scaled model of implementation, where they can start with a simple system that limits costs, allows the organization to build skills and knowledge, and identify important use cases.

User Stories (Personas)

Limited Resource Organizations: Many organizations, including universities or research collaborations, either have limited financial or personnel resources or be unwilling to commit needed resources until a proof-of-concept system has been established and evaluated and yet are intrigued by the possibilities offered by a RIM system. For these groups, they are interested in research information about their organization. Since the RIM system is being tested in a trial period, the project is often helped by early project successes.

Organizations Requiring Software as a Service: Some organizations have adopted a strategy of or software as a service and are interested in a similar VIVO system as the *Limited Resource Organizations* but want VIVO to be a cloud-based systems.

Full-Service Organizations: Organizations with the financial resources and expertise to support customized implementations of VIVO that is designed to meet specific institutional needs. These systems often utilize commercial systems, such as Symplectic Elements, or hosted systems like those offered by Clarivate.

Organizations with Limited Resources	Organizations with Limited Personnel	Organizations Seeking to Support More Robust Services
• Easy system installation on local servers • Limited customization • Branding • No commercial systems needed • Early wins	• Software as a service • Limited customization • Branding • No commercial systems needed • Data is managed locally • Even earlier wins	• Customization of profiles to serve institutional context • Interoperable system • Data reuse through API • Visualization/research dashboards

Project Goals

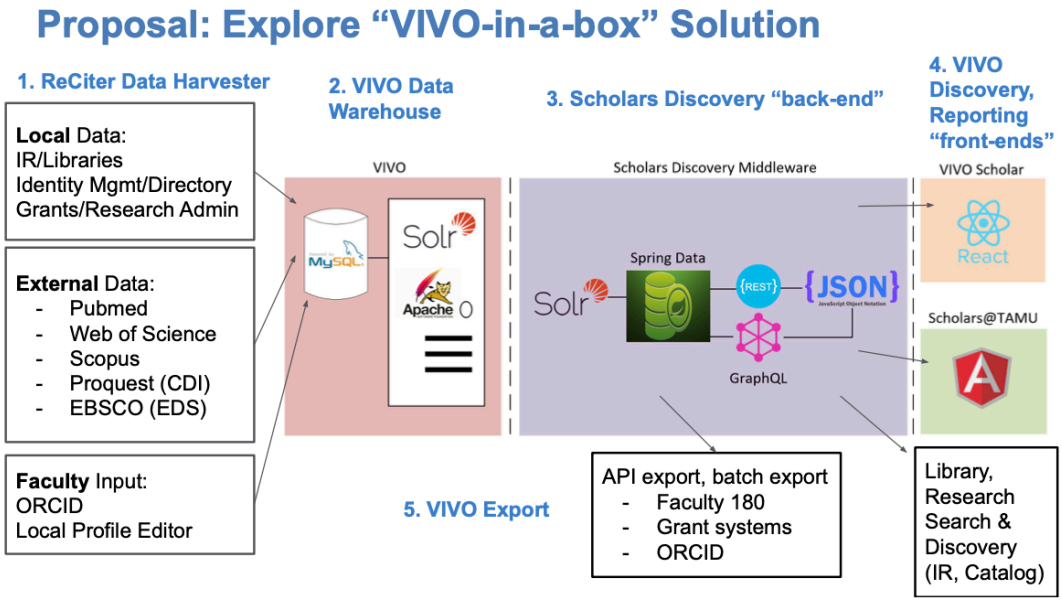
Development of a simple VIVO, VIVO-In-A-Box, would support the VIVO community by increasing the number of VIVO implementations and increasing the number of potential members. Therefore, we propose the following strategy:

1. Develop a simplified VIVO-In-A-Box version of VIVO that reduces the costs and personnel needs for implementation and maintenance. VIAB features include:
    - All open-source components
    - Easy installation
    - Simplified data harvesting and ingestion from a limited number of sources to fill simplified researcher profiles
    - Limited customization
    - User interface optimized for accessibility
    - Improved profile editor
    - Support data reuse and/or reporting
2. Explore developing a software as a service (SaaS) solution
3. Develop training programs that can support university teams in getting their system up and running

Damaris outlined three key benefits that will set VIAB apart from VIVO:

1. Viable option for smaller budgets - Low-cost, high-reward. Easy to implement. Quick win for research intelligence, advancement, and communications efforts at your university.
2. Out of the box features for a wide range of stakeholders – enables reporting and visualization options that benefit various audiences.
3. Building the academic community by creating connections – focus on a practical/non-threatening/social approach to using linked data. Help distinguish this tool from benchmarking tools.

VIAB System Description



The VIAB project provides an opportunity to tackle some important development needs. Andrew Woods described this as an opportunity to decouple the application layers, embrace scalable architectural patterns, and support clear ingest patterns. Additionally, a modern reboot of VIVO would likely draw broader development interest from the community.

Data Harvesting and Ingestion. We propose that VIAB includes a data harvesting and ingestion system that harvests publication data from a limited set of sources. In particular, we discussed developing a system that harvests from Web of Science, allowing Clarivate to help market VIVO to organizations that subscribe to WoS (~9000 organizations). This, of course, also brings benefits to Clarivate – which is a good thing as they are one of our most important partners. One option is to base the data harvester on Weill Cornell's Reciter program.

Paul Albert (Weill Cornell) offered some important insights on the use of Reciter as a data harvesting system. Goals for the development project would include:

- Offer an open source software system which allows institutions to use locally owned identity data and machine learning to quickly/easily maintain current publication lists for their scholars
- System is offered via software as a service. Institutions can set it up and host it independently, but they can also pay a service provider (Clarivate) to manage the AWS hosting and/or some of the technical setup work.
- Publication data can be exposed through set of APIs, a reporting database, and via VIVO scholarly profiles
- Even with service provider, system could be inexpensive (excluding setup costs: ~$15k/year???) to operate.

    - Hosting costs (depending on number of scholars): $5k/year. Also, there are significant (70%??) potential cost savings if multiple sites share a single AWS instance. Kubernetes allows sites to use the same instance without sacrificing security or privacy.
    - Duraspace / future development: $5k/year
    - Service provider: $5k/year (???)

Major software development required for application:

- Complete user interface improvement (WCM in progress)
- Integrate with VIVO 1.11 (WCM in progress)
- Fine tune the CloudFormation template for automatically creating all the components / dependencies for the system across AWS (mostly done?)
- Create a connector to Web of Science
- Abstract ReCiter data model so it is not PubMed centric (significant work required)
- Create an article disambiguation layer (significant work required depending on approach chosen)
    - Question: do we even need this? Web of Science could be configured to be the exclusive data source at least in the early going
    - Question: hard-code (easier) vs. allow user overrides (harder)?
    - Question: allow system to be matching like Source A-B, A-C, and B-C vs. only A-B and A-C but not B-C
- Update UI so it allows for manual lookups of Web of Science

Major work required for each new adopting institution:

- Get identity data from source systems
- Configure the properties file (relatively easy)
- Write script to recurrently import data into ReCiter
- Write script to run ReCiter APIs as desired
- If institution wants a VIVO profile system, grants, appointments, educational background, etc. would need to be collected and imported into VIVO

Minimal Viable Researcher Profile. Based on Don Joon Lee's research at Texas A&M, the metadata elements that have the most **value** in terms of faculty usage in RIMS are:

1. faculty name (and headshot)
2. affiliation & contact information
3. position title
4. ~~research~~ overview
5. research areas/keywords, and, of course (faculty input vs harvesting keywords from objects)
6. publications.

Profiles with the six parts, above, would be the minimal profiles that meet common use cases for a RIM or CRIS. If the community wants to add more sections, possibilities:

1. Teaching activities for instructional faculty (mainly support teaching faculty's reputation only; no value as institutions' data yet)
2. Grant (for faculty reputation only, not yet serve as data (lack of relationship between publications and grants))
3. Instruments/facilities – for industry – conversations around this idea
4. Patents
5. Innovation work (Duke) – future work, research directions

Profile Editor: Since faculty remain the authoritative source for their research information (though their metadata often contains errors), the RIM system would benefit from a profile editor.

Implementation Process

Our implementation plan needs to involve all stakeholders and build buy-in among the wider VIVO community. Here are more details:

1. Bruce authors the VIAB proposal. This is sent to the Leadership Group for comments and (once any changes or clarifications are made) approval.
2. Bruce sends the VIAB proposal to all TF and committees asking for their input and ideas on a plan to create the product and associated services/programs. (tentative: 1 month). Bruce and other Leadership members (Robert C.? Weill-Cornel team for Reciter?) is available to meet with TF and Committee to answer questions and discussion.
3. TF/Comms submit recommended design/implementation plans to Leadership Group for discussion and evaluation of how everything fits together. Decisions are made and questions answered.
4. Steps 2 & 3 are repeated until both Leadership Groups and TFs buy into the design and implementation plan.
5. At a time when the VIAB design and implementation plan are starting to take shape - but before it is finalized, a small group from the Leadership and developer groups hold a VIVO community town hall to discuss the developing plan and seek input from the wider VIVO community.
6. Plan is implemented. Marketing commences. We seek early adopters who help us by providing feedback.

References

Bryant, R., Clements, A., de Castro, P., Cantrell, J., Dortmund, A., Fransen, J., Gallagher, P., and Mennielli, M., 2018, Practices and Patterns in Research Information Management: Findings from a Global Survey: OCLC Research.

Bryant, R., Dortmund, A., and Lavoie, B., 2020, Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise: OCLC Research.

Radecki, J., and Schonfeld, R. C., 2021, Academic Research Budgets: A Look Ahead with Special Emphasis on Research Enablement and Suppor: Ithaka S+R.

Rieger, O. Y., and Schonfeld, R. C., 2020, The Senior Research Officer: Experience, Role, Organizational Structure, Strategic Directions, and Challenges: Ithaka S+R.