

Call for Interest: Integrating DSpace with VIVO

December 2021

Overview

LYRASIS is seeking those interested in working on a two phase project to further a VIVO DSpace integration.

Background

Depositing of open-access publications in a repository such as DSpace can be a source for monitoring an open-access policy compliance at an institution. On the other side, research information systems such as VIVO enable reporting and monitoring on research activities and achievements. Integration of those two types of systems can lead to an integrated all-in-one platform for unique and comprehensive monitoring of the research domain at one institution. Also, this integration avoids duplicated efforts for cataloguing information about publications and researchers (authors) in those two types of platforms. This document defines specification of a short-term project which should result in implementation of features in the VIVO platform which should enable depositing metadata and files in DSpace repositories. The project will be funded by VIVO. The project will be implemented in two phases. This is a call for expression of interest in participation in the first phase of the project.

LYRASIS

LYRASIS [1] is a non-for-profit organization which is a leader in open technologies, hosting, data-migration, content licensing, and community supported software programs for libraries, archives, museums and research organizations worldwide. The organization catalyzes and enables equitable access to the world's knowledge and cultural heritage. Moreover, LYRASIS helps its members succeed by working with them to identify their needs, issues and challenges and providing products, services and learning experiences to address them. LYRASIS brings together several critical open source technologies (including DSpace and VIVO) all under one roof, giving members and users shared infrastructure, enhanced development of the software and a strong backbone for sustainability.

DSpace

DSpace [2] is an open source repository software package which is used for creating digital repositories for scholarly institutions' outputs. There are over three thousand instances of the DSpace platform around the world [3].

VIVO

VIVO [4] is a member-supported, open source software and an ontology for representing scholarship. VIVO supports recording, editing, searching, browsing and visualizing scholarly activity. VIVO encourages research discovery, expert finding, network analysis and assessment of research impact. VIVO is easily extended to support additional domains of scholarly activity.

Goals of integration

- Streamline process for academics
 - Avoid duplicated bibliographic data management
- Adding semantic web aspect to existing DSpace repositories
- Adding depositing files and monitoring of Open-access policy compliance to existing VIVO instances. VIVO instance might be used as a front-end for researchers where they can be motivated to create their own profile page including a list of research results (publications, datasets, etc.). For any single research results researchers might provide DOI/URL in the metadata and/or initiate depositing files to DSpace through VIVO user interface, while the rest of the files processing will take place in Dspace by librarian/officers.
- Growing community for both platforms
 - Knowledge transfer between selected team members, improving their capacity through collaboration, and engaging new developers for the VIVO and DSpace community.

Requirements

The first phase

Functional requirements

- VIVO doesn't store full text articles and other research outputs (e.g. dataset) - it transfers them to a digital repository
 - Adding a file (or files) to VIVO entities (publications, datasets, etc.) through VIVO UI
 - Use DSpace REST API to deposit file items and update metadata in DSpace
 - URL to file deposited in DSpace is preserved in the VIVO and visible to VIVO users

- Crosswalks - definition of the mapping between VIVO ontology and DSpace internal model
 - XML settings files used to specify how data fields are mapped between the two systems
 - Default crosswalks supplied for standard fields
 - Mappings for both directions
 - Inbound DSpace -> VIVO
 - Outbound VIVO -> DSpace
- Migration batch
 - Harvest all DSpace items and ingesting in VIVO with clear database

Non-functional requirements

- Communication between VIVO and DSpace should be based on CSRF Tokens [7]
- All new features should be covered with
 - Logging
 - Tests
 - Wiki documentation

The second phase

Functional requirements

- Support to copyright checking in DSpace
 - VIVO deposits items to the DSpace workflow
 - VIVO checks and shows status of an item (e.g. First deposit, Published), when DSpace item is published, the URL is visible to VIVO users
- Enable adjusting crosswalks to take advantage of evolving systems
 - Crosswalks capabilities
 - String manipulation (split, concatenate, format, etc.)
 - Boolean algebra
 - Map multivalued compound data types
 - Dictionary lookups
 - Registry lookups (based on identifiers such as ORCID)
 - Regex
- Support multiple DSpace repository connections in VIVO
 - e.g. separate repositories for datasets & publications, or
 - in different departments
 - selection of a DSpace repository in which metadata and a file (or files) should be deposited
- Migration batch
 - Harvest all DSpace items and matching to existing publications in VIVO
 - Monitoring for changes in DSpace
 - VIVO preserves in its database the date of last DSpace harvest
 - daily or weekly update might be configured
 - OAI-PMH ListIdentifiers request with defined *from* parameter to collect identifiers [5], and to use them for collecting item information through DSpace REST endpoint [6]

Non-functional requirements

- All new features should be covered with
 - Logging
 - Tests
 - Wiki documentation
- Create a dockerfile for VIVO-DSpace custom build

Application & selection procedure

Eligible applicants

Anyone interested in the topic of the project. However, applicants' skills stated in CVs and letters of interest will be analyzed and candidates with following skills, knowledge and experience will have advantage in the selection process:

- Digital documents management
- The VIVO platform
- The DSpace repository
- Java
- Semantic web technologies
- Formats' mappings/crosswalks

Proposed timeline

The first phase

- **Application deadline - January 17th**
- First round of selection/review results - January 24th
- Interviews as needed - February 7th
- Final selection - February 10th
- Implementation - February 14th - May 14th
- Reviewing of developed code - May 28th
- Code correction - June 21st
- Documentation- June 30th

The second phase

To be defined after completion of the first phase.

Application

All interested applicants should submit a short CV and letter of interest via vivo@lyrasis.org email by **January 17th**. The application is only for the first phase of the project, although selected candidates will have priority in the negotiation for the second phase of the project.

Selection

We are targeting to select a team of applicants representing a good synergy and complementary team members, ideally a combination of VIVO core committers, DSpace core committers and developers outside of those groups interested to join VIVO and DSpace communities. The team should contain 2-3 members.

The first round of selection will be conducted based on CVs and letters of candidates. The final selection will be based on online interviews where the selected candidates in the first round should present their availability, cost and plan for the implementation.

Funding

The total budget of the project depends on experience and plans of selected candidates. The budget for the first round will be in the range 6,000 - 10,000 USD.

Eligible costs

- honorarium fee for software developments

Reporting and Monitoring

All project team members have to participate in regular weekly calls for the VIVO developers interest group and to report about the project's progress and discuss the issues.

The developed source code should be a VIVO contribution delivered in the form of a GitHub pull request in accordance with the guideline for contributing to VIVO [8]. After reviewing the pull request by at least two VIVO core committers, the project participants have to correct their code in accordance with reviewers' suggestions. The developed solution should be tested (unit testing and smoke testing) and documented.

References

- [1] LYRISIS, <https://www.lyrasis.org/>
- [2] DSpace, <https://wiki.lyrasis.org/display/DSPACE/>
- [3] The registry of DSpace instances, <https://duraspace.org/registry/>
- [4] VIVO, <https://wiki.lyrasis.org/display/VIVO>
- [5] The OAI-PMH ListIdentifiers request, <http://www.openarchives.org/OAI/openarchivesprotocol.html#ListIdentifiers>
- [6] The DSpace 7.x REST endpoint specification, <https://github.com/DSpace/RestContract>
- [7] CSRF Tokens, <https://github.com/DSpace/RestContract/blob/main/csrf-tokens.md>
- [8] Contributing code to VIVO, <https://wiki.lyrasis.org/display/VIVO/Contributing+code+with+a+fork%2C+branches%2C+and+pull+requests>