Initiatives for making VIVO platform more interoperable

euroCRIS Strategic Membership Meeting,

Nijmegen, The Netherlands

December 2nd, 2022

Background

- VIVO is a semantic web platform enabling research discovery
- It is based on the VIVO ontology and might be used as a research portal for an institution or region
- Moreover, it can be customized to support some other roles in the research information ecosystem
- Furthermore, because of its nature and the fact it is based on semantic web technologies it might be part of a distributed research information system enabling cross-linking and cross-searching of information
- Following the VIVO product roadmap established for this year, some initiatives have been launched in line with the strategy to enhance the interoperability of the VIVO platform

Mapping between CERIF and VIVO ontology

CERIF and VIVO

- In July 2021 a Memorandum of Understanding between VIVO and euroCRIS was signed
- One specific action started after the signing of the MoU is updating the mapping between the CERIF data model and the VIVO Ontology from 2013
- The goal is to make possible communication between the VIVO platform and research information system capable of producing or ingesting research information in the CERIF format or its profiles such as one used in the OpenAIRE guidelines for CRIS managers

Status and plans

- In one direction at the moment - CERIF to VIVO
- Matching of structure
- Mapping between used vocabularies
- Review (help needed)
- https://docs.google.com/spreadsheets/d/19vmYxsyawNi3W0JkToCc7CDUvjTSluoULPBsAgb-nEU/edit?usp=sharing
- Formal language for representation of mapping
  - Maintaining and versioning of the mapping

Integration of VIVO platform and DSpace repository

Goal

- Streamline process for academics
  - Avoid duplicated bibliographic data management
- Adding semantic web aspect to existing DSpace repositories
  - Reasoning - infer new facts from existing data based on inference rules or ontologies
- Adding depositing files and monitoring of Open-access policy compliance to existing VIVO instances

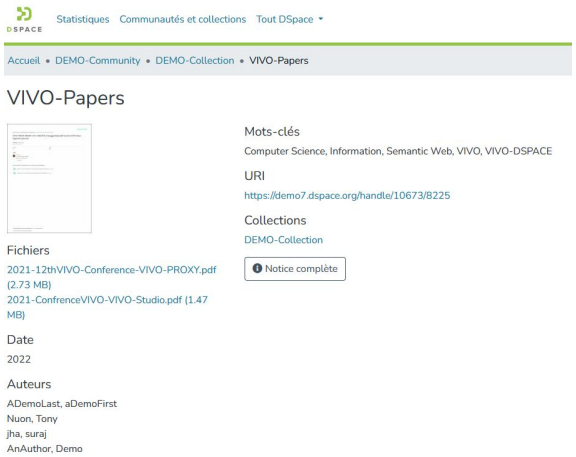
Status

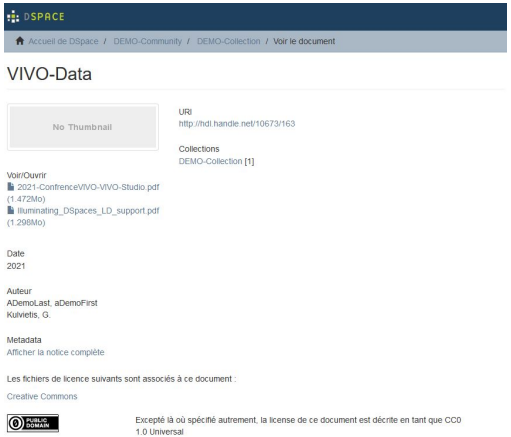
- The first phases completed in June
- Usage scenario
  - A new fresh and empty instance of VIVO is installed. There is a running instance of DSpace. Configure and run the migration process by using bash scripts as it is described at https://wiki.lyrasis.org/display/VIVO/Migrate+DSpace-Metadata+from+DSpace+to+VIVO. After initial migration of data from DSpace, VIVO UI for editing and adding new items has been turned off (**VIVO is used in read only mode**). Weekly or daily, there is the update of information in VIVO during the night by dropping triplestore, and running migration with a fresh copy of DSpace data. The update is defined as a bash script run as a scheduled job (for instance a cron job).
- Planning of the follow up phases is in progress

The first phase of the project

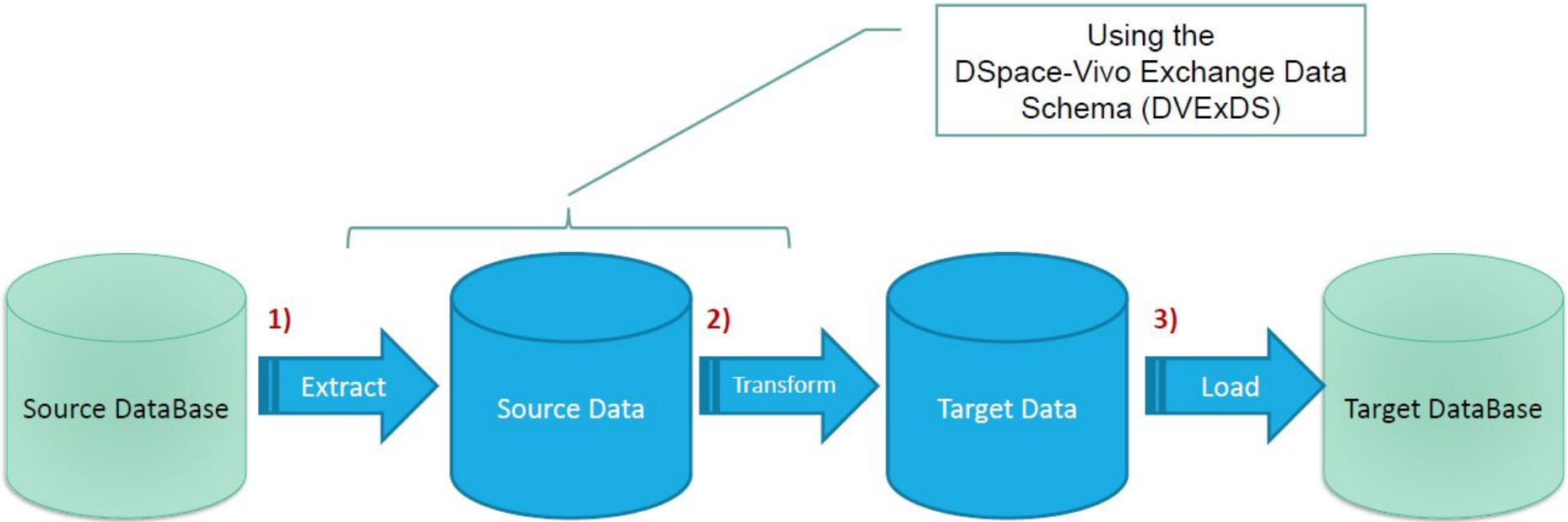
- Call for expression of interest published - beginning of January 2022
- Procedure for selection of candidate
  - Michel Héon, Canada
  - José Ortiz, Ecuador
  - DSquare, India
    - Abhishek Raval and Narendra Kumar
- The project started in February
- The end of the first phase - June 30th

- The goal - migration of data from operational DSpace to VIVO in read-only mode
  - A new presentation of DSpace items and semantic web aspect to existing DSpace repositories
  - DSpace 6 or 7





ETL approach



https://github.com/vivo-community/DSpace-VIVO

https://wiki.lyrasis.org/display/VIVO/DSpace-VIVO+Technical+Documentation

DVExDS

YAML Metamodel

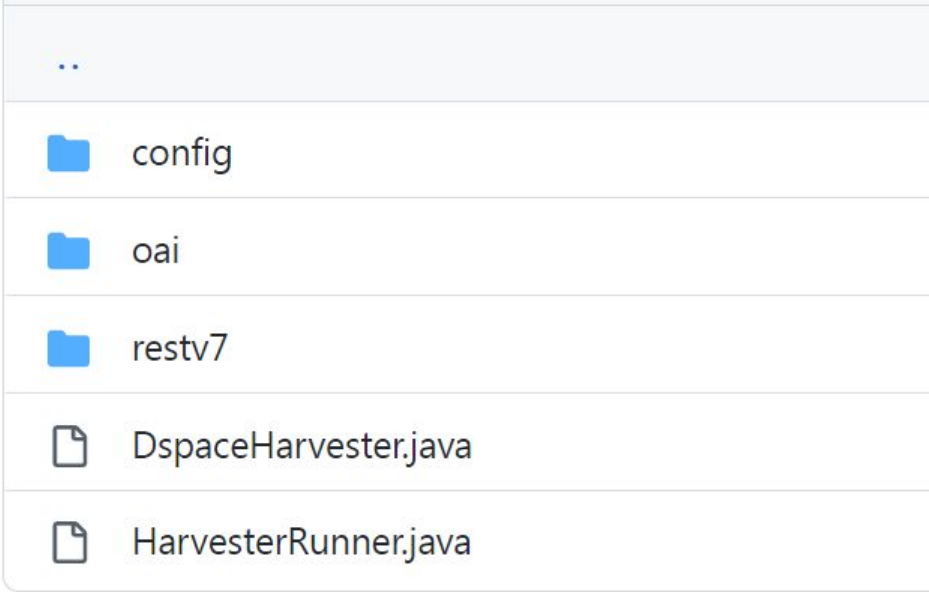
Collapse source

```yaml
openapi: "3.0.2"
info:
  version: 1.1.0
  title: DSpace-VIVO EXchange Data Schema (DVExDS)
  description: |-
    This is the "DSpace-VIVO EXchange Data Schema (DVExDS)" based on the OpenAPI 3.0.2 specification.
    You can find out more about Swagger at [http://swagger.io](http://swagger.io). In the third iteration of the pet store, we've switched to the design first approach!
    You can now help us improve the API whether it's by making changes to the definition itself or to the code.
    That way, with time, we can improve the API in general, and expose some of the new features in OAS3.
    Some useful links:
    - [DSpace-VIVO - Integration project of DSpace metadata with VIVO](https://github.com/vivo-community/DSpace-VIVO)
    - [The Pet Store repository](https://github.com/swagger-api/swagger-petstore)
    - [The source API definition for the Pet Store](https://github.com/swagger-api/swagger-petstore/blob/master/src/main/resources/openapi.yaml)
  termsOfService: http://swagger.io/terms/
  contact:
    name: Michel Héon PhD
    email: vivo@uqam.ca
    url: https://github.com/vivo-community/DSpace-VIVO
  license:
    name: BSD 3-Clause License
    url: http://www.apache.org/licenses/LICENSE-2.0.html
servers:
  - url: /api
tags:
```

https://wiki.lyrasis.org/display/VIVO/DSpace+Communities+taxonomy+for+VIVO

Extraction of data

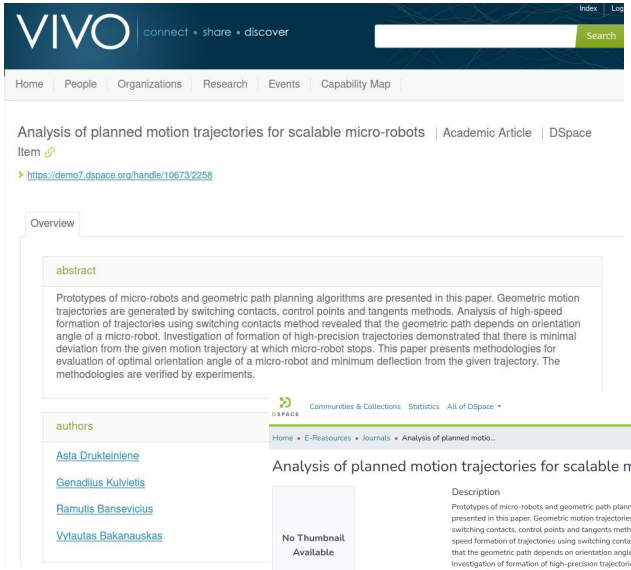
- OAI-PMH
- REST API - DSpace 7



https://github.com/vivo-community/DSpace-VIVO/tree/main/bundles/org.vivoweb.dspacevivo.transformation/src/main/java/org/vivoweb/dspacevivo/transformation/harvester

Transformation & Loading

- Java
- Bash scripts



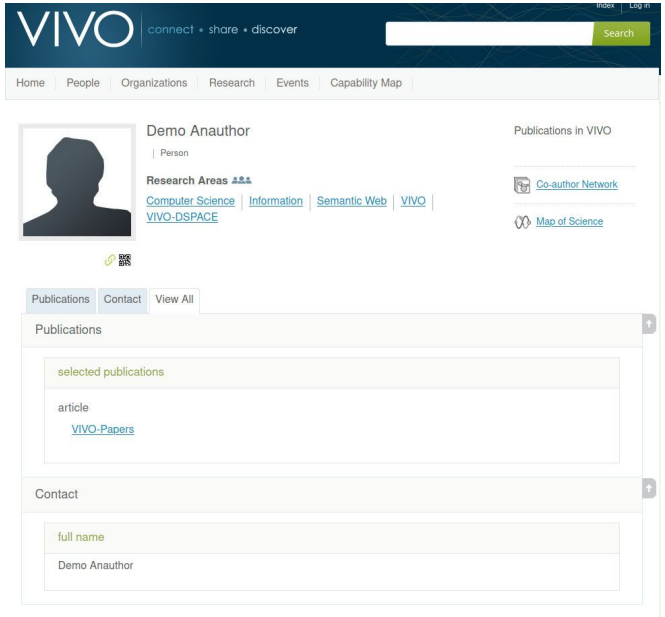

```
@prefix obo:   <http://purl.obolibrary.org/obo/> .
@prefix ro:    <http://purl.obolibrary.org/obo/ro.owl#> .

<https://dspace7.org/resource/10673_8225>
       a                      obo:BFO_0000031 , bibo:Document ,
                              owl:Thing , bibo:Article , obo:BFO_0000001 ,
                              vivo:Dataset , obo:IAO_0000030 , p1:Item , obo:BFO_0000002 ;
       rdfs:label             "VIVO-Papers" ;
       obo:ARG_2000028        <https://dspace7.org/resource/10673_8225-vcard> ;
       dcterms:creator        "Nuon, Tony" , "AnAuthor, Demo" , "jha, suraj" ;
       dcterms:date           "2022-05-09T10:44:07Z" , "2022" ;
       dcterms:subject        "Computer Science, Information, Semantic Web, VIVO, VIVO-DSPACE " ;
       dcterms:title          "VIVO-Papers" ;
       dcterms:type           "Article" ;
       vitro:mostSpecificType bibo:Article , p1:Item ;
       vivo:relatedBy         <http://dspacevivo.vivoweb.org/individual/nuon_tony> ,
              <http://dspacevivo.vivoweb.org/individual/anauthor_demo> ,
              <http://dspacevivo.vivoweb.org/individual/jha_suraj_https_dspace7_org_resource_10673_8225> ,
              <http://dspacevivo.vivoweb.org/individual/jha_suraj> ,
              <http://dspacevivo.vivoweb.org/individual/nuon_tony_https_dspace7_org_resource_10673_8225> ,
              <http://dspacevivo.vivoweb.org/individual/anauthor_demo_https_dspace7_org_resource_10673_8225> ;
       p1:hasBitstream  "https://demo7.dspace.org/bitstreams/20de77fb-acea-4aaf-8fd3-6dbf087da109/download"^^xsd:anyURI ,
                        "https://demo7.dspace.org/bitstreams/65bbb69a-2497-47c3-a8e5-a9b87415a37b/download"^^xsd:anyURI .
```

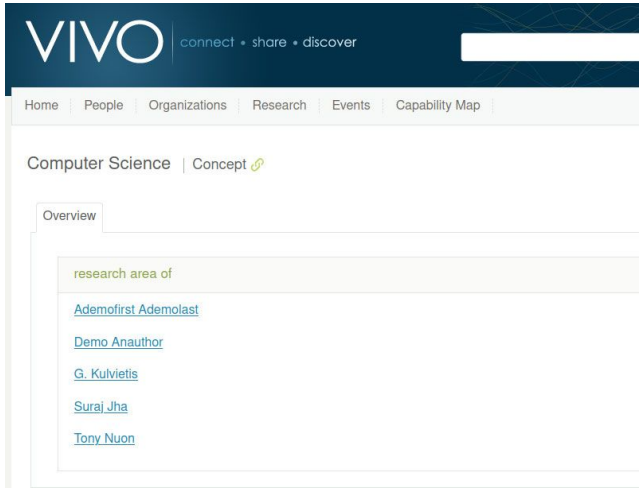
Several Dspace metadata are translated into the VIVO vocabulary:

- The corresponding DSpace Url (Clikable)
- The abstract
- The authors
- The keywords translated in vivo:Concept

Authors -> foaf:Person

- DSpace creators are translated into foaf:Person in VIVO
  - The author's permanent URI is built from a processing on the name String
  - The topics associated with an Item-Keywords are translated into 'Reseach Area' for the item's author - grouping researchers with a common area of expertise



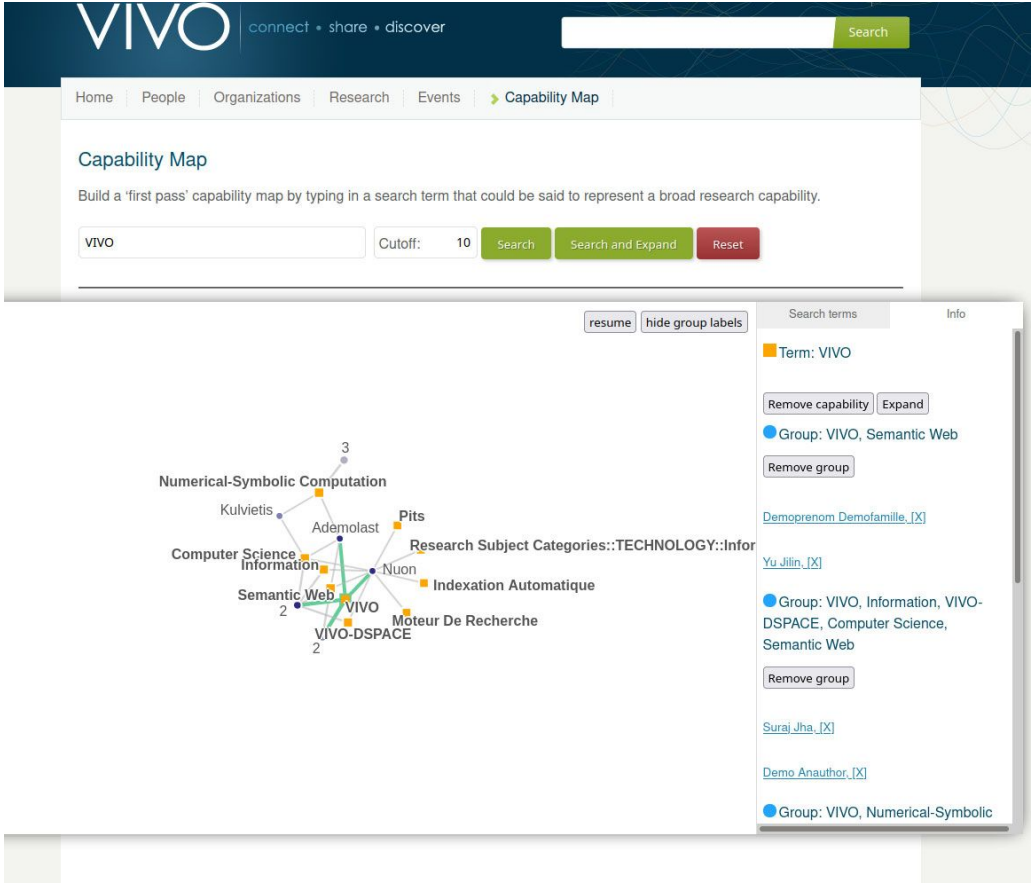


Multiple sources

- DSpace Items can come from several DSpace sources
  - University VIVO with integrated data from Department 1 DSpace, Department 2 DSpace, etc.
- The author's 'Research Area' (expertise) is enriched with publications from various data sources

Capability Map

The Keyword transformation into a concept allows the expertise and experts mapping by crossing the information coming from the multiple DSpace instances



Next phases

- What?
  - Completion of exchange information task
    - Transformation of other entities - repositories, communities, collections
    - Integration of project within VIVO codebase
    - Establishing stable links between DSpace and VIVO items and terms
  - Description of crosswalks/mapping in formal language
    - Adjusting default crosswalks for customized VIVO and DSpace instances
  - Sending notifications to DSpace from VIVO
    - Depositing pdf files from VIVO UI into DSpace
  - Receiving notifications from DSpace
    - Changes in DSpace items reported to VIVO
- When?
  - ASAP
- How?
  - Crowdfunding
- Who?
  - Call for expressing interest for participation in the project

https://wiki.lyrasis.org/display/VIVO/DSpace-VIVO+integration+task+force

Building a REST endpoint for getting and ingesting data from/into VIVO

Background

- At the moment VIVO platform supports SPARQL API endpoint which can be used for getting/ingesting data by using SPARQL select and construct queries
- Problem: semantic web expertise needed
- Solution: the json like REST endpoint for common entities

JSON like REST endpoint

- Dynamic API
  - Simplify VIVO extension
    - Build on top of an ontology
      - Enables VIVO users to develop endpoints in accordance with needs at their institutions
  - Decouple backend and frontend
  - Provide RESTful API to VIVO users
    - Ingestion of data
    - Endpoints for person and organization units are planned to be implemented first

Organizational level

- Two VIVO sprints were organized
  - https://wiki.lyrasis.org/display/VIVO/Sprint+-+Dynamic+API+based+on+an+ontology
  - https://wiki.lyrasis.org/display/VIVO/Sprint+-+Dynamic+API+based+on+an+ontology+2
- A Task force
  - https://wiki.lyrasis.org/display/VIVO/Dynamic+API+Task+Force

Status and plans

- Dynamic API ontology has been defined
- The engine for executing API description has been partially implemented
- The pilot version of a GUI has been started
- Hopefully, the first version of the JSON like endpoint for organization units and persons should be part of VIVO 1.15.0 (2023)
- Extension of expressiveness of dynamic API ontology