

CRIS 2022, Dubrovnik, 12 - 14 May, 2022

# CERIF 2 VIVO mapping

Dragan Ivanovic (University of Novi Sad, Serbia)

Christian Hauschke (TIB, Germany)

Tatiana Walther (TIB, Germany)

Anna Guillaumet (SIGMA Gestión Universitaria, A.I.E, (M.P.))

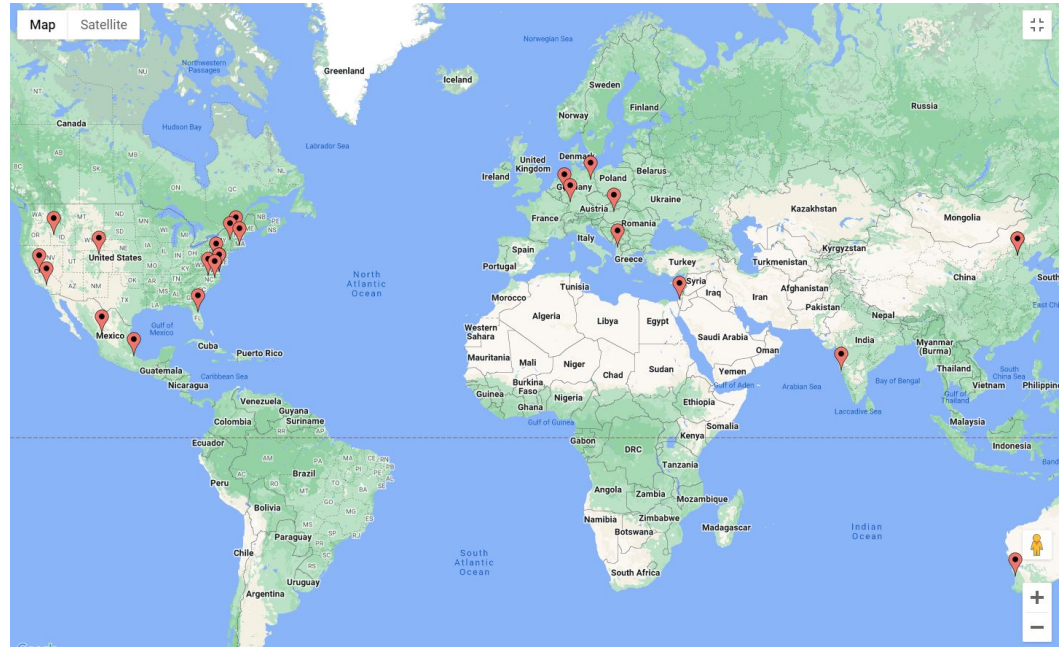
# Agenda

- Background
  - VIVO
  - CERIF
  - MoU - VIVO + euroCRIS
  - Project definition
- Motivation
- Challenges
- Approach
- Progress

# Background

# VIVO

- VIVO [Pronunciation: vee-voh] is member-supported, an ontology and open-source software supporting recording, editing, searching, browsing and visualizing scholarly activity
- At the beginning it was a US community and project, but it is a global community today
  - Join us -  
<https://vivo.lyrasis.org/community/>



# CERIF

- CERIF is considered as a standard interchange format for research domain recommended by the European Union to its Member States
- This model represents information about research domain entities such as research results, researchers, organizations, events, etc.
- More information about CERIF can be found at <https://eurocris.org/services/main-features-cerif>
- Ongoing CERIF related projects
  - OpenAIRE Guideline for CRIS managers
  - CERIF refactoring project

# Memorandum of understanding

- In July 2021 a Memorandum of Understanding between VIVO and euroCRIS was signed
- The euroCRIS and VIVO representatives recognized a shared goal and the need for joint action in integration of research domain information across the world



## Memorandum of Understanding

The purpose of this Memorandum of Understanding (MoU) is to establish and promote a strategic and cooperative partnership between VIVO and euroCRIS.

euroCRIS is a not-for-profit, statutory association ( <https://www.eurocris.org> ) established in 2002, governed by Dutch law and dedicated to the development and implementation of efficient and effective institutional, national and international research information systems and their interoperability, based on CERIF (Common European Research Information Format). One of euroCRIS's main objectives is the promotion of cooperation and exchange of expertise between

# Mapping project

- One specific action started after the signing of the MoU is definition of mapping between the CERIF data model and the VIVO Ontology
- The first attempt of definition mapping happened in 2013, but it is outdated because both formats evolved meanwhile
- The representative of both communities (VIVO and euroCRIS) should work on the project
- Expertise of both communities is needed
- Different perspectives should be taken into account
- No strict deadline for the project

# Motivation



# Benefits of the mapping

- Interoperability between VIVO platforms and CERIF-compatible CRIS systems
- Knowledge transfer
  - Improvement of CERIF model and VIVO ontology by analyzing the other side
    - Extensions of the data models
    - Addition of descriptions and annotations
  - Machine-executable mapping for various purposes and in various notations, for example for a CERIF-compliant data export from VIVO

# Challenges

# European vs USA point of view

- CERIF started as a European format, meaning it is aligned with European academic landscape
- VIVO started with a focus on the USA, meaning its ontology is aligned with USA academic landscape
- For instance:
  - European academy recognizes project which might or might not be funded by third parties, while USA academy recognizes only grant/funding
  - Organizational types (especially scholarly organizations) differ widely between countries (colleges are US-specific, Lehrstühle only exist in German-speaking countries)
  - Different qualification procedures require different types of documents, e.g. the habilitation thesis is only required in a few countries.

# Data model vs Ontology point of view

- Properties (“attributes”) are owned by the entity (“class”) that contains them
- Associations represent both sides of a relationship.
- Closed World Assumption
- Properties are independent of classes, they just combine with them through the `rdfs:domain` and `rdfs:range` statements.
- An object property represents one side of a relationship, the other side is represented by an inverse object property.
- Open World Assumption
  - URI

# CERIF vs VIVO

- XML scheme? Available Formats?
- Entities and attributes from one and the same data model
- Types of relationships? 1:n?
  - But also there are linked entities
- Standard vocabularies with sub-types of entities (e.g. organization types)
- An OWL ontology
- Uses several integrated OWL ontologies (FOAF, BIBO, OBO, Vcard, ...)
- Apart from the simple relations, uses complex constructs with context nodes, e.g. person -> authorship -> publication
- Categorization of entities through sub-classes

# Approach

# Collaboration

- CERIF task group independent meeting
- VIVO ontology group independent meeting
- Periodical virtual joint meeting of those two groups
- Anyone is welcome to join us on any of those side
  - the invitation to join the initiative was shared through the complete VIVO and euroCRIS community.

# Definition of matchings

- In the first phase, mappings will be presented in a free form by using tables with predefined list of columns: CERIF entity to VIVO class, CERIF data type to VIVO data type, each qualified with mapping expression (equivalent, constraint, similar, shoehorn).
- Equivalent - the target concept has the same semantics as the source
  - e.g. cfPers -> foaf:Person
- Similar - the target and source concept have similar semantics,
  - e.g. cfProj.cfAcro - > vivo:Project.vivo:vivo:abbreviation
- Constraint - the target concept is restriction of the source concept,
  - e.g. cfResProdDescr -> bibo:abstract
- Shoehorn - the target concept is broader than the source,
  - e.g. cfOrgUnit\_Class.cfStartDate -> foaf:Organization.vivo:dateTimeInterval



# Machine-executable mapping

- Existing Cerif2VIVO and VIVO2Cerif XSLT and Java-based translator - <https://github.com/ieru/cerif2vivo>
- Available notations for formal representation of mapping have to be investigated and selected
- The matching from the previous step will be expressed in the selected notation

# Versioning of the mapping

- The joint group (VIVO and euroCRIS) is also planning to work on the definition of an infrastructure for further maintenance of the mapping and for keeping the history of those mappings
- The mapping will be expressed in the selected formal language for definition of machine-executable mapping and made available through the aforementioned infrastructure
- After some changes in CERIF data model or VIVO ontology, the formal representation of the mapping will be updated, and the history of the mapping will be also available

# Implementation of a tool/plugin

- The last phase will be exploitation of those mapping through implementation of a software module capable to read and execute mapping rules presented in the selected notation and to perform automatic transformation of CERIF records to VIVO records
- So far, in this action, the mapping from CERIF to VIVO has been carried out, but the opposite direction, i.e. the mapping from VIVO to CERIF is also to be taken into account for a next phase of this joint work

Progress

# Definition of the matching

- Matching entities
  - cfProj -> vivo:Project
  - cfPers -> foaf:Person
  - cfOrgUnit -> foaf:Organization
  - cfResPubl -> bibo:Document
  - cfResProduct ->  
obo:ERO\_0000071
  - cfResPatent -> vivo:Patent
  - cfFacil -> vivo:Facility
  - cfEquip -> vivo:Equipment
  - cfSrv -> obo:ERO\_0000005
  - cfEvent -> event:Event
  - cfCountry -> vivo:Country
  - cfFund -> vivo:Grant

# Definition of the matching

- Elements for matching
  - CERIF attributes
  - Multilingual entities
  - Link Entities
  - Classifications (vocabularies)
  
- NOTE: we are matching only in one direction CERIF -> VIVO

# Status of matching

- We are still in the first phase of the project
- Our progress can be found at

[https://docs.google.com/spreadsheets/d/1yauv3iUZe-o9OyTcyn9TI8cGIWoCL\\_IOKklxpmYRug/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1yauv3iUZe-o9OyTcyn9TI8cGIWoCL_IOKklxpmYRug/edit?usp=sharing)