

(Some thoughts on...)

VIVO Data Integration at TIB



Tools, workflows & connections with VIVO
(as a CRIS and a RIM)

EuroCRIS session at VIVO Conference 2019

(Podgorica, Montenegro)

Christian Hauschke & Tatiana Walther (TIB)



Some background...

- TIB is an emerging research institute,
 - high topic diversity
 - lots of grey publications
- No sufficient single data sources for any kind of data (ORCID, citation databases like Dimensions, open access repository etc.)
- A system for research administration:
 - Yearly report
 - Reporting to Leibniz Association
- A system for researchers:
 - showcasing research
 - enabling discovery of TIB's research output

Prerequisite: Advanced Role Management

- need for more sophisticated / specialized roles
- one of the ...

*Four Horsemen
of the
VIVocalypse!*



Publications

- High diversity of researchers with some well-known problems regarding citation database coverage
 - ◆ non-english publications
 - ◆ high diversity of subjects, including publications from social sciences and humanities
 - ◆ publications without DOI / Pubmed-ID
- TIB makes use of the self-editor role
- DOI and Pubmed ID claiming tool
- Publication editor group -> notification and import workflows and responsibilities distributed across several departments

Publication editor group

Group consists of members of Publication Service, Controlling, and Open Science Lab

Workflow

1. Self-reporting by authors in VIVO or internal wiki
 - import of all publications in VIVO
 - if author profile(s) exists, publication will be connected to it
2. First approval (part of professional publication activity, no violence of laws etc.)
3. Revision and enrichment (e.g. peer review and open access status)
4. Curation of either imported via DOI or otherwise ingested metadata by librarians
 - > TIB-affiliation, publication date, class, authors and editors, publication venue etc.
5. Completion / correction of metadata in the GUI or in OpenRefine Project
6. Publication has status “Validated” and can be considered for reporting

Example of KDSF-driven enrichment

The Effect of Gender in the Publication Patterns in Mathematics Academic Article 

<https://doi.org/10.1371/journal.pone.0165367> 

Overview

Identity

Additional Document Info

Kdsf

View All

access

Open Access

DOI and Pubmed ID Claiming (Self-Editor)

Claim publications by

DOI

PubMed ID

Suggested Works

- Import of publications with DOI or PubMed ID by researchers
- “Suggested works” show publications where one author matches the self-editors label

Challenges and opportunities of using VIVO as a reporting tool

- ☒ Hauschke, C ; *Triggs, G* ; ☐ Walther, T
- ☐ Unlisted Author
- ☐ Editor
- ☐ This is not my work

Enter DOI:

You may enter one or more DOIs to match, and can be entered either as an ID or URL:

e.g.

ID: 10.1038/nature01234

URL: <https://doi.org/10.1038/nature01234>

Currently, DOIs issued by Crossref, DataCite and mEDRA are supported.

Each DOI should be separated by a comma or new line.

10.21105/joss.01182

<https://doi.org/10.31263/voebm.v72i2.2808>

DOI: 10.31263/voebm.v7i2.2808

Academic Article



Open Science und die Bibliothek – Aktionsfelder und Berufsbild

Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 2019; 72 (2)

☐ Blümel, I ; ☐ Drees, B ; ☒ Hauschke, C ; ☐ Heller, L ; ☐ Tullney, M

☐ Unlisted Author

☐ Editor

☐ This is not my work

DOI: 10.21105/joss.01182

Academic Article



VIVO: a system for research discovery

Journal of Open Source Software 2019; 4 (39) :1182

☐ Conlon, M ; ☐ Woods, A ; ☐ Triggs, G ; ☐ O'Flinn, R ; ☐ Javed, M ; ☐ Blake, J ; ☐ Gross, B ; ☐ Ahmad, Q ; ☐ Ali, S ;

☐ Barber, M ; ☐ Elsborg, D ; ☐ Fofack, K ; ☒ Hauschke, C ; ☐ Ilik, V ; ☐ Khan, H ; ☐ Lawless, T ; ☐ Levernier, J ; ☐

Lowe, B ; ☐ Martin, J ; ☐ McKay, S ; ☐ Porter, S ; ☐ Walther, T ; ☐ White, M ; ☐ Wolff, S ; ☐ Younes, R

☐ Unlisted Author

☐ Editor

☐ This is not my work

Confirm

Data management with OpenRefine

- OpenRefine - an open source spreadsheet tool for data management
- Normalizing, filtering, cleaning, transforming data
- Fetching data from other Web sources by means of URI / URL
- RDF-Extension -> Use of ontologies and alignment to VIVO data model
- Reconciliation to VIVO and other data sources, e. g. lobid GND data
- GREL (General Refine Expression Language) -> allows definition of special conditions and rules for each column
- Data export -> various formats (CSV, TSV, Excel, RDF, etc.)
- Installation on server -> collaborative data management and automatization of workflows

Data management with OpenRefine

Facet / Filter

Undo / Redo 465

Refresh

Reset All

Remove All

DOI

change invert reset

1 choices

Sort by: name count

true 7

exclude

Facet by choice counts

Publikationsdatum

2019

case sensitive regular expression

6 matching records (415 total)

Extensions: Freebase RDF

Show as: rows records

Show: 5 10 25 50 records

« first < previous 1 - 6 next > last »

All	Publikationstyp	VIVO-Klasse	VIVO-URI	Autor/innen 1	Autor/innen 2	Autor/innen 3	Autor/innen 4	Autor/innen 5
126.	Konferenz-Paper	http://vivoweb.org/ontology/core#ConferencePaper		Plank, Margret Choose new match	Drees, Bastian Choose new match	Hauschke, Christian Choose new match	Kraft, Angelina Choose new match	Leinweber, Christian Choose new match
127.	Ausstellungskatalog	http://purl.org/ontology/bibo/Book						
128.	Konferenz-Poster	http://vivoweb.org/ontology/core#ConferencePoster						
129.	Eprint	http://purl.org/ontology/bibo/Article						
131.	Beitrag	http://purl.org/ontology/bibo/Chapter						
132.	Blog Posting	http://vivoweb.org/ontology/core#BlogPosting						

Reconcile column "Autor/innen 3"

Freebase Query-based Reconciliation

Sindice

VIVO Reconciliation Service

GND reconciliation for OpenRefine

Reconcile each cell to an entity of one of these types:

DocumentPart

Agent

AcademicArticle

BFO_0000004

tib-vivo:Local

Person

tib-vivo:TIBMitarbeiter

Slideshow

core:Concept

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Also use relevant details from other columns:

Column	Include?	As Property
Publikationstyp	<input type="checkbox"/>	
VIVO-Klasse	<input type="checkbox"/>	
VIVO-URI	<input type="checkbox"/>	
Autor/innen 1	<input type="checkbox"/>	
Autor/innen 2	<input type="checkbox"/>	
Autor/innen 4	<input type="checkbox"/>	
Autor/innen 5	<input type="checkbox"/>	
Autor/innen 6	<input type="checkbox"/>	
Autor/innen 7	<input type="checkbox"/>	
Autor/innen 8	<input type="checkbox"/>	
Autor/innen 9	<input type="checkbox"/>	
Herausgeber	<input type="checkbox"/>	

Add Standard Service...

Add Namespaced Service...

Start Reconciling

Cancel

Data management with OpenRefine

RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: <https://vivo.tib.eu/fis/individual/n> [edit](#)

RDF Skeleton

RDF Preview

Available Prefixes: [kdsf-vivo](#) [rdf](#) [owl](#) [xsd](#) [rdfs](#) [bibo](#) [vivo](#) [vcard](#) [obo](#) [foaf](#) [tib-vivo](#) [+add prefix](#) [manage prefixes](#)

(row index) URI

[add rdf.type](#)

[x](#) [->rdf.type->](#)

[x](#) [->rdfs.label->](#)

[x](#) [->vivo.relatedBy->](#)

VIVO-Klasse URI

[add rdf.type](#)

[x](#) [->rdfs.label->](#)

(row index) URI

[x](#) [vivo:Authorship](#)

[add rdf.type](#)

[x](#) [->vivo.relates->](#)

[x](#) [->vivo.relates->](#)

Autor/innen 1 URI

[add rdf.type](#)

Autor/innen 1 URI

[x](#) [vcard:Individual](#)

[add rdf.type](#)

[x](#) [->vivo.rank->](#)

1

[Add another root node](#)

RDF Node

Use content from cell...

- ☐ (row index)
- ☐ Publikationstyp
- ☐ VIVO-Klasse
- ☐ VIVO-URI
- ☒ Autor/innen 1
- ☐ Autor/innen 2
- ☐ Autor/innen 3
- ☐ Autor/innen 4
- ☐ Autor/innen 5
- ☐ Autor/innen 6
- ☐ Autor/innen 7
- ☐ Autor/innen 8
- ☐ Autor/innen 9
- ☐ Herausgeber
- ☐ Sonstige Beteiligte
- ☐ Titel

The cell's content is used ...

- ☒ as a URI
- ☐ as text
- ☐ as language-tagged text
- ☐ as integer number
- ☐ as non-integer number
- ☐ as date (YYYY-MM-DD)
- ☐ as dateTime (YYYY-MM-DD HH.MM.SS)
- ☐ as boolean
- ☐ as custom datatype (specify type URI)
- ☐ as a blank node

[Use custom expression...](#)

```
forNonBlank(cells["Autor/innen  
1"].value,v,cell.recon.match.id,  
(if(isBlank(cells["Autor/innen 1"].value),v,""))  
preview/edit
```

[OK](#) [Cancel](#)

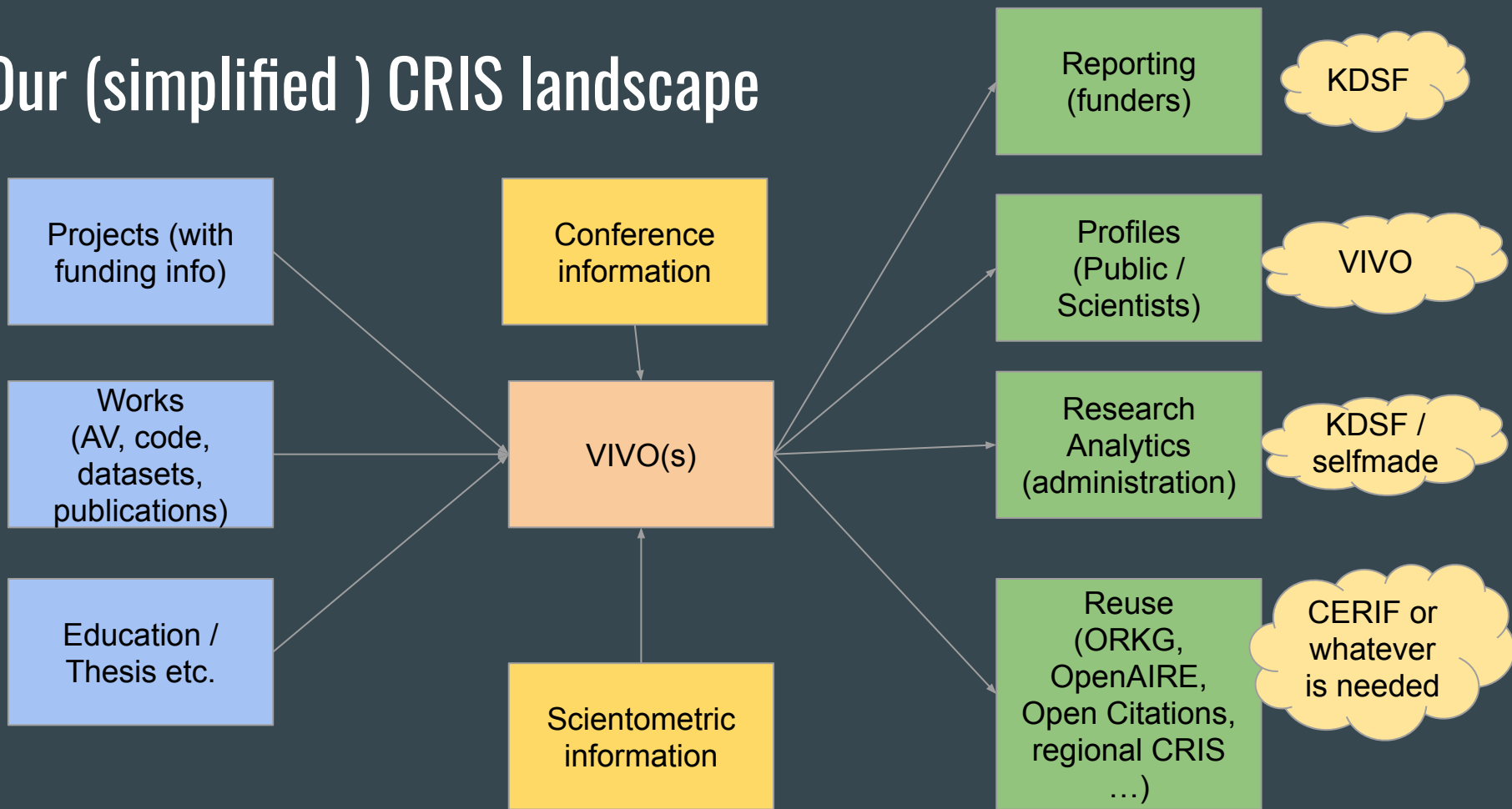
Projects and Third Party Funding

- Concept of grants vs projects
- German funding specialties -> extension with additional categories, e.g. project call (Förderlinie)
- Special rights and roles in VIVO -> research project editor
- Data source at the TIB - a collaboratively edited Excel sheet
- A **uniform, standardized input of information** for further script operations is required :
 - ◆ Names of funders, organisational units, project calls, etc.
 - ◆ Dates
 - ◆ Unique internal IDs serve as a basis for URIs in VIVO
- Python libraries:
 - ◆ [csvkit](#)
 - ◆ [VIVOPump](#)
 - ◆ [VIVO-rdflib-sparqlstore](#)

Projects and Third Party Funding - Workflow

1. Inserting project information by responsible staff members (e. g. research project leads) in Excel Sheet
2. Extraction and filtering of the required information, conversion into csv with Python library csvkit
3. Normalizing headers, merging data, assigning URIs with own python scripts
4. Conversion csv into rdf, aligning to the VIVO data model by means of VIVOPump
5. Automatized ingest and update of project data in VIVO by means of VIVO-rdflib-sparqlstore
6. Data check in VIVO by research project editors
7. Changes (except for deletion of a project) in the Excel sheet

Our (simplified) CRIS landscape



Is VIVO a CRIS, a RIMS or something else at TIB?

- VIVO at TIB is not used on top of a monolithic (Sebastian Herwig) CRIS system.
- But is it a **system** with **current information** about **research**?

Yes.

- Is it a **system** for **research information management**?

Yes.

What about research analytics?

- common task of CRIS
- another *Horseman of the VIVOcalypse!*
- Visualizations in customized dashboards
 - ◆ Presentation by Qazi Asim Ijaz Ahmad:
“Curated Data Visualization and Reporting from VIVO”
Friday, 11 AM, Crystal Ball B
- Reporting via Vitro Query Tool
 - ◆ Presentation by Graham Triggs:
“A Library of Queries and Reports -
introducing the Vitro Query Tool”
Friday, 15:30, Crystal Ball A
- Reference implementation for Open
Scientometric Indicators (Poster)



More information?

Thanks for your attention! :-)

Let's talk now...

... or in the breaks!

<https://vivo.tib.eu/fis>

(Tatiana & Christian)

Image credits and licenses

- The Four Horsemen of the Apocalypse: The Rider on the Pale Horse George Frederic Watts (1817–1904) Walker Art Gallery Public Domain according to https://commons.wikimedia.org/wiki/File:The_Rider_on_the_Pale_Horse_-_George_Frederic_Watts.jpg
- “The Four Horsemen of the Apocalypse: The Rider on the White Horse”, c.1874–1883. Walker Art Gallery, Liverpool. Medium oil on canvas
Public domain according to <https://commons.wikimedia.org/wiki/File:Horseman-of-the-apocalypse-george-frederick-watts.jpg>.