OER Search Index for international repositories

Axel Klinger (TIB)
Hannover, Germany

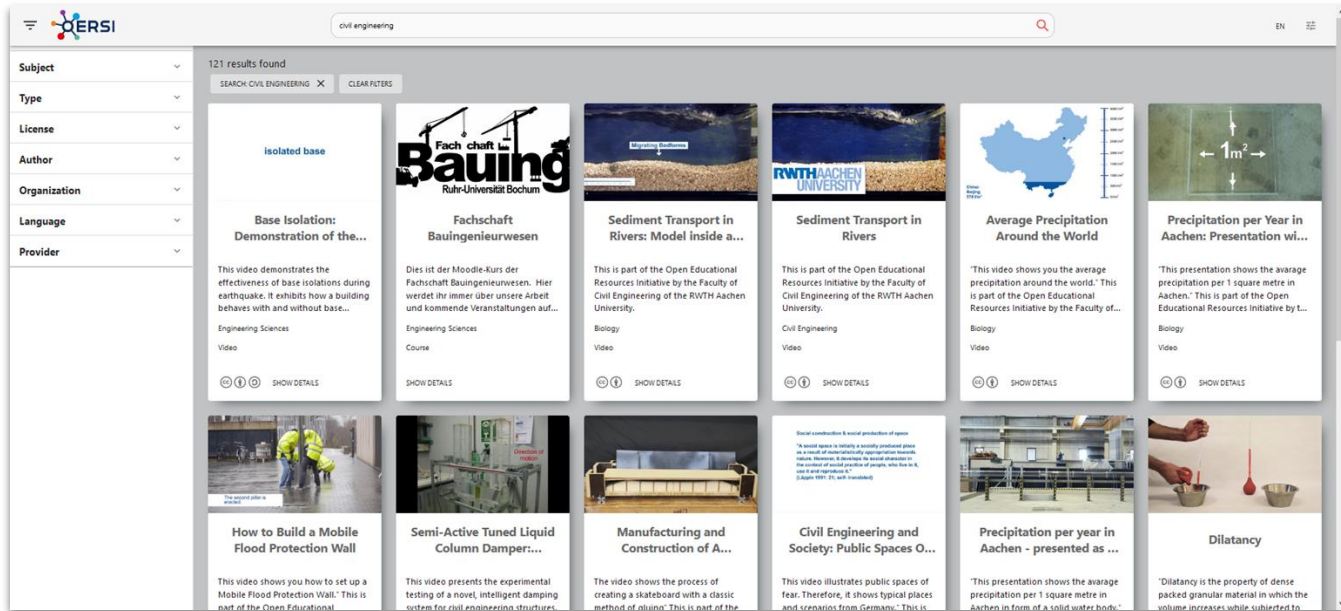OE Global Conference, 23rd May 2022

Agenda

- What is OERSI
- Why we need OERSI
- How to extend OERSI
- Internationalisation
- GitLab for OER
- Contribution

What is OERSI

- Open Source search index for distributed heterogeneous OER repositories in higher education https://oersi.org

What is OERSI

- Technology Stack:
  - ElasticSearch backend
  - ReactiveSearch frontend
  - Metafacture / Python import
- Open Source on GitLab
  - https://gitlab.com/oersi
- Cooperative project
  - North Rhine-Westphalia University Library Center (hbz)
  - Technische Informationsbibliothek (TIB)

Why we need OERSI

- **increasing number of regional or institutional repositories**
  results in more effort for teachers and learners to search in different places
- **need for a single point of entry**
  to reduce the effort of finding OER
- **get more relevant results**
  by summarization of content

Example repository - twillo

- regional OER Portal for Lower Saxony, Germany
- funded by the Ministry of Science and Culture
- operated by Technische Informationsbibliothek (TIB)
- based on edu-sharing software
- https://www.twillo.de
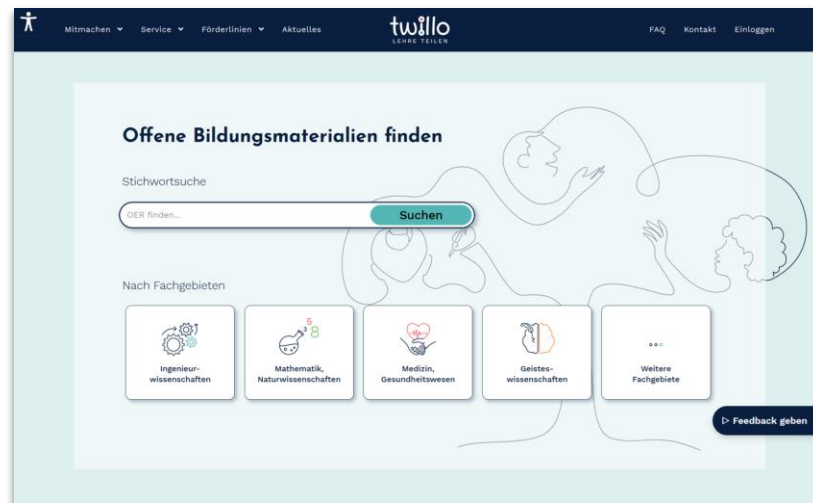
Regional OER Repositories for Higher Education in Germany

- OpenLearnWare (TU Darmstadt)
- HHU Mediathek (HHU Düsseldorf)
- ZOERR (Baden-Württemberg)
- twillo (Lower Saxony)
- HOOU (Hamburg)
- vhb (Bavaria)
- ORCA (North Rhine-Westphalia)
- …

Why we need OERSI - current approach

- create a network of OER repositories
- **bidirectional connections -> benefit for each repository**
- **integrate search into existing repositories or websites**
- selective search in OERSI

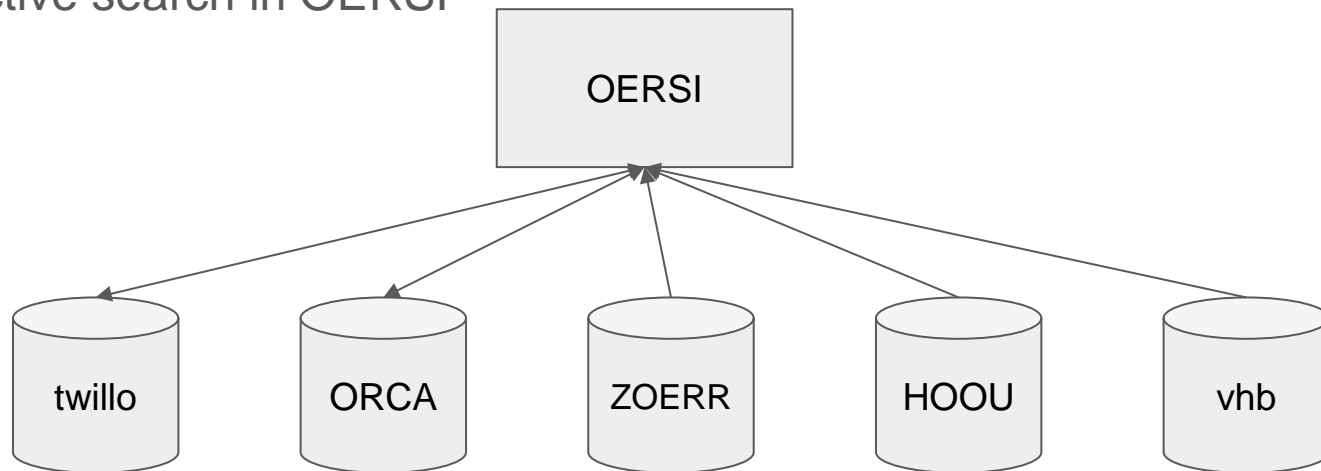# How to extend OERSI

- **Lightweight interface for the connection**
  of further repositories
- **Modular ETL import process**
  for extraction, transformation and loading
  of metadata based on Metafacture framework
- **optional import processes**
  via Python or other scripting languages
  repositories should ideally have an API or a sitemap
- **simply suggest further sources**
  via https://oersi.org/resources/services/contact

Internationalisation - the Idea

- support cross-lingual education
  by finding related resources for topics in other languages
- connect international repositories
  and enable cross-lingual search
- for international students to learn
  and teachers to support them

The idea is, instead of translating the content, to map to similar content
in another language by the **mapping** of **keywords** from **WikiData**

Internationalisation - current state

- translation of the **user interface** in de, en, uk
- translation of the **vocabulary for subjects and resource types** in de, en, uk
- mapping of **keywords** to de, en, fr, es, nl, dk, uk
  - regular updates
  - coverage of about 60-70% for more widespread languages
  - coverage of about 45-55% for more regional languages
- minimize time to translate UI and vocabulary
  into another language by pre-filling the translation
- vocabulary is open source
  - https://github.com/dini-ag-kim/hochschulfaechersystematik
  - https://github.com/dini-ag-kim/hcrt

Internationalisation - example

- search for: climate change (en), Klimawandel (de), зміна клімату (uk), klimaatverandering (nl), cambio climático (es), klimaændring (dk)
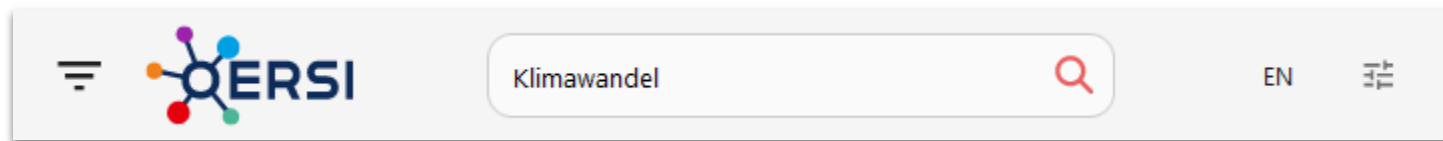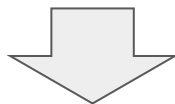
262 results in different languages

Language	
English	187
German	55
Undefined	16
Spanish	4
French	1

GitLab as an OER Repository

- **GitLab is an open source version control system**
  especially for text documents like source code, Markdown or LaTeX
- **collaboration platform**
  for open and closed projects
- **can be self-hosted**
  by institutions or alliances of institutions
- **great features for automation (enrichment, transformation, …)**
  of text processing pipelines -> easy to share as templates
- **for students and teachers**
  in thesis, projects, textbooks, interactive courses

Examples from GitLab

- write Markdown and generate several output formats
  e.g. HTML, PDF, EPUB

https://gitlab.com/axel-klinger/studienarbeit

Examples from GitLab

- … or dynamic websites like LiaScript
  past URL to RAW-Markdown file into LiaScript https://liascript.github.io/

GitLab Metadata

- simple form as GitLab page generates consistent YAML metadata

https://gitlab.com/oersi/metadata-form

```yaml
---
name: "Untersuchung eines Fassadenelements mit zweifarbiger Beschich
creator:
  - givenName: "Leif-Arne"
    familyName: "Peterson"
  - givenName: "Axel"
    familyName: "Klinger"
publisher:
  givenName: "Axel"
  familyName: "Klinger"
keywords:
  - "Solare Wärmegewinne"
  - "passive Fassade"
description: "Nachbildung einer Studienarbeit in Markdown"
inLanguage: "de"
funder: "Universität Hannover"
subjectOf: "Bauingenieurwesen"
version: 1.0
license: "https://creativecommons.org/licenses/by-sa/4.0/deed.de"
thumbnailUrl: http://placehold.it/100x100
```

How to contribute

- Contact us for questions, feedback or suggestions
  https://oersi.org/resources/services/contact
- Rise an issue for feature request or bug report
  https://gitlab.com/groups/oersi/-/issues
- Feel free to participate in the development process
  https://gitlab.com/oersi/oersi-setup/-/blob/master/CONTRIBUTING.md
- Integrate OERSI in your own platform
  https://oersi.org/resources/services/contact (API in progress)
- Install your own index for a regional solution
  https://gitlab.com/oersi/oersi-setup

Conclusion

OERSI + GitLab
offer a pragmatic solution for an
open, reusable, sustainable and scalable
international infrastructure for OER

Thanks for your attention!

- Contact: axel.klinger@tib.eu
- https://oersi.org

**… and if you know about a GitLab instance or a regional OERSI
to connect, please tell us https://oersi.org/resources/services/contact**