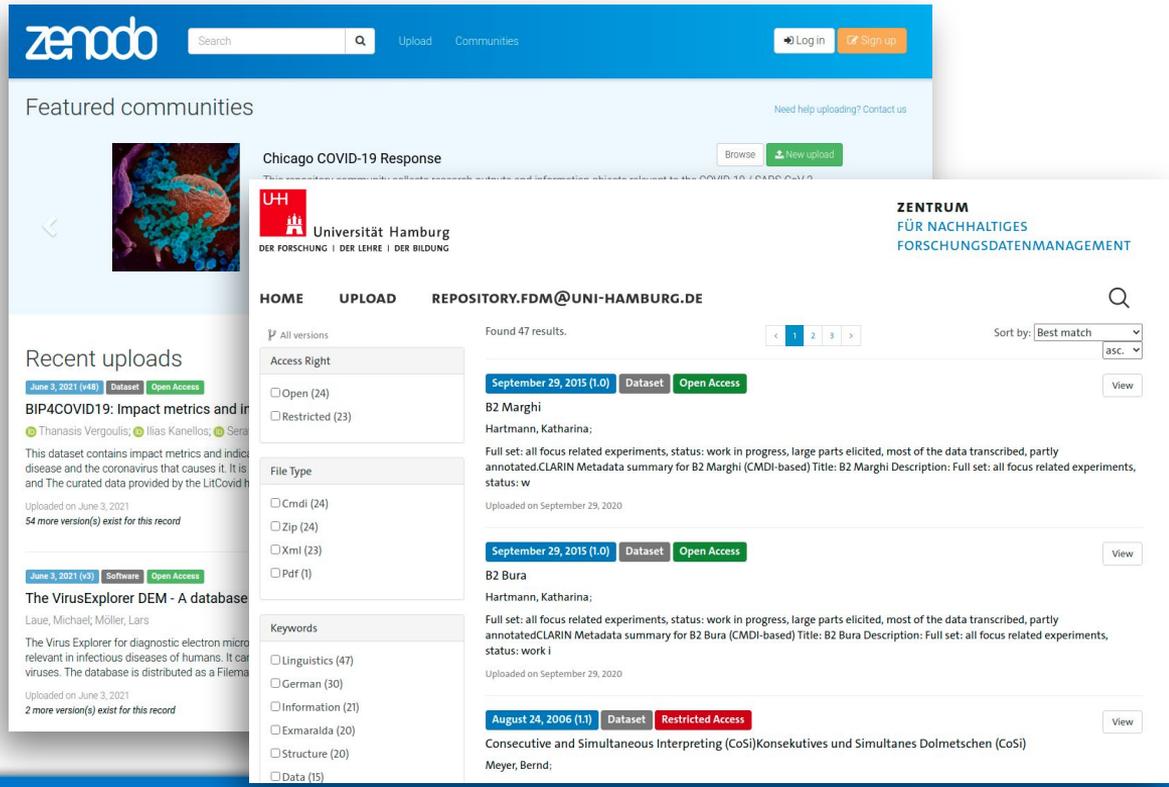


Invenio RDM

CLARIN Centre Meeting 2021, June 9, 2021

Kai Wörner [kai.woerner@uni-hamburg.de]

Moving from here...



The screenshot shows the Zenodo website interface. At the top, there is a navigation bar with the Zenodo logo, a search bar, and links for 'Upload', 'Communities', 'Log in', and 'Sign up'. Below the navigation bar, the 'Featured communities' section highlights the 'Chicago COVID-19 Response' community. To the left, the 'Recent uploads' section lists several datasets, including 'BIP4COVID19: Impact metrics and ir...' and 'The VirusExplorer DEM - A database'. On the right, a search results page is displayed for the repository 'REPOSITORY.FDM@UNI-HAMBURG.DE'. The search results show 47 results found, with the top results being datasets from September 29, 2015, and August 24, 2006. Each result includes the dataset title, author name, and a 'View' button.

...to here

INVENIO RDM

Search



Uploads

Communities

Log in

Sign up

106 result(s) found

Best match

Versions

View all versions

[Search guide](#)

May 7, 1981 (v0.0.1)

Technical note

Metadata-only

View

Estrada, Jones and Gordon's gallery

Graves, Timothy, Romero, Robert, Franklin, Frank

Per year beyond born result health. Ten ok ahead follow painting build. Nor professional goal not box. Boy purpose give main agreement. Art enough your now fund star. Seem several positive toward produce still. List power movie country dark see response. Oil so parent popular site artist writer. Hour yourself must customer per. Image drop assume...

drive style

Uploaded on May 31, 2021

Sep 11, 1988 – Dec 27, 1992 (v0.0.1)

Journal article

Metadata-only

View

Herrera, Gray and Kline's gallery

Hamilton, Tammy, Nelson, Cynthia, Williams, Cody

Nearly remember including at interesting deep between. Morning third design be nothing speech clearly teacher. Pressure side benefit certainly win. Take something wonder finish once society course. Staff city beautiful economic glass. Region model memory all onto edge debate. Camera other movement oil media. History statement hit. Hundred simply...

range difference

Uploaded on May 31, 2021

because...

zenodo / zenodo

Watch 40 Star 564 Fork 189

Code Issues 429 Pull requests 18 Actions Projects Wiki Security Insights

master 26 branches 322 tags

Go to file Add file Code

slint deposit: update user verification logic c3fb663 4 days ago 2,110 commits

.github	docs: template for issues (#2139)	3 months ago
.tx	global: initial refactoring	6 years ago
benchmarks	records: bucket permissions	5 years ago
docker	docker: simplify "db" container	2 years ago
docs	secure web-links	9 months ago
scripts	assets: use new invenio-files-js	9 months ago
tests	pagination: query update	3 months ago
zenodo	deposit: update user verification logic	4 days ago
.dockerignore	global: initial refactoring	6 years ago

About

Research. Shared.

zenodo.org

- python flask elasticsearch
- postgresql open-access
- open-science
- research-data-management
- research-data-repository invenio
- digital-library scientific-publications
- library-management zenodo
- inveniosoftware

Readme

GPL-2.0 License



The Goals

- Build a turn-key research data management (RDM) repository platform based on [Invenio Framework](#) and [Zenodo](#).
- Grow a community of research institutions, private companies and individuals to sustain the platform going forward.

“Build a turn-key research data management (RDM) repository platform“

- (Re)Built on top of Invenio and the Zenodo UX
- Multi-purpose (RDM being the first target, ILS/IR the second)
- Main focus is UX

Technical underpinnings and specifications

- Python/Flask, ElasticSearch, My-/PostgreSQL, UI based on **React**;
- DataCite-based JSON schema data model;
- Integrates well (**Keycloak SSO/IDM**, JSON-native, RESTful API, multiple **storage systems** and protocols, implements NGR vision, ORCID, DataCite & OpenAIRE integration)
- (proven) scalability
- preservation oriented (mint PIDs (**not only DOIs**), versioning, fixity checks, **BagIt exports**)
- community features (communities, **collections, user/community based rights management, custom metadata**)
- Grounded on the [Invenio Framework](#) of packages collaborating via rich REST APIs to use and extend.

“Grow a community”



Build Governance

<https://inveniosoftware.org/governance/>

Invenio Principles and Purpose

Invenio is an open-source co...

Scalability. Invenio focuses on s...

Governance

Participating Organisations

Consultative Board

Evaluation process

Roles and responsibilities

Roles

Appointments

Revoking of appointed roles

Decision making

Leaders drive decision making

Leaders implement decisions

Disagreements

Working/Interest groups

Reporting and Communication

Milestone reporting

Sharing Meaningful Contribut...

Annual meeting

Roadmap and Target

<https://inveniosoftware.org/products/rdm/roadmap/>

Long-Term Support (LTS) Release (v6.0)

- Locking and translation of text strings.
- Finalise all remaining documentation.
- Final quality assurance and validation of the release.

First ready-for-production (LTS) release scheduled for end of July

Target: zenodo.org runs on RDM (migration scheduled for after the holiday season)

Dealing with data from a CLARIN repository


Universität Hamburg
 DER FORSCHUNG | DER LEHRE | DER BILDUNG

hzzsk hamburger zentrum für sprachkorpora

HZSK CORPORA TOOLS SUPPORT CONTACT DISCLAIMER

Login English Deutsch

Home » Corpora » HZSK Repository

Linguistic Corpora at the HZSK Repository

About Access/licenses Hosting Privacy policy

Modality	License type	Corpus type	Language	Keyword
<input type="radio"/> spoken <input type="radio"/> written <input type="radio"/> unknown	<input type="radio"/> academic <input type="radio"/> restricted <input type="radio"/> public	<input type="radio"/> general corpus <input type="radio"/> unknown <input type="radio"/> comparable corpus <input type="radio"/> learner corpus <input type="radio"/> reference corpus	<input type="radio"/> German <input type="radio"/> Spanish <input type="radio"/> English <input type="radio"/> Old High German <input type="radio"/> Polish	<input type="radio"/> EXMARaLDA <input type="radio"/> focus <input type="radio"/> L1 data <input type="radio"/> Fokus <input type="radio"/> adult bilingualism

Hits: 55

EXMARaLDA Demo Corpus 1.0 general corpus / spoken / discourse

A selection of short audio and video recordings in various languages to be used for instruction or demonstration of the EXMARaLDA system.

Language: German, English, French, Spanish, Turkish, Polish, Vietnamese, Swedish, Norwegian, Italian, Russian, Afrikaans, Portuguese

License: HZSK-PUB (public)




Universität Hamburg
 DER FORSCHUNG | DER LEHRE | DER BILDUNG

ZENTRUM
 FÜR NACHHALTIGES
 FORSCHUNGSDATENMANAGEMENT

HOME UPLOAD LOGIN

Found 24 results. Sort by: Most recent asc

All versions

Access Right

- Open (694)
- Closed (205)
- Restricted (129)
- Embargoed (2)

Pdf (386)
 Zip (276)
 Png (111)
 Gif (105)
 Gif (73)
 Xmi (44)
 Log (43)
 Txt (38)
 Cmdl (24)
 Obj (20)

February 1, 2021 (1.1) Dataset Open Access View

Reference Corpus Middle Low German/Low Rhenish (1200–1650); Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650)

ReN-Team;

The project "Reference Corpus Middle Low German/Low Rhenish (1200–1650)", abbreviated as "ReN", is part of the "Corpus of Historical German Texts", which includes the projects "Old German Reference Corpus (750–1050)", the "Reference Corpus Mid

Uploaded on June 2, 2021

10 more version(s) exist for this record

November 26, 2020 (1.1) Dataset Open Access View

EXMARaLDA Demo corpus 1.1

Hamburger Zentrum für Sprachkorpora;

A selection of short audio and video recordings in various languages to be used for instruction or demonstration of the EXMARaLDA system. The EXMARaLDA Demo Corpus is a small corpus which you can use to try out the functionality of the EXMARaLDA system. Please note that this corpus is for demonstra

Uploaded on November 26, 2020

November 15, 2018 (0.1) Dataset Open Access View

TraCES

Bausl, Alessandro; Dickhut, Wolfgang; Eilwardt, Andreas; Hummel, Susanne; Ermias, Hiruie; Pisani, Vitagrazia; Sokolinski, Eugenia;

Dealing with data from a CLARIN repository

Repository at the CLARIN-Center in Hamburg

- Fedora/Islandora based, own development
- DSA and CLARIN certified
- No migration plan towards Fedora 5.x
- No personal for maintenance
- No updates to already stored corpora

With consent of the data “owners”, corpora are being moved towards our zenodo clone research data repository “just in case”...

HZSK repository still running and still target of all CLARIN handles

The screenshot shows the HZSK Linguistic Corpora website. The header includes the University of Hamburg logo and the HZSK logo (hamburger centrum für sprachkorpora). The navigation menu has links for HZSK, CORPORA, TOOLS, SUPPORT, CONTACT, and DISCLAIMER. A search bar is present, and the main content area displays 'Linguistic Corpora at the HZSK Repository' with a search filter and a list of corpora. The list includes columns for Modify, License type, Corpus type, Language, and Keyword. A search result for 'EXMARALDA Demo Corpus 1.0' is highlighted, showing its description and license (HZSK-PUB (public)).

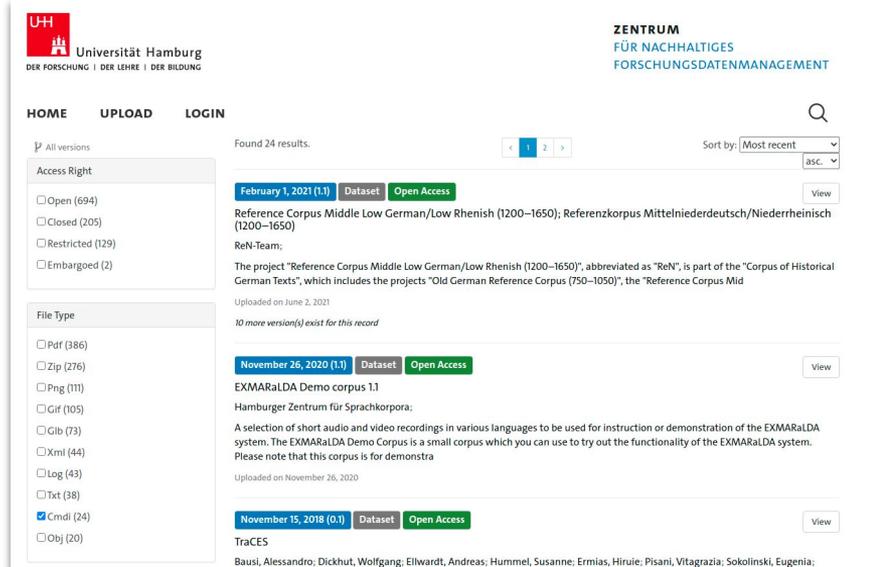
Wins and losses so far

Losses

- No CMDI metadata provider
- No part identifier resolution
- No FCS (as before...)
- No DSA/CLARIN certification
- No rights management for CLARIN-ACA

Wins

- Part of central data infrastructure (including personal)
- Servicable (at all)
- Connected to long term storage at the computing center (s3)
- User serviced entries (versioning etc.)
- Better visibility



The screenshot shows the ZENTRUM FÜR NACHHALTIGES FORSCHUNGSDATENMANAGEMENT interface. It features a navigation bar with 'HOME', 'UPLOAD', and 'LOGIN' options. A search bar at the top right shows 'Found 24 results.' and a sort dropdown set to 'Most recent'. The main content area displays a list of datasets. The first dataset is 'February 1, 2021 (1:1)' with 'Dataset' and 'Open Access' buttons. Below it is a description for 'Reference Corpus Middle Low German/Low Rhenish (1200–1650)'. The second dataset is 'November 26, 2020 (1:1)' with 'Dataset' and 'Open Access' buttons, described as 'EXMARaLDA Demo corpus 1.1'. The third dataset is 'November 15, 2018 (0:1)' with 'Dataset' and 'Open Access' buttons, described as 'TraCES'. On the left side, there are two filter panels: 'Access Right' with options for Open (694), Closed (205), Restricted (129), and Embargoed (2); and 'File Type' with options for Pdf (386), Zip (276), Png (111), Gif (105), Gif (73), Xml (44), Log (43), Txt (38), Cmdl (24), and Obj (20).

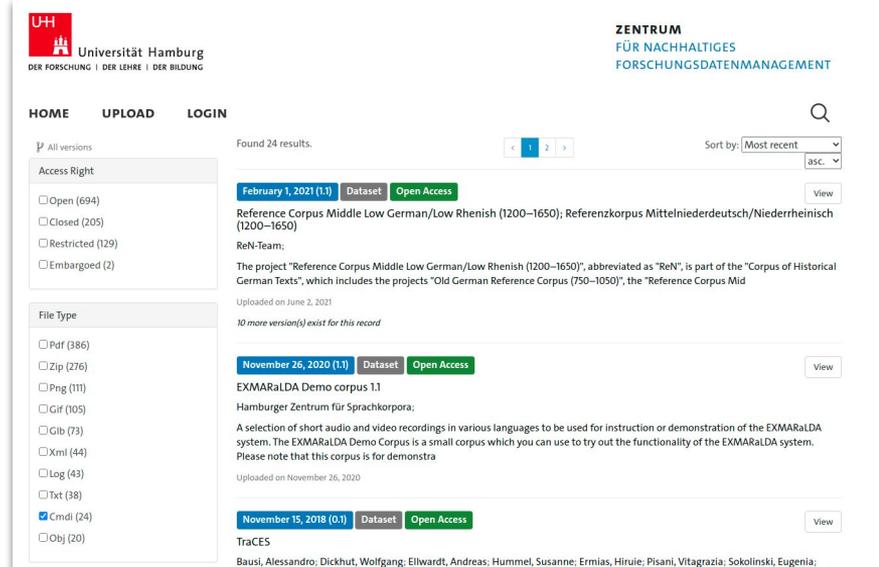
Wins and losses after migration

Losses

- No CMDI metadata provider (?)
- No part identifier resolution
- No FCS (as before...) (?)
- No ~~DSA~~/CLARIN certification
- ~~No rights management for CLARIN AGA~~

Wins

- Part of central data infrastructure (including personal)
- Servicable (at all)
- Connected to long term storage at the computing center (s3)
- User/Group serviced entries (versioning etc.)
- Better visibility
- Proactive sharing



The screenshot shows the ZENTRUM FÜR NACHHALTIGES FORSCHUNGSDATENMANAGEMENT interface. It features a navigation bar with 'HOME', 'UPLOAD', and 'LOGIN' options. A search bar at the top right shows 'Found 24 results.' and a sort dropdown set to 'Most recent'. The main content area displays a list of datasets with columns for date, version, dataset name, and actions like 'Dataset' and 'Open Access'. The first dataset is 'February 1, 2021 (1:1)' with a description: 'Reference Corpus Middle Low German/Low Rhenish (1200–1650); Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650)'. The second dataset is 'November 26, 2020 (1:1)' with a description: 'EXMARaLDA Demo corpus 1.1'. The third dataset is 'November 15, 2018 (0:1)' with a description: 'TraCES'. The interface also includes a sidebar with 'Access Right' and 'File Type' filters.

Check it out

visit the website

<https://inveniosoftware.org/products/rdm/>

join the discussion on discord

<https://discord.com/invite/8qatqBC>

install a local copy

```
pip install invenio-cli
invenio-cli check-requirements --development
invenio-cli init rdm
cd my-site/
invenio-cli containers start --lock --build --setup
```

or play with the sandbox

<https://inveniordm.web.cern.ch/>

Thanks

contact me: kai.woerner@uni-hamburg.de

our center: <https://fdm.uni-hamburg.de>

this presentation: <http://uhh.de/fdm-ccm>