

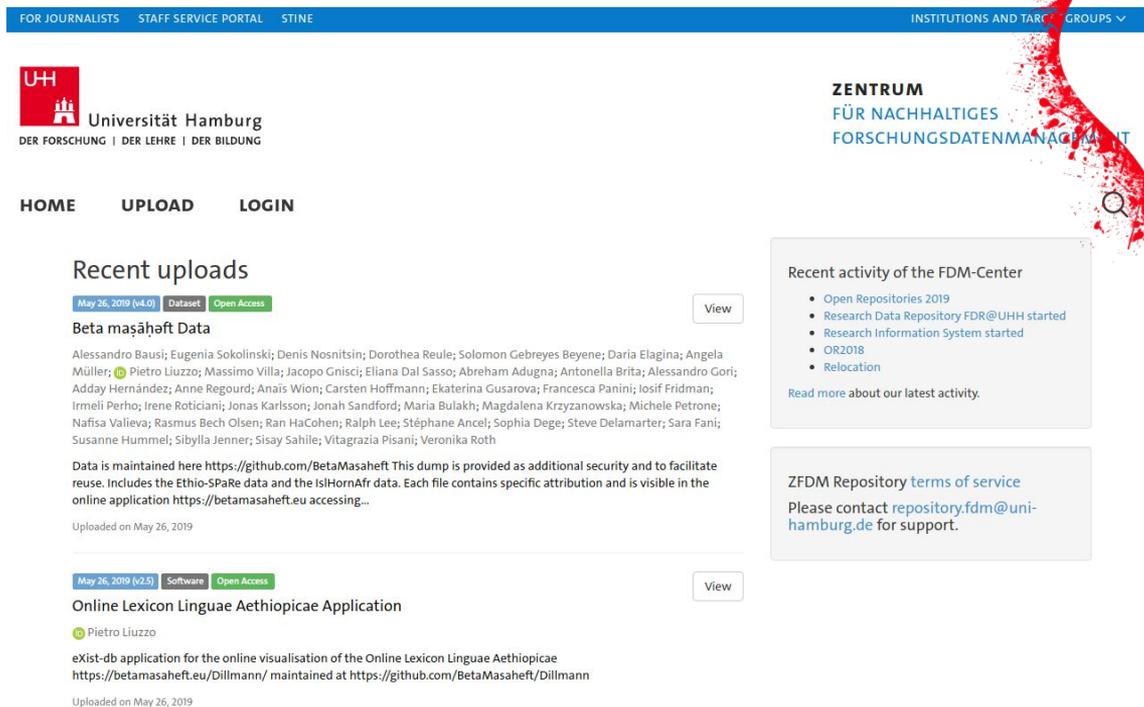
# Stripping down Zenodo to build a research data repository @ Universität Hamburg

Kai Wörner [[kai.woerner@uni-hamburg.de](mailto:kai.woerner@uni-hamburg.de)]

Center for Sustainable Research Data Management

Universität Hamburg

# We chose Zenodo because “UX and attractiveness are important for the acceptance of a repository.”



The screenshot shows the Zenodo repository interface. At the top, there is a navigation bar with links for 'FOR JOURNALISTS', 'STAFF SERVICE PORTAL', 'STINE', and 'INSTITUTIONS AND TARGET GROUPS'. Below the navigation bar is the University of Hamburg logo and name. The main content area is divided into sections: 'HOME', 'UPLOAD', and 'LOGIN'. The 'Recent uploads' section features two entries:

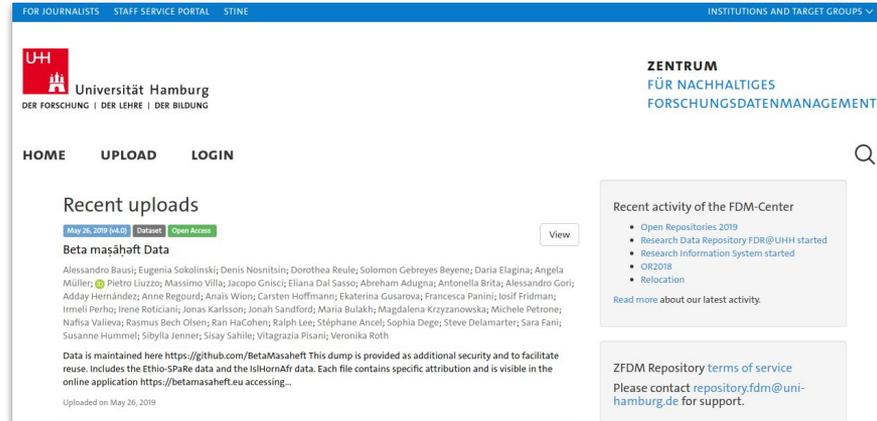
- Beta maṣāḥēft Data** (May 26, 2019, v4.0) - Dataset, Open Access. Description: Data is maintained here <https://github.com/BetaMasaheft>. This dump is provided as additional security and to facilitate reuse. Includes the Ethio-SPaRe data and the IslHornAfr data. Each file contains specific attribution and is visible in the online application <https://betamasheft.eu> accessing...  
Uploaded on May 26, 2019
- Online Lexicon Linguae Aethiopiae Application** (May 26, 2019, v2.5) - Software, Open Access. Description: eXist-db application for the online visualisation of the Online Lexicon Linguae Aethiopiae <https://betamasheft.eu/Dillmann/> maintained at <https://github.com/BetaMasaheft/Dillmann>  
Uploaded on May 26, 2019

On the right side, there is a section titled 'ZENTRUM FÜR NACHHALTIGES FORSCHUNGSDATENMANAGEMENT' with a large red heart graphic. Below this, there are two boxes:

- Recent activity of the FDM-Center**
  - Open Repositories 2019
  - Research Data Repository FDR@UHH started
  - Research Information System started
  - OR2018
  - Relocation[Read more about our latest activity.](#)
- ZFDM Repository terms of service**

Please contact [repository.fdm@uni-hamburg.de](mailto:repository.fdm@uni-hamburg.de) for support.

# Two ways towards a Zenodo-like repository

**ZENTRUM FÜR NACHHALTIGES FORSCHUNGSDATENMANAGEMENT**

**Recent uploads**  
 May 26, 2019 (v4.0) Dataset Open Access View

**Beta mašáhft Data**  
 Alessandro Bausi; Eugenia Sokolinski; Denis Nosnitsin; Dorothea Reule; Solomon Gebreyes Beyene; Daria Elagina; Angela Müller; Pietro Liuzzo; Massimo Villa; Jacopo Gnisci; Eliana Dal Sasso; Abreham Adugna; Antonella Brita; Alessandro Gori; Adday Hernández; Anne Regourd; Anais Wilo; Carsten Hoffmann; Ekaterina Gusarova; Francesca Panini; Iosif Fridman; Irmeli Perho; Irene Roticiari; Jonas Karlsson; Jonah Sandford; Maria Bulakh; Magdalena Krzyzanowska; Michele Petrone; Naftsa Valleja; Rasmus Bech Olsen; Ran HaCohen; Ralph Lee; Stéphane Ance; Sophia Dege; Steve Delamarter; Sara Fani; Susanne Hummel; Silylia Jenner; Sisy Sahle; Vítagratiá Piani; Veronika Roth

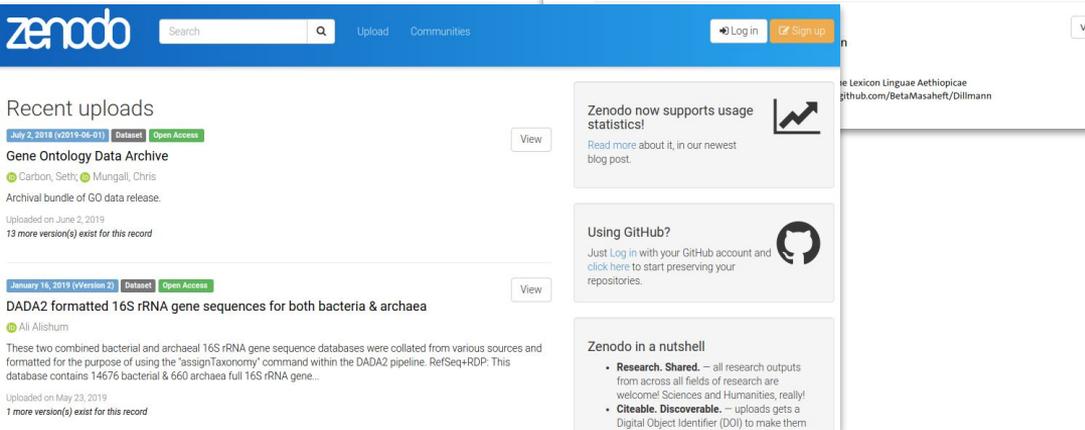
Data is maintained here <https://github.com/BetaMasahft> This dump is provided as additional security and to facilitate reuse. Includes the Chivo-SpARc data and the ISHomA4r data. Each file contains specific attribution and is visible in the online application <https://betamasahft.eu> accessing...  
 Uploaded on May 26, 2019

**Recent activity of the FDM-Center**  

- Open Repositories 2019
- Research Data Repository FDR@UHH started
- Research Information System started
- OR2018
- Relocation

 Read more about our latest activity.

ZFDM Repository [terms of service](#)  
 Please contact [repository.fdm@uni-hamburg.de](mailto:repository.fdm@uni-hamburg.de) for support.

**zenodo** Search Upload Communities Log in Sign up

**Recent uploads**  
 July 2, 2018 (v2019-06-01) Dataset Open Access View  
**Gene Ontology Data Archive**  
 Carbon, Seth; Mungall, Chris  
 Archival bundle of GO data release.  
 Uploaded on June 2, 2019  
 13 more version(s) exist for this record

January 16, 2019 (vVersion 2) Dataset Open Access View  
**DADA2 formatted 16S rRNA gene sequences for both bacteria & archaea**  
 Ali Alilshum  
 These two combined bacterial and archaeal 16S rRNA gene sequence databases were collated from various sources and formatted for the purpose of using the 'assignTaxonomy' command within the DADA2 pipeline. RefSeq+RDP. This database contains 14676 bacterial & 660 archaea full 16S rRNA gene...  
 Uploaded on May 23, 2019  
 1 more version(s) exist for this record

**Zenodo now supports usage statistics!**  
 Read more about it, in our newest blog post.

**Using GitHub?**  
 Just Log in with your GitHub account and click here to start preserving your repositories.

**Zenodo in a nutshell**  

- Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- Citable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them

```

$ curl -k --header "Content-Type: application/json" \
  https://localhost:5000/api/records/1?prettyprint=1
  
```

```

{
  "created": "2018-05-23T13:28:19.426206+00:00",
  "id": 1,
  "links": {
    "self": "https://localhost:5000/api/records/1"
  },
  "metadata": {
    "contributors": [
      {
        "name": "Doe, John"
      }
    ]
  },
  "type": "dataset"
}
  
```

# What's good about building from Invenio/Scratch

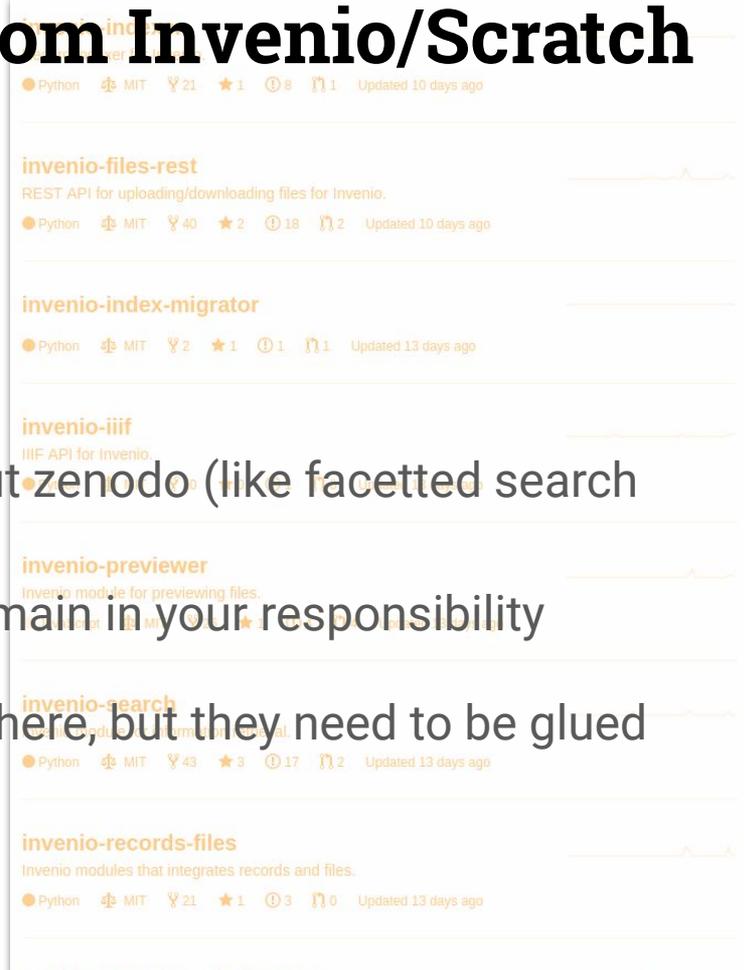
- build upon well maintained Invenio modules
- easier to keep up with Invenio development
- only create (and maintain) code and module you really need
- “clean slate” codebase
- probably easier to understand what you actually do

# What's wrong with building from Invenio/Scratch

It's a lot of work!

- no out-of-the-box UI functionality
- no datamodel\*
- no Deposit-UI\*
- none of the nice UI features people love about zenodo (like faceted search etc)\*
- modules that you build for your repository remain in your responsibility

\* most components can be grabbed from somewhere, but they need to be glued together by you



# What's good with cutting zenodo.org to size

- have Zenodo's functionality and UX right away
- use a community-proven product that's actually used ->
- all rough edges or problems have at least been addressed, if not solved
- benefit from new zenodo.org features (like DOI versioning, stats)
- it's still an Invenio repository, so you can mostly use Invenio and often general Flask modules to improve your repository
  - We used shibboleth-authenticator to plug zenodo into our IDMS
  - We\* contributed invenio-S3 to the Invenio codebase to connect zenodo to our storage

 **records: new field for methods** **Enhancement** **Needs design** **Triage: Accepted** **i - Arcadia**

#1695 opened on 5 Feb by Inielsen  4 of 5 

# What's wrong with cutting zenodo.org to size



Lars Holm Nielsen @lnielsen

Jun 12 2018 14:25

Also, please note that Zenodo is a service, and we don't make any gurantees about backward compatibility etc....

\*10<sup>3</sup>

# What's wrong with cutting zenodo.org to size

- some things are not for everyone (special APIs for Zenodo, e.g. for OpenAIRE)
- some code is pretty hardwired to the service in the Zenodo codebase
- things can radically change and it's no one's fault
- cherry picking changes into your altered codebase is less fun than picking cherries from trees
- how things are done in Zenodo is sometimes hard to grasp for unexperienced Invenio/Flask "developers"

# What's wrong with cutting zenodo.org to size

- some things are not for everyone (special APIs for Zenodo, e.g. for OpenAIRE)
- some code is pretty hardwired to the service in the Zenodo codebase
- things can radically change and it's no one's fault
- cherry picking changes into your altered codebase is less fun than picking cherries from trees
- how things are done in Zenodo is sometimes hard to grasp for unexperienced Invenio/Flask "developers"



# So what's the better way to do it?

- the “cut-to-size”-approach provides you with a working repository much earlier

```
$ git clone https://github.com/zenodo/zenodo.git
$ git checkout master
$ docker-compose build
$ docker-compose up
$ docker-compose run --rm web bash /code/zenodo/scripts/init.sh
$ docker-compose run --rm statsd bash /init.sh
$ docker-compose run --rm web zenodo index reindex --yes-i-know
$ docker-compose run --rm web zenodo index run -d
```



this actually works!

- the “build from scratch”-approach probably results in a product that’s easier to maintain
- both are doable if you have enough Flask or even better Invenio experience

# So what's the better way to do it?

- the “cut-to-size”-approach provides you with a working repository much earlier

```
$ git clone https://github.com/zenodo/zenodo.git
$ git checkout master
$ docker-compose build
$ docker-compose up
$ docker-compose run --rm web bash /code/zenodo/scripts/init.sh
$ docker-compose run --rm statsd bash /init.sh
$ docker-compose run --rm web zenodo index reindex --yes-i-know
$ docker-compose run --rm web zenodo index run -d
```



this actually works!

- the “build from scratch”-approach probably results in a product that’s easier to maintain
- both are doable if you have enough Flask or even better Invenio



Hi there!

# What would be an even better way?



Lars Holm Nielsen @lnielsen

Jun 12 2018 14:25

Also, please note that Zenodo is a service, and we don't make any gurantees about backward compatibility etc....

# What would be an even better way?



Lars Holm Nielsen @lnielsen

Jun 12 2018 14:25

Also, please note that Zenodo is a service, and we don't make any gurantees about backward compatibility etc.... if you're interested in an Invenio Research Data Management flavour based on Zenodo (which would have nice upgrade path etc), let us know - we're trying to gather of group of interested parties since there's a couple of other zenodo/invenio installations out there and if we pool resources we could make it happen

# What would be an even better way?



Also, please note that Zenodo is a service, and we don't

ma

yo

fla

03

17

zenodo/invenio installations out there and if we pool

resources we could make it happen

**InvenioRDM Project Kick-Off**

24 Jun 2019, 11:45 → 26 Jun 2019, 16:00 Europe/Zurich

CERN

# Invenio RDM

## Project Goals:

1. Build an open-source turn-key research data management (RDM) platform based on Zenodo and Invenio v3.
2. Build a community of research institutions and private companies to grow and sustain the platform going forward.

Inveniocrdm project

ESA-E-121, Universität Hamburg

Lars Holm Nielsen

15:30 - 15:45

# Steps ahead:

- Define requirements and build Invenio RDM based on Zenodo's functionality
- Move our repository to the Invenio RDM codebase