

Open Discovery: Forschungsdaten & Zenodo

...

Leif Klemstein, Jonas Engeland

Gliederung

- Zenodo
- Forschungsdaten
- Catmandu
 - Catmandu Docker Container
- Zenodo OAI
- Forschungsdatendownload
- OpenRefine
- Solr
- Typo3-Find
- Quellenverzeichnis

Zenodo

- CERN-Dienstleistung & OpenAIRE
- Finanziert über die Europäische Kommission.
- Forschungsergebnisse in jedem Umfang, jedem Format, aus jeder Fachrichtung zu teilen und aufbewahren
- wissenschaftliche Datensätze, aber auch für wissenschaftsbezogene Publikationen, Berichte, Präsentationen, Videos etc. verwendet werden kann.

Forschungsdaten

- Forschungsdaten bezeichnen (digitale) Daten, die je nach Fachkontext Gegenstand eines Forschungsprozesses sind, während eines Forschungsprozesses entstehen oder sein Ergebnis sind.

Catmandu

Catmandu ist ein Tool, mit dem man per Kommandozeile Zugang zu Daten erhält und diese Daten konvertieren kann. Das Toolkit war ursprünglich Teil des LibreCat-Projekts und ist jetzt international aufgestellt.

Mit Catmandu kann man folgende Aufgaben durchführen:

- Daten-Downloads per Protokoll oder Standard, wie z. B. OAI-PMH, SRU, SPARQL und Linked- Data-Fragmenten.
- Konvertieren von verschiedenen Formaten wie MARC, MODS, Dublin Core und andere sowie JSON, YAML, XML, Excel und viele mehr

Catmandu

- RDF erstellen und mit Semantic Web arbeiten
- Daten in Datenbanken indexieren, wie Solr, Elasticsearch und MongoDB.
- Metadaten in jedes mögliche Format konvertieren

Catmandu wird im LibreCat-Projekt dazu verwendet, Repositorien und Suchmaschinen zu erstellen. Mit Catmandu kann man schnell Berichte erstellen und größere Programmierprojekte durchführen.

Catmandu - Docker Container

1. `mkdir catmandu -->`
2. `chmod a+rwX catmandu`



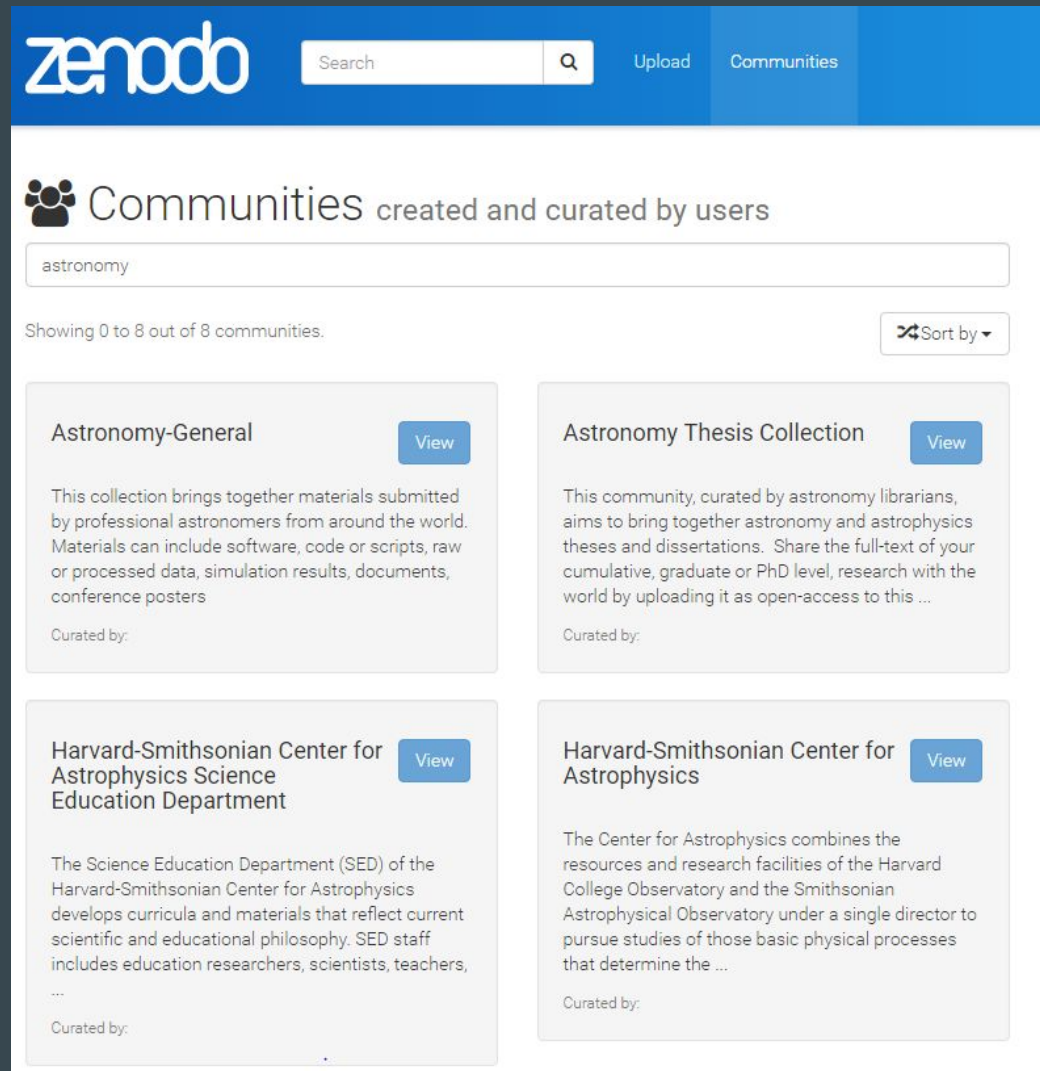
Name	Größe	Geändert	Rechte	Besitzer
catmandu		07.01.2017 14:29:10	rwxrwxrwx	stud
download		07.01.2017 15:38:09	rwxr-xr-x	stud
oaip-master		05.01.2017 10:23:09	rwxr-xr-x	stud
refine		07.01.2017 14:45:38	rwxr-xr-x	stud
refine-backup_07_5-3		20.11.2016 19:34:01	rwxr-xr-x	stud
refine-backup_07_5-8		21.11.2016 00:59:18	rwxrwxr-x	stud
testverzeichnis		11.10.2016 13:19:46	rwxr-xr-x	stud

3. `sudo docker run --rm -v /home/stud/catmandu:/home/catmandu/ -it librecat/catmandu`

```
stud@lohmeier6:~$ sudo docker run --rm -v /home/stud/catmandu:/home/catmandu/ -it librecat/catmandu
catmandu@829202ef46a7:~$
```

Zenodo - OAI

- 1015 Anzahl Communities
- Unterschiedliche Fachgebiete - von Astrophysik bis Zoology
- Insgesamt: 126.780 Forschungsdaten, die auf unterschiedliche Dokumententypen verweisen



zenodo

Search

Upload Communities

Communities created and curated by users

Showing 0 to 8 out of 8 communities. Sort by

Astronomy-General [View](#)

This collection brings together materials submitted by professional astronomers from around the world. Materials can include software, code or scripts, raw or processed data, simulation results, documents, conference posters

Curated by:

Astronomy Thesis Collection [View](#)

This community, curated by astronomy librarians, aims to bring together astronomy and astrophysics theses and dissertations. Share the full-text of your cumulative, graduate or PhD level, research with the world by uploading it as open-access to this ...

Curated by:

Harvard-Smithsonian Center for Astrophysics Science Education Department [View](#)

The Science Education Department (SED) of the Harvard-Smithsonian Center for Astrophysics develops curricula and materials that reflect current scientific and educational philosophy. SED staff includes education researchers, scientists, teachers, ...

Curated by:

Harvard-Smithsonian Center for Astrophysics [View](#)

The Center for Astrophysics combines the resources and research facilities of the Harvard College Observatory and the Smithsonian Astrophysical Observatory under a single director to pursue studies of those basic physical processes that determine the ...

Curated by:

Curated by:

AstroCurator

Curation policy:

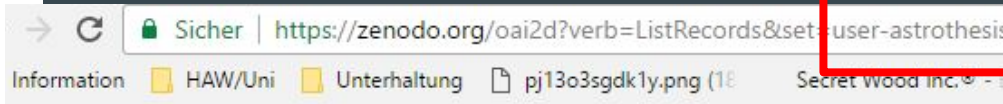
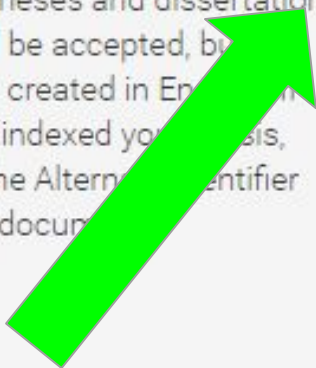
This collection includes theses and dissertations written to complete Masters or PhD requirements on topics in the fields of astronomy and astrophysics. Other publication types and topics will not be accepted; however, documents that supplement your thesis, such as data sets or videos, are welcome. Although Zenodo accepts all file formats, PDF is strongly preferred for publication submissions. Theses and dissertations written in any language will be accepted, but a metadata record should be created in English. If the NASA ADS has already indexed your thesis, provide us the Bibcode in the Alternative Identifier field before submitting the document.

Created:

April 1, 2015

Harvesting API:

OAI-PMH Interface



OAI 2.0 Request Results

[Identify](#) | [ListRecords](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browser.

Datestamp of response	2017-01-07T13:32:40Z
Request URL	https://zenodo.org/oai2d

Request was of type ListRecords.

Username = astrothesis → 1 Community

Username = Zenodo → alle Communities

Forschungsdatendownload




1. `sudo docker run --rm -v /home/stud/catmandu:/home/catmandu/ -it libreca/catmandu`

```
stud@lohmeier6:~$ sudo docker run --rm -v /home/stud/catmandu:/home/catmandu/ -it libreca/catmandu
catmandu@06717d6c0dbd:~$ catmandu convert OAI --url https://zenodo.org/oai2d --set user-astronomy-general --metadataPrefix oai_dc to JS
ON > astronomy-general.json
catmandu@06717d6c0dbd:~$ █
```

2. `catmandu convert OAI --url https://zenodo.org/oai2d --set user-BEISPIELUSER --metadataPrefix oai_dc to JSON > BEISPIELDATEINAME.json`

z.B. astronomy-general oder astrothesis

DOWNLOAD 

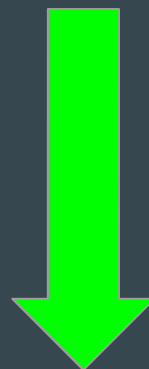
 astro.json	48 KB	07.01.2017 14:25:23	rw-r--r--	thomi
 astronomy-general.js...	48 KB	07.01.2017 16:01:54	rw-r--r--	thomi
 astrothesis.json	301 KB	07.01.2017 16:06:52	rw-r--r--	thomi
 lory.json	571 KB	07.01.2017 14:10:52	rw-r--r--	thomi
 zenodo-ecfunded.json	571 KB	07.01.2017 14:09:38	rw-r--r--	thomi
 zenodo-ecfunded-jso...	5.858 KB	07.01.2017 12:34:50	rw-r--r--	stud

3. `exit`

OpenRefine

vs records Show: 5 10 25 50 records

oai:zenodo.org:19197	2016-09-12T05:55:48Z	oai:zenodo.org:19197	10.5281/zenodo.19197	Sesar, Branimir	openaire_data	info:eu-

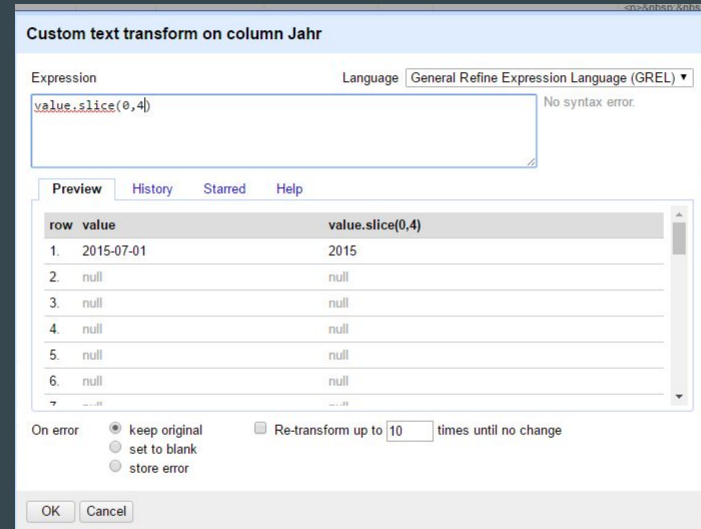
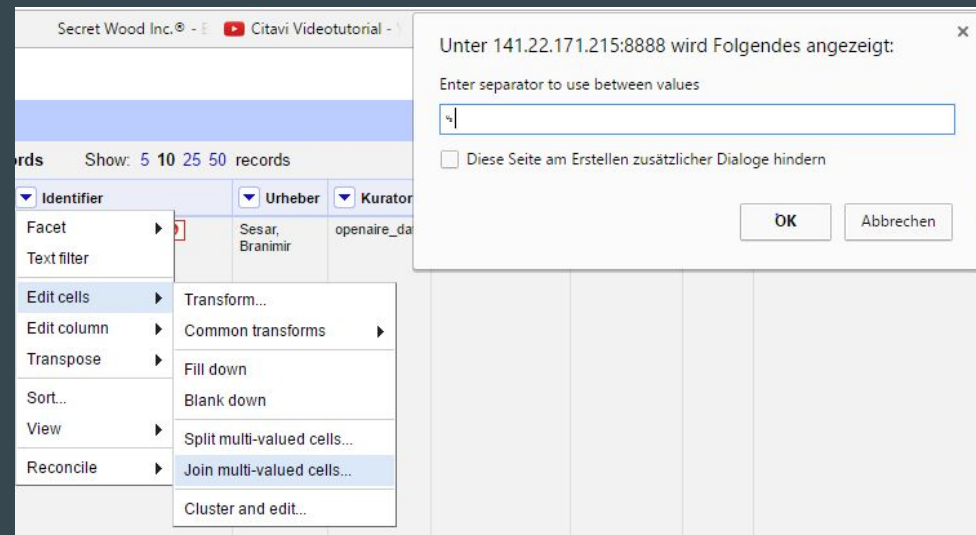


Show as: rows records Show: 5 10 25 50 records

All	id	Identifier	Urheber	Kurator	Schlagwoerter
1.	19197	10.5281/zenodo.19197 us https://zenodo.org/record/19197 us oai:zenodo.org:19197	Sesar, Branimir us Bovy, Jo us Bernard	openaire_data us user-ecfunded us user-astronomy-general	

OpenRefine

- Unit Separator: mehrere Zeilen zusammenfassen und die Daten mit dem Unit Separator u_s trennen
- per `value.slice()`-Befehl die Jahreszahl aus dem Datum extrahieren
- Rename: die Spaltennamen umbenennen. Orientierung an Dublin Core.



Solr

curl

```
"http://localhost:8983/solr/gettingstarted/update/csv?commit=true&separator=%09&f.DDC.split=true&f.DDC.separator=%E2%90%9F&f.Beschreibung.split=true&f.Kurator.split=true&f.Beschreibung.separator=%E2%90%9F&f.Schlagwoerter.split=true&f.Schlagwoerter.separator=%E2%90%9F&f.Beitragende.split=true&f.Beitragende.separator=%E2%90%9F&f.Link.split=true&f.Link.separator=%E2%90%9F" --data-binary @/home/stud/catmandu/BEISPIELDATEINAME --H 'Content-type:text/plain; charset=utf-8'
```

```
␣%90%9F" --data-binary @astronomy-general-json.tsv -H 'Content-type:text/plain; charset=utf-8'  
Warning: Couldn't read data from file "astronomy-general-json.tsv", this makes  
Warning: an empty POST.  
<?xml version="1.0" encoding="UTF-8"?>  
<response>  
<lst name="responseHeader"><int name="status">400</int><int name="QTime">1</int></lst><lst name="error"><lst name="metadata"><str name="error-class">org.apache.solr.common.SolrException</str><str name="root-error-class">org.apache.solr.common.SolrException</str></lst><str name="msg">Expected fieldnames in CSV input</str><int name="code">400</int></lst></response>  
stud@lohmeier6:~$ █
```

Typo3-Find

- Medientyp

On forced and free atmospheric oscillations near the 27-day periodicity

id

[zenodo.org:59601](https://zenodo.org/59601)

Urheber

- Schanz, Ansgar
- Hocke, Klemens
- Kämpfer, Niklaus

Schlagwoerter

- Solar forcing

Jahr

- 2016

Medientyp

- info:eu-repo/semantics/article

Titel

- On forced and free atmospheric oscillations near the 27-day periodicity

Beschreibung

- The Whole Atmosphere Community Climate Model was used to investigate the

MEDIENTYP

undefiniert 149741

Elektronische

Ressource 61674

info:eu-

repo/semantics/article 1682

info:eu-

repo/semantics/other 614

Bildtontraeger 484

info:eu-

repo/semantics/conferencePaper

show all

Quellenverzeichnis

CERN, 2016: Zenodo. [Online] Stand: 2016 [Zugriff am: 2017-01-08] Verfügbar unter: <https://github.com/Zenodo/zenodo>

KINDLING, M., und SCHIRMBACHER, P., 2013: „Die digitale Forschungswelt“ als Gegenstand der Forschung. In: *Information : Wissenschaft und Praxis* 64 (2/3), S. 127–136. ISSN 1619-4292.

LIBRECAT: *Catmandu*. [Online; Zugriff am: 2017-01-08]. Verfügbar unter: <https://hub.docker.com/r/librecat/catmandu/>

OPENAIRE, 2016: *Project factsheets*. [Online] Stand: 2016-09-05 [Zugriff am: 2017-01-08] Verfügbar unter: <https://www.openaire.eu/project-factsheets>