

# SSH Open Marketplace

View into  
Social Sciences and Humanities research practice

Laure Barbot, Matej Ďurčo  
ACDH-CH Austrian Academy of Sciences  
DARIAH & the SSH Open Cluster

26.10.2022  
Danish e-infrastructure Cooperation (DeiC) Conference

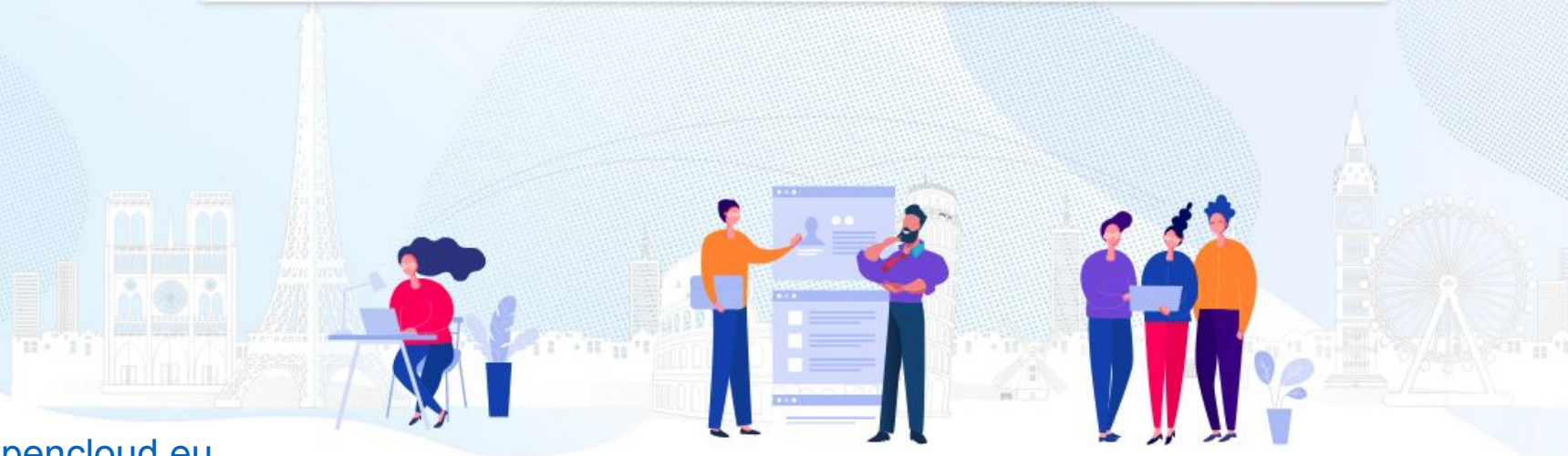
This presentation is the result of a collaborative work! [Meet the team!](#)



## Social Sciences & Humanities Open Marketplace

Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets. [Read more...](#)

All categories

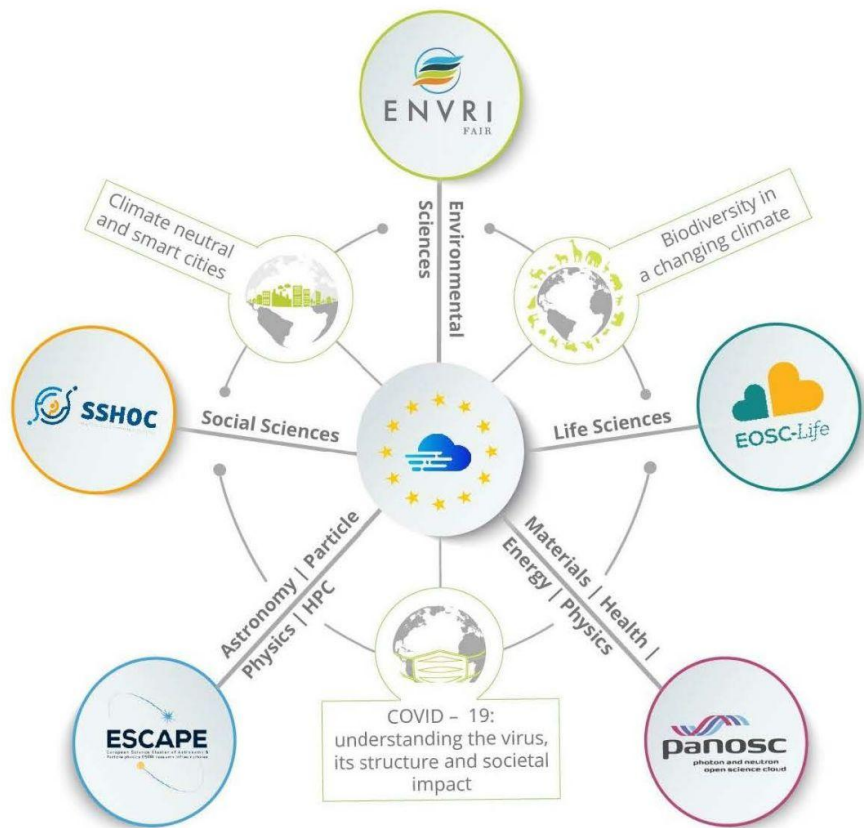


[marketplace.sshopencloud.eu](https://marketplace.sshopencloud.eu)

The SSH Open Marketplace is maintained and will be further developed by three European Research Infrastructures - DARIAH, CLARIN and CESSDA - and their national partners. It was developed as part of the "Social Sciences and Humanities Open Cloud" SSHOC project, European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782.



# The SSH Open Cluster in the EOSC



## 33 SSHOCing Key Exploitable Results to support your research in SSH and Beyond



SSHOC. (2022). SSHOC Legacy booklet. Zenodo.  
<https://doi.org/10.5281/zenodo.6394462>

# SSH Open Marketplace Governance “scheme”

## 3 ERICs service agreement

DARIAH, CLARIN & CESSDA signing a collaborative and **binding agreement** and giving resources to maintain and further develop the SSH Open Marketplace after the end of the SSHOC project.

**Service provision** ensured by ACDH-CH and PSNC



## Research communities / User contributions

The SSH Open Marketplace is designed as to support community curation. Contributors can add or enrich existing content via a user-friendly interface

Hands-on sessions are the main means envisioned to support contributions.

## Editorial Board & the day-to-day operations

Incl. moderators and administrators who take care of the maintenance, data quality and curation of the SSH Open Marketplace. They are also in charge of the community engagement.

Clara Petitfils, Suzanne Dumouchel, Nicolas Larrousse, Edward J. Gray, Laure Barbot, Arnaud Roi, Matej Ďurčo, Klaus Illmayer, Stefan Buddenbohm, & Tomasz Parkola. (2021). D7.5 Marketplace - Governance. <https://doi.org/10.5281/zenodo.5608487>

DaCy is a Danish text processing pipeline built using SpaCy. At the time of writing, it has achieved State-of-the-Art performance on part-of-speech (POS) tagging, named-entity recognition (NER) and Dependency parsing for Danish.

[Go to Tool or service](#)

## Details

### ACCESS

License [Apache License 2.0](#)

### CATEGORISATION

Activity [Analyzing](#) [Enriching](#) [Named Entity Recognition](#) [POS-Tagging](#) [Natural Language Processing](#) [Parsing](#)

Keyword [SpaCy](#) [DARIAH National Resource](#)

Discipline [Digital humanities](#) [Computer Sciences](#)

Language [Danish](#)

Intended audience [Researcher](#) [Data creator](#) [Data engineer](#)

Resource category [Data Analysis](#) [Software Libraries](#)

### CONTEXT

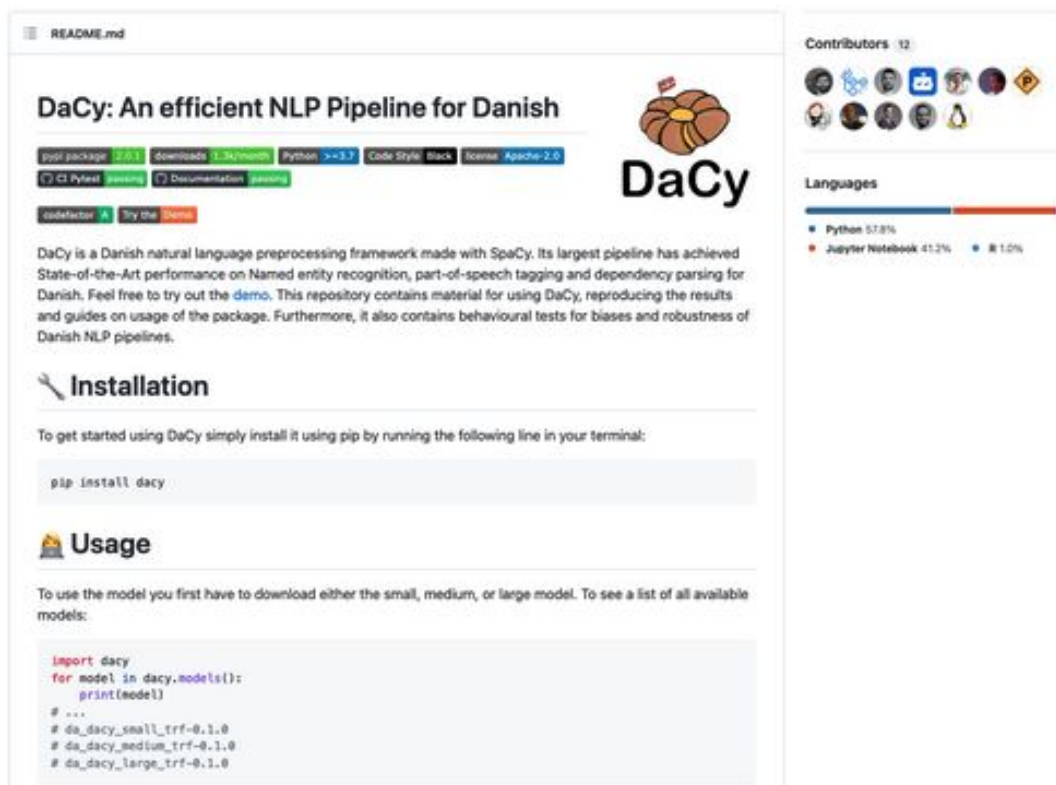
See also [http://ceur-ws.org/Vol-2989/short\\_paper24.pdf](http://ceur-ws.org/Vol-2989/short_paper24.pdf)

### CREATOR

[Kenneth C. Enevoldsen](#)

[Aarhus University](#)

[ORCID](#)



The screenshot shows the GitHub README for DaCy. It features the DaCy logo, a list of badges for package version (2.1.1), downloads (1.7k/month), Python version (>=3.3), code style (Black), and license (Apache-2.0). The text describes DaCy as a Danish natural language preprocessing framework made with SpaCy, highlighting its State-of-the-Art performance on NLP tasks. It includes sections for 'Installation' with the command 'pip install dacy' and 'Usage' with a code snippet for listing models.

```
import dacy
for model in dacy.models():
    print(model)
# ...
# da_dacy_small_trf-0.1.0
# da_dacy_medium_trf-0.1.0
# da_dacy_large_trf-0.1.0
```

DaCy screenshot

DaCy is a Danish text processing pipeline built using SpaCy. At the time of writing, it has achieved State-of-the-Art performance on part-of-speech (POS) tagging, named-entity recognition (NER) and Dependency parsing for Danish.

[Go to Tool or service](#)

## Details

### ACCESS

License [Apache License 2.0](#)

## Related items (1)



spaCy

spaCy is an open-source software library for advanced natural language processing, written in the programming languages Python and Cython. The library is published under the MIT license and currently offers statistical...

[Read more](#)[Show more](#)[Eusebio Hernandez](#)

Aarhus University

[ORCID](#)

Kristoffer L. Nielbo

Aarhus University

[dblp](#) [ORCID](#) [Wikidata](#)

### CURATOR

DARIAH-EU

<https://www.dariah.eu/>

### REVIEWER

DARIAH-DK

<https://dighumlab.org/dariah-dk/>[GitHub](#)

# Workflows



## Create a dictionary in TEI

[Copy to clipboard](#)

This scenario sets out the best practices for creating a born-digital dictionary, especially with the TEI (Text Encoding Initiative). However, building a standardized lexicographical dataset is not only a data format problem, it is also an intellectual and technical process where one has to choose how to model their data, and with which tools operate in order to create an easy-to-use and sustainable resource.

[Go to Workflow](#)

## Details

### ACCESS

License [Creative Commons Attribution 4.0 International](#)

### CATEGORISATION

Activity [Encoding](#)

Keyword [Linguistics](#) recommended

Language [English](#)



# Workflows



## Create a dictionary in TEI

### Workflow steps (5)

- 1 Build the model of the dictionary
- 2 Create a corpus of useful resources for the dictionary.
- 3 **Setting up the editing environment**
- 4 Setting up the publishing environment
- 5 Provide for long-term availability

### 3 Setting up the editing environment

Collapse ▲

Choose a tool for editing the dictionary, and also consider the back-end. For the latter one, we can discern two prototypical approaches: systems based on relational databases (MySQL, PostgreSQL etc.) and native XML databases (eXist, BaseX).

#### Related items (4)



#### Dictionary Writing Systems \& Corpus Query Systems

No description provided.

[Read more](#)



#### BaseX Documentation

BaseX is a light-weight, high-performance and scalable XML Database and an XQuery 3.1 Processor with full support for the W3C Update...

[Read more](#)



#### Introducing Lexonomy: an open-source dictionary writing and publishing system

This demo introduces Lexonomy [www.lexonomy.eu](http://www.lexonomy.eu), a free, open-source, web-based dictionary writing and publishing system. In...

[Read more](#)



#### VLE - Viennese Lexicographic Editor

VLE is an experimental piece of software that first came into existence as a by-product of a collaborative glossary editing project carried out ...

[Read more](#)





# Content types (see [about/data-population](#))



- **Tools & services** Materials or products used to perform activities: Desktop clients solutions (to be installed locally), Browser-based or command-line based resources, Mobile apps, Programming libraries or APIs, Data catalogues



- **Training materials** Tutorials, lessons or didactic resources explaining how to perform an action or highlighting the learning outcomes one would gain from engaging with the material.



- **Publications** Research results published in academic journals or repositories. The SSH Open Marketplace references only publications that can be connected to other resources and is not an exhaustive collection of papers.



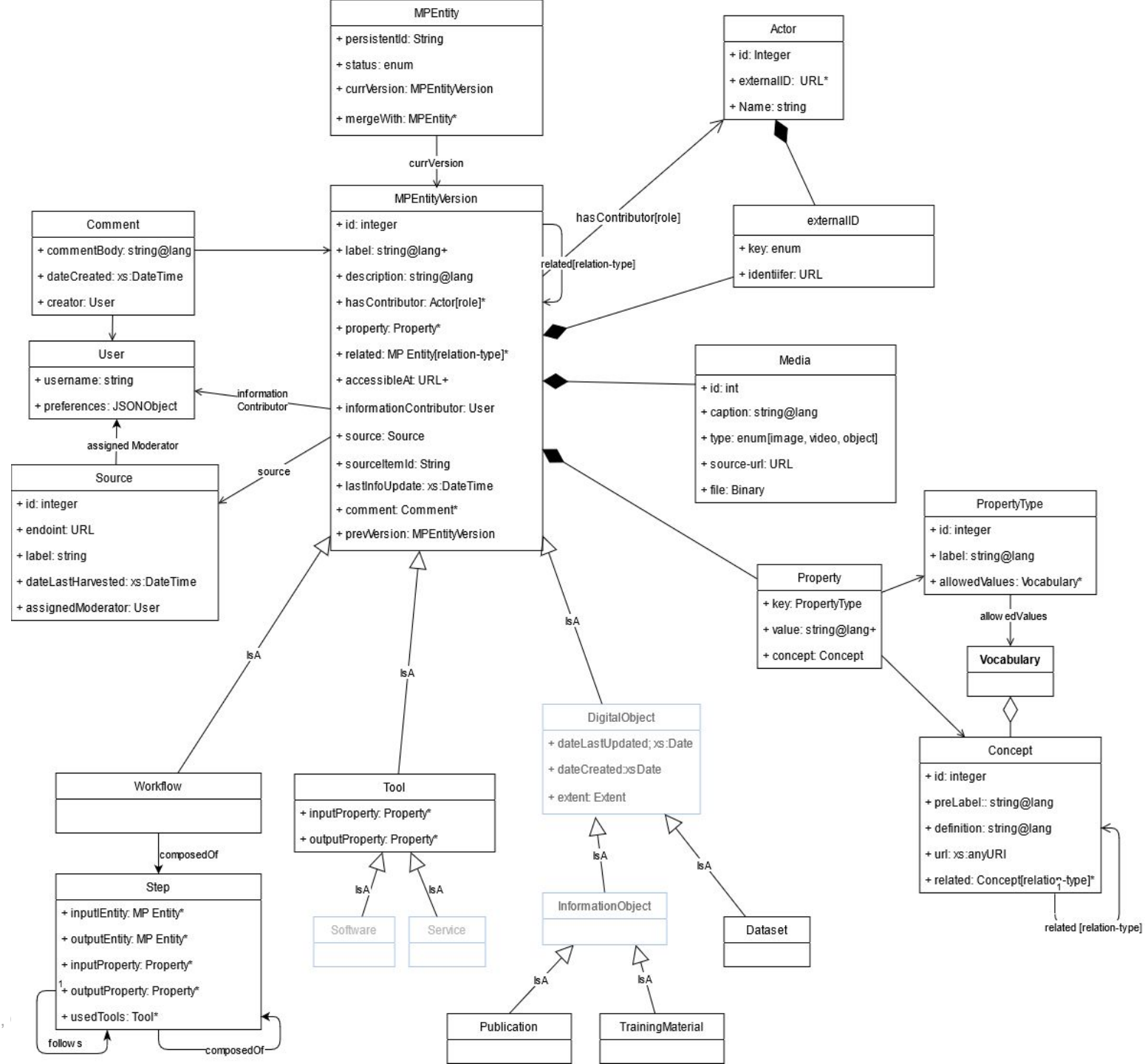
- **Datasets** Single digital objects or collections of data, records, or information that is kept as a persistent unit of information in the knowledge generation process. Datasets are used as evidence for some phenomena.



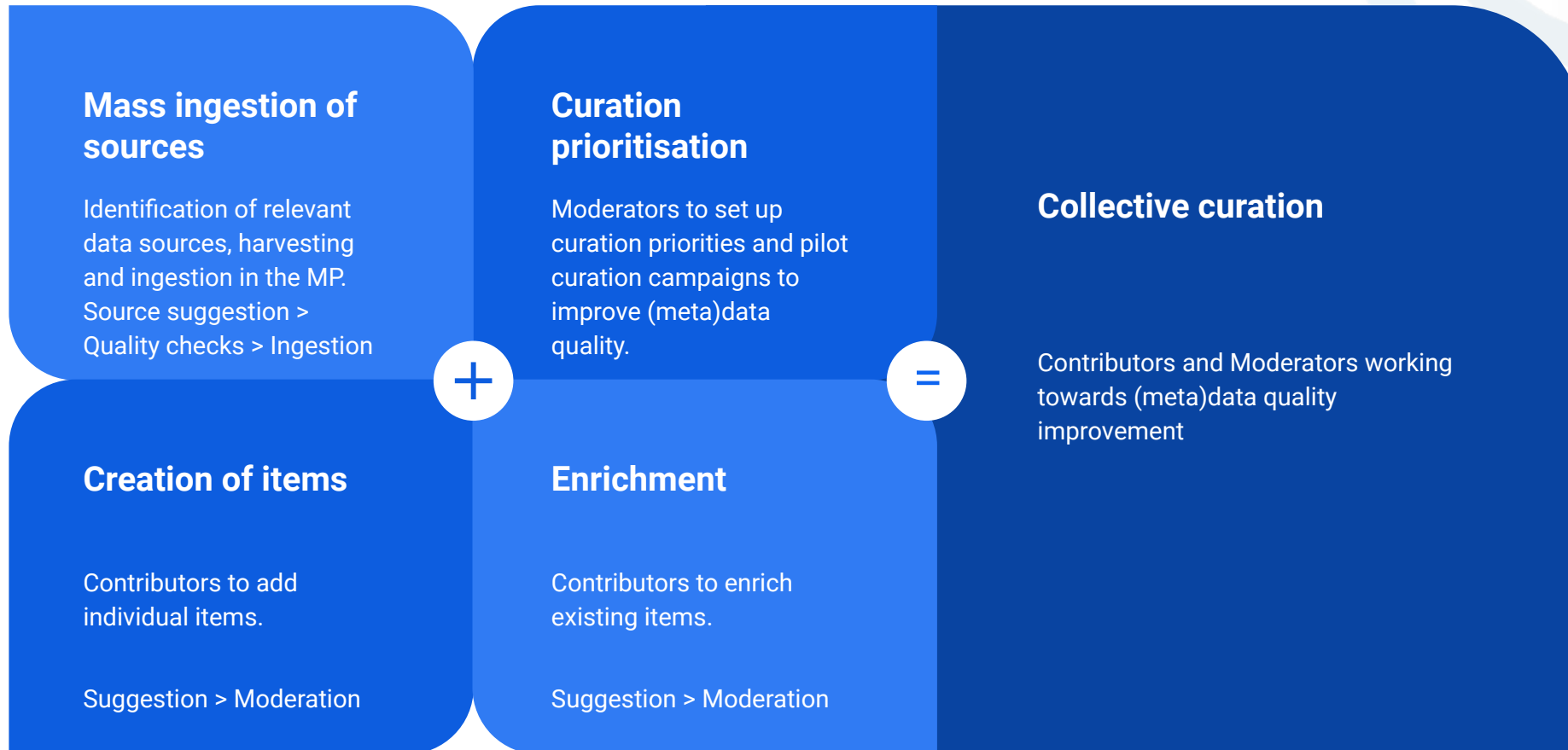
- **Workflows** Sequences of operation/steps performed on research data during their lifecycle. Workflows can be achieved by using diverse tools and facilities, and useful resources are connected to each step.

# Data model

Matej Ďurčo, Laure Barbot, Klaus Illmayer, Sotiris Karampatakis, Frank Fischer, Yoann Moranville, Joshua Teteh Ocansey, Stefan Probst, Michal Kozak, Stefan Buddenbohm, & Seung-Bin Yim. (2021). 7.2 Marketplace – Implementation (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.5749465>



# Curation puzzle



# Editorial Board missions

“An Editorial Team ensures the **day-to-day maintenance** and the **data quality** of the SSH Open Marketplace. The Editorial Team, liaising with the service providers and the **end-users** of the service (SSH researchers and support staff for researchers), ensures the technical running of operation, the effectiveness of the curation process and the editorial policy’s successful implementation.”

[.../about/service#governance--sustainability](#)

## User roles

### Visitor

Consult, browse, discover, re-use. No account needed

### Contributor

Consult, browse, discover, re-use. Promote, create, enrich.  
Account creation via the [sign in function](#).

### Moderator

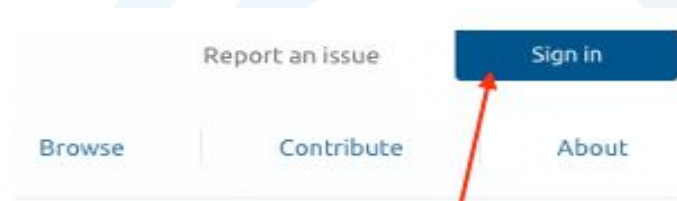
Moderate, curate, advise

### Admin.

Moderate, curate, advise. Users and sources management.

# Federated login via EOSC

[.../auth/sign-in](#)



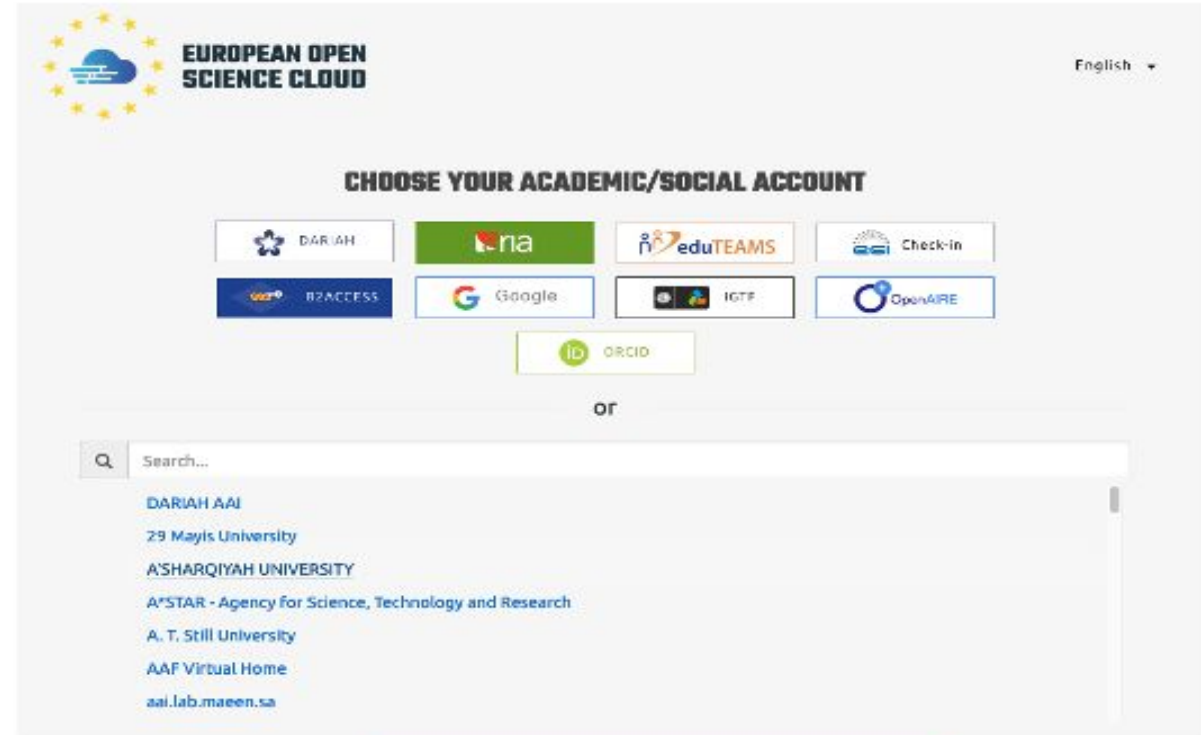
1. Access the sign in page

## Sign in

Sign in with EOSC using existing accounts such as **Google**, **Dariah**, **eduTEAMS** and multiple academic accounts.



2. Click on Sign in with EOSC



3. Choose your identity provider

# Curation - Editing

## Details

### ACCESS

License [Creative Commons Zero v1.0 Universal](#)

### CATEGORISATION

Keyword [research notebook](#) [newspapers](#) [NLP](#) [python](#) [language resources](#)

Language [English](#)

Intended audience [Trainer](#) [Researcher](#) [Student](#)

Resource category [Education & Training](#)

### CONTRIBUTOR

[Europeana](#)  
<https://pro.europeana.eu/ROR>

### CREATOR

[CLARIN ERIC](#)  
<https://www.clarin.eu/>  
[Twitter](#)

[Twan Goosen](#)  
[dblp](#) [ORCID](#)

[Michał Gawor](#)  
[CLARIN ERIC](#)

[GitHub](#)

### Actors

Role	Name
<input type="text" value="Creator"/>	<input type="text" value="CLARIN ERIC"/> <small>Person/group/body who helped create the resource.</small>
<input type="text" value="Creator"/>	<input type="text" value="Twan Goosen"/> <small>Person/group/body who helped create the resource.</small>
<input type="text" value="Creator"/>	<input type="text" value="Michał Gawor"/> <small>Person/group/body who helped create the resource.</small>
<input type="text" value="Contributor"/>	<input type="text" value="Europeana"/> <small>Person/group/body who helped create the resource.</small>

[+ Add Actor](#)

ID Service	Identifier
<input type="text" value="GitHub"/>	<input type="text" value="clarin-eric/europeana-newspaper-notebooks"/>

[+ Add External ID](#)

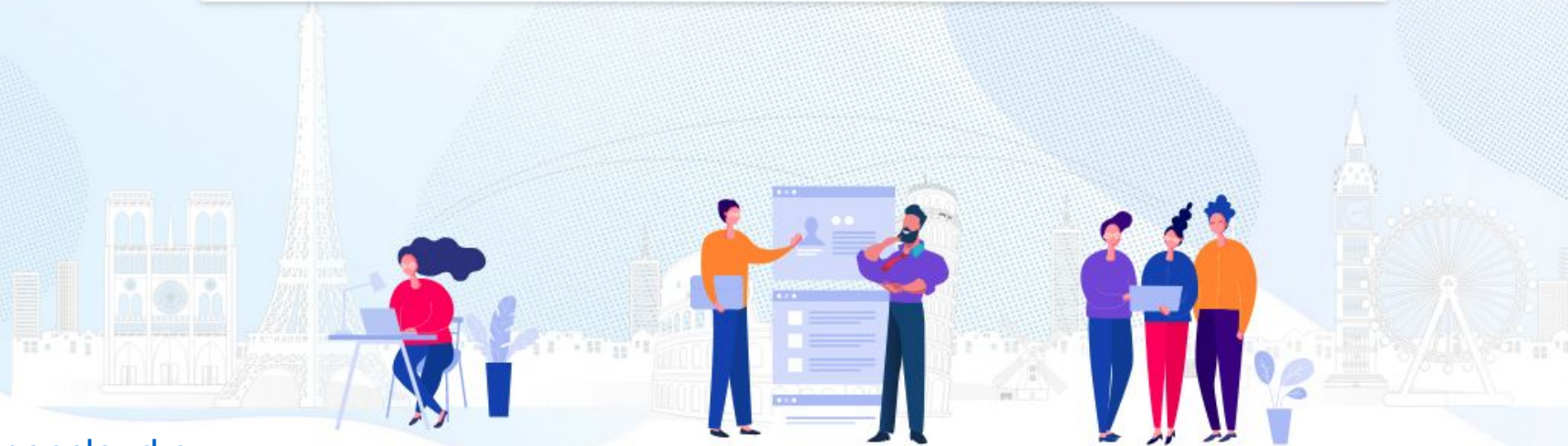
## Properties

Property type	Concept
<input type="text" value="Keyword"/>	<input type="text" value="research notebook"/> <small>Concept or term related to MP entry. See <a href="#">Keywords from SSHOC MP</a>.</small>
Property type	Concept
<input type="text" value="Keyword"/>	<input type="text" value="newspapers"/> <small>Concept or term related to MP entry. See <a href="#">Keywords from SSHOC MP</a>.</small>
Property type	Concept
<input type="text" value="Language"/>	<input type="text" value="English"/> <small>Language(s) in which a resource is available. See <a href="#">ISO 639-3 Language Co</a></small>
Property type	Concept
<input type="text" value="License"/>	<input type="text" value="Creative Commons Zero v1.0 Universal"/> <small>Select license pertaining to resource. See <a href="#">Software License</a>.</small>
Property type	Concept
<input type="text" value="Intended audience"/>	<input type="text" value="Trainer"/> <small>See <a href="#">Intended audience</a>.</small>
Property type	Concept
<input type="text" value="Resource category"/>	<input type="text" value="Education &amp; Training"/> <small>See <a href="#">EOSC Resource Category List</a>.</small>
Property type	Concept
<input type="text" value="Intended audience"/>	<input type="text" value="Researcher"/> <small>See <a href="#">Intended audience</a>.</small>
Property type	Concept
<input type="text" value="Intended audience"/>	<input type="text" value="Student"/> <small>See <a href="#">Intended audience</a>.</small>

## Social Sciences & Humanities Open Marketplace

Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets. [Read more...](#)

All categories



[marketplace.sshopencloud.eu](https://marketplace.sshopencloud.eu)

The SSH Open Marketplace is maintained and will be further developed by three European Research Infrastructures - DARIAH, CLARIN and CESSDA - and their national partners. It was developed as part of the "Social Sciences and Humanities Open Cloud" SSHOC project, European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782.





Thank you!

<https://marketplace.sshopencloud.eu/>

[sshopenmarketplace@sshopencloud.eu](mailto:sshopenmarketplace@sshopencloud.eu)

