SSH Open Marketplace

Public Consultation with the DARIAH Community

The SSH Open Marketplace team

Virtual meeting
3. July 2020
10 AM CEST–11:30 AM CEST

Your speakers

Laure Barbot

European project officer at DARIAH
Coordination Office

laure.barbot@dariah.eu

Nicolas Larrousse
Huma-Num / CNRS
nicolas.larrousse@huma-num.fr

Frank Fischer

Associate Professor for Digital Humanities at
Higher School of Economics, Moscow, and
DARIAH-EU Director

frank.fischer@dariah.eu

Klaus Illmayer

Developer at the
Austrian Academy of Sciences

Klaus.Illmayer@oeaw.ac.at

Stefan Buddenbohm

Project officer, Göttingen State and University Library

buddenbohm@sub.uni-goettingen.de

Clara Petitfils

European Project officer

clara.petitfils@huma-num.fr

Introduction

- Why are we meeting today and what is the objective?
    - (Non public) alpha version of the SSH Open Marketplace on 30 June 2020.
    - We want to involve your DARIAH expertise.
    - You may be valuable influencers and multiplicators.
    - And most important: you are needed as our critical advisers!

- Technical procedures for the webinar (as easy as possible):
    - 10:00–11:30h via this Zoom room (which is being recorded).
    - Raise hand to participate in the discussion, you are allowed to speak by the moderator.
    - Use the chat to post your questions at any time.
    - You may switch on your camera while you speak (but you don't have to).
    - Usual documentation incl. slides will hit your inbox afterwards.

Schedule (10:00–11:30h)

1. Introduction & housekeeping (Laure), 5 min

2. *Interactive: Mentimeter questionnaire, 5 min*

3. **Talk**: A bit of context: DARIAH, SSHOC & EOSC (Frank), 10 min

4. **Talk**: SSH Open Marketplace & walkthrough alpha release instance (Stefan), 10 min

5. *Interactive: Q&A session wrt the walkthrough (Laure/Stefan), 10 min*

6. **Talk**: Curation & Content aspects (Klaus), 10 min

7. **Talk**: Curation & Governance aspects (Clara, Nicolas), 10 min

8. *Interactive: Name your challenge session & discussion (Laure/ Stefan), 20 min*

9. Wrap up (Laure/ Stefan), 5 min

Let's get introduced to one another - a mentimeter questionnaire

Why the SSH Open Marketplace?

- DH & tools registries - Predecessors: TAPoR, DiRT Directory

- DARIAH Strategic Plan 2019–2026
  - "DARIAH As Creator: Build a Marketplace to facilitate fluid exchange of tools, services, data and knowledge"

- European Open Science Cloud (EOSC) as funding opportunity to transfer the idea into a services
- SSHOC as SSH contribution to the EOSC

SSHOC
social sciences & humanities open cloud

Type of action & funding:
Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 45
(20 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Project budget:
€ 14,455,594.08

Project website:
www.SSHopencloud.eu

Duration: 40 months
(January 2019 – 30 April 2022)

Objectives:

- creating the social sciences and humanities **(SSH)** part of European Open Science Cloud **(EOSC)**
- maximising **re-use** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

DARIAH & the SSH Open Marketplace

- Governance & sustainability schema

- DARIAH sources to be linked to the SSH Open Marketplace: DH sources (TAPoR, SSK, Programming Historian, …); in-kind contributions; DARIAH-Campus

- From December 2020: community curation & DARIAH network

The SSH Open Marketplace

- *"An integrated platform that allows easy access to lots of tools, services and datasets for Social Sciences and Humanities. This place is the SSH Open Marketplace."*

- Target groups: Social Sciences AND Humanities and adjacent disciplines; researchers and resource providers
- Content: Open and wherever feasible FAIR in its offerings > frequented updated stock of metadata on the various resource types (tools, services, datasets, tutorials, standards, etc.) coming from already existing sources
- Curation: two lines: (automatic) harvesting and refining AND community-induced curation
- EOSC context: SSH contribution to the EOSC, close collaboration with CESSDA, CLARIN and DARIAH but also ESS, SHARE, E-RiHS, LIBER

Participatory Design & Community Involvement

- *"We collaborate closely with the communities, the future users of the marketplace, researchers, librarians, research infrastructure members, or service providers, policy-makers and citizen scientists, to develop a marketplace that will serve their needs."*

- We try to be: transparent, inclusive, target-group-specific
- SSHOC D7.1 System Specification and SSHOC community on Zenodo
  - User stories, UX design, Feedback on content and curation
- Information about the development progress and constant feedback opportunities
  - Including a dedicated website: Public consultation platform
- Series of webinars/workshops from the very beginning on
- Vertically integrated community involvement group
  - most important wrt the DARIAH context: Eliza Papaki & the rich DARIAH expertise

ALPHA RELEASE: Walkthrough

Timeline for the SSH Open Marketplace

| Jun 2020 | First release of the SSH Open Marketplace is only available to testers! | Dec 2020 | Beta and public release of the SSH Open Marketplace | Dec 2021 | Final release of the SSH Open Marketplace |

- (non public) alpha instance of the marketplace: https://marketplace.sshopencloud.eu/ (please don't share at this time)

- Landing page, Browsing facets, Result list,

  Detail page, Related items

Please note your questions during the walkthrough.
And keep in mind: it's only the alpha.

ALPHA

ALPHA

SSH Open Marketplace
ALPHA
Social Sciences and Humanities Open Marketplace

Landing page

Social Sciences & Humanities Open Marketplace

Discover new resources for your research in Social Sciences and Humanities: tools, services, training materials and datasets, contextualised. Read more...

**SSH Open Marketplace is under development and the current content is subject to change.
Alpha release is planned for June 2020.**

| All Categories ▾ | | Search |

Researchers in political science

Researcher in linguistics

Researchers in literature studies

SSH Open Marketplace ALPHA
Social Sciences and Humanities Open Marketplace

Tools   Training Materials   Datasets   Workflows   Browse | Community   About

Lorem ipsum dolor ✕   Search

Search results (2 600)

Result list

Refine your search   Clear Filters

Sort by name ▾   ‹ Previous   4   z   118   Next ›

CATEGORY ▲

☐ Tools	1 344
☐ Tool	1 341
☑ Software	2
☑ Service	1
☐ Training Materials	159
☑ Training material	156
☑ Paper	1
☐ Tutorial	1
☐ Webinar	1
☑ Workflows	0
☑ Datasets	3
☐ Data Catalogue	2
☐ Dataset	1

ACTIVITY ▲

☐ Recording	2 100
☑ Transcription	567
☐ Discovering	78

Name of the Tool   🔗 Software

Creator: Visually, Inc.
Discipline: Name of the Discipline
Topic: Name of the Topic

Copy the link to the object

Basic informations about the tool, What for is this tool and who is use it. Maximum two lines of the text. Lorem ipsum dolor sit amet, consectetur elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris...

Read more

Name of the Training material   🔗 Online course

Creator: Visually, Inc.
Discipline: Name of the Discipline
Topic: Name of the Topic

Basic informations about the training material, What for is this tool and who is use it. Maximum two lines of the text. Lorem ipsum dolor sit amet, consectetur elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris...

🕐 20m   Read more

Name of the Dataset   🔗 Dataset

Creator: John Smith, John Williams, Dave Taylor
Discipline: Name of the discipline
Topic: Name of the Topic

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

Tools   Training Materials   Datasets   Workflows   Browse   Community   About

Font size   Contrast   Sign up   Log in

Detail page - item

Search

Home / Tools / Sketch2Code (Ver. 1.2)

Sketch2Code (Ver. 1.2)

🔗 Copy Link

Go to Service⧉

Weka (Waikato Environment for Knowledge Analysis) is a free Java-based data mining workbench of machine learning algorithms, offered by the Machine Learning Group of the University of Waikato. It includes tools for data pre-processing, classification, clustering and visualization, among others. Development on Weka began in the early 1990s and it is extensively documented. Read more...

What's new in version 1.1

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Preview

Tool details

Curator: Username

Creator: John Smith (Email, Website⧉)

Discipline: Name of the Discipline

Topic: Name of the Topic

Type: Software

License: Apache License 2.0⧉ (Free; Open Source)

Language: pl

Topic Classification: Psychology; Lorem ipsum

Activity: Annotating

Source: TAPoR⧉

+ Add tags

People also accessed

Software

Visual.ly / Ver 2.1

Easy-to-use web-based software for creating infographics and data visualization, including

Microsoft | Microsoft AI
Learn more at AI Lab
⬆ UPLOAD DESIGN   ⧉ SKETCH2CODE IS AT WORK   ⬇ DOWNLOAD YOUR HTML

Sketch2Code
Transform any hands-drawn design into a HTML code with AI.

Use a sample from gallery or upload your own design

< >

Use this sample

‹ ›

Related

Related Datasets, Tools, Training Materials

oXygen XML editor Tool

Contributors:

A cross-platform XML editor that may be used to create and validate XML documents and associated schema. It fully supports XSL (both XSLT and FO), DTD, Schema (Relax RNG and W3C), Database, XQuery ...

Das DTA-Basisformat in neuem Gewand Training Material

Contributors:

Das DTA-Basisformat (kurz: DTABf) erscheint seit kurzem in zweierlei Hinsicht „in neuem Gewand". Zum einen wurde die Dokumentation vollständig überarbeitet und dabei um verschiedene Features ...

Generating an Ordered Data Set from a Text File Training Material

Contributors:

This tutorial illustrates strategies for taking raw OCR output from a scanned text, parsing it to isolate and correct essential elements of metadata, and generating an ordered data set (a python ...

Extract textual content from images Workflow

Contributors:

How may textual content be extracted from images? A SSK scenario describing in 11 steps a possible workflow. The workflow starts with necessary definitions for text, images, examining other existing ...

Questions from the walkthrough (10 min)

# SSH Open Marketplace (MP) content: sources and technical aspects

MP content: sources

- MP aims to collect **metadata** from different **sources** that are **relevant** for SSH communities

- Preference on sources holding **already curated data**

- **Types of items** to ingest includes: **tools** (software and services), **digital objects** (information objects like publications and training material as well as exemplary datasets), **workflows** (that may be composed of steps)

- **Sources** that we already ingested: TAPoR (tools), The Programming Historian (digital objects), Standardization Survival Kit (workflows), manual collected items (mixture)

- **Current focus** on **well-known sources** where we get the data easily and do have a good mixture of different items to ingest

MP content: data model

- Continuous development of a data model (current version 1.2) = important outcome of SSHOC MP
- We call our main content: entities
- There are **basic fields** that are shared between the entities
- There are **derivations** of the **base entity type** forming specific types
- Additionally, **properties** allow to add more detailed information and enrichment - gives flexibility but on cost of harmonizing efforts (properties ideally backed up by vocabularies like TaDiRAH)
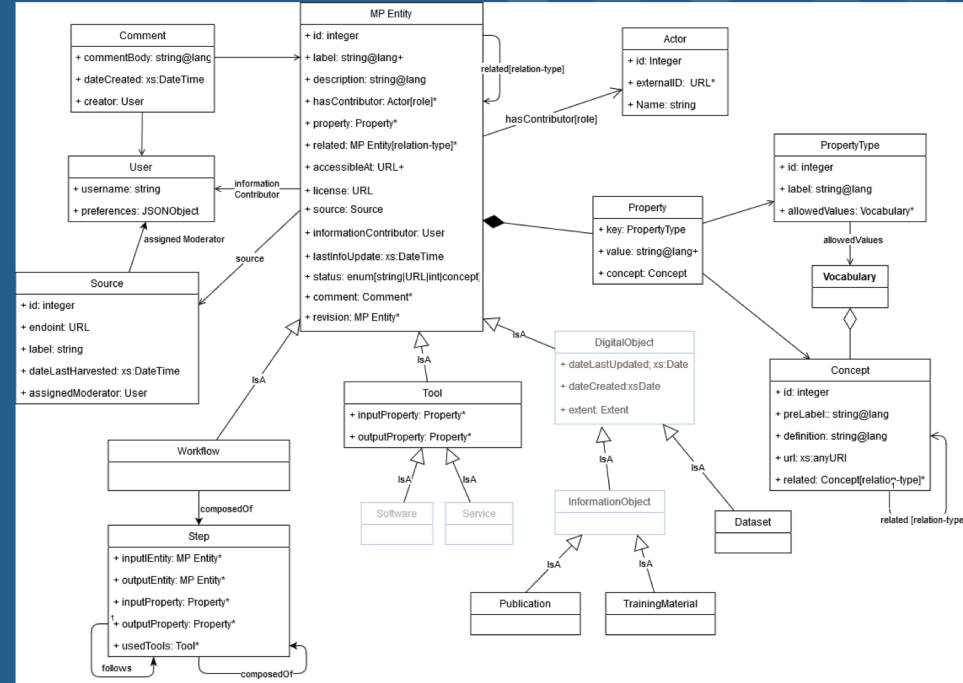
MP content: technical background

- MP is build up on a **decoupled approach**
- REST API is connection point
- Currently frontend and ingestion tools are running against the REST API
- For writing data into the MP you need an authentification
- Reading data is open to all
- **Flexible approach**: all kind of tools possible that are able to communicate with REST API
- Tools can be connected to share results (not shown in illustration)
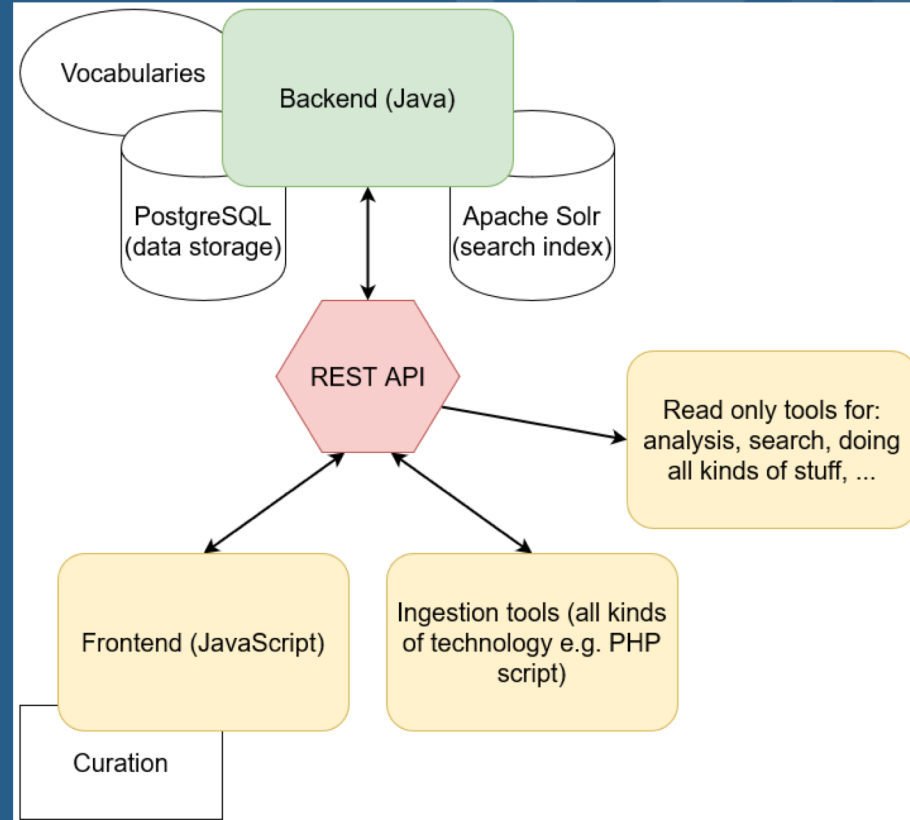
MP content: observations & discussion points

- **Data model** is one key to success: feedback welcome
  - **Flexibility** of data model comes with **data harmonization** efforts
  - Properties can be **expanded easily**, but information not always collected at source
  - Different ways for **interlinking** (relating entities together) are currently in test phase (in the end an important outcome but also a great challenge)
- **Ingested sources** also essential: recommendations welcome
  - Sources to be ingested: DARIAH Campus, dariahTeach, OpenAIRE, EOSC marketplace, Switchboard, DH conferences, TERESAH, Awesome lists, ...
  - Challenges are in identifying and handling of **overlaps**
  - Clever **enrichment** necessary
  - Connection to **authority files** one of the next tasks (e.g. Wikidata)
- **Technical setup** well thought out and tested
  - **Adjustments** will be done **continuously** (e.g. integration of vocabularies)
 - Things will get exciting with new tools using the API

# SSH Open Marketplace (MP) curation: workflows and challenges

MP curation: workflow mapping

- Mappings necessary to ingest data from a source
- Finding overlaps between data model of source and MP data model
- Basic information usually available (title, contributor, url, …)
- Additional information can be integrated with MP properties (for better curation effort new properties needs to be registered)
- Still, data wrangling is necessary
- Resolving against vocabularies as an ongoing effort
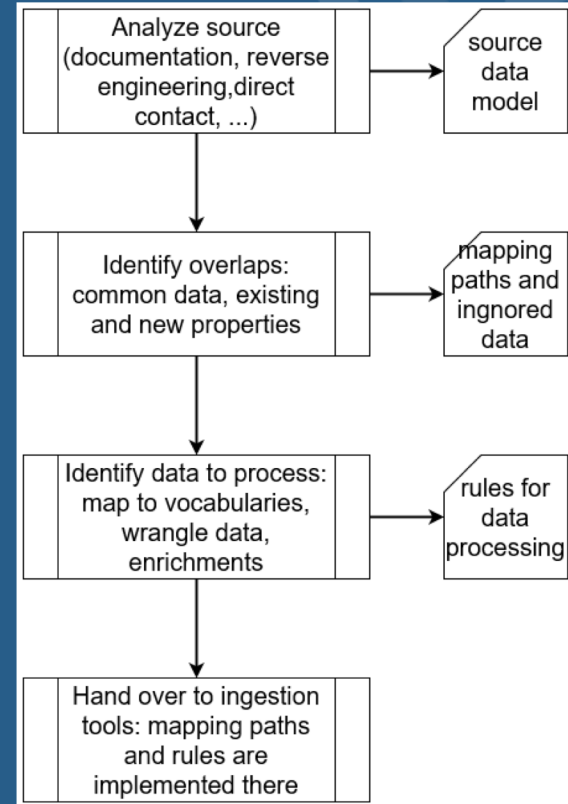
MP curation: workflow ingestion

- Ingestion of data done with the help of dedicated tools: PoolParty Semantic Suite and custom script
- Harvesting data from different types of source: API, database dump, GitHub repository, spreadsheet, web scraping
- Ingestion against MP REST API
- Backend delivers tailored API calls
- Basic enrichment as part of ingestion, e.g. creating information on authors of entities
- Ingestion creates first version, curation then takes over
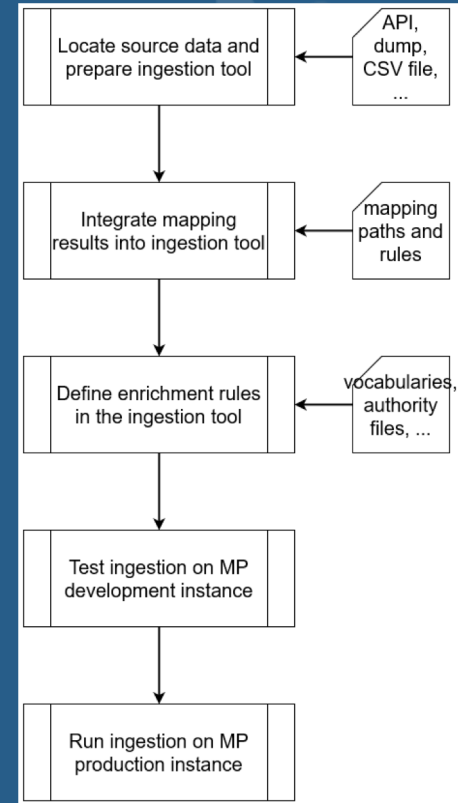
MP curation: challenges mapping and ingestion

- Workflows build up on well established practice in digital SSH domain
- Sharing same issues: reconciliation of data, different versions of truth about an entity, handling of bad data quality
- High quality of mapping and ingestion crucial for users (gives first impression of MP) and for curators (building up on dense information makes curation easier)
- Mapping and ingestion as entry point for preparing FAIRness of data
- Curation efforts are still necessary and time consuming
- Mapping and ingestion as evolving process that feeds the guidelines for curators (sharing experience of intensive study of the data from the source)

Governance & Curation (10 minutes)

How to better sustain the SSH Open Marketplace and ensure its quality contextualisation, curation and community-driven approach ?

A quick update on the SSH Open Marketplace Governance :
The "Macro-Governance"

→ Governance hypothesis chosen : ERICs to handle the Marketplace, at least one year after the project's end.
→ Three governing bodies :

1) Governing Board (GB) : committed to support the sustainability of the SSH Open Marketplace as its ultimate/decision-making body.
2) Editorial Board (EB) : committed to the SSH Open Marketplace editorial policy (curation)
3) Community Forum (CF) : ensures the visibility of SSH research communities and practices within the Marketplace.

⇒ Community-driven added value : CF + a dedicated role in the curation process.

Governance & Curation

How to foster community participation in the curation process?
The "Micro-Governance" / Curation

Two curation roles :
1) Curator : able to suggest new content in the discovery platform but also to enrich existing items.
2) Chief Editor/Moderator : a dedicated member of the SSH Open Marketplace's editorial board, which missions and objectives are being developed at the moment.

Curation criteria for a curator: SSH scope, Usage in research communities, Up-to-date/maintained, Openness, FAIRness, Type of licence etc.

How to ensure a quality curation process aside from curator's involvement ?

The possible roles of the Chief Editor/Moderator:
- Makes the final decision on publication and/or depublication
- Check the consistency of the content with the help of different expertises from the editorial board
- Ensures compliance with the editorial policy
- Makes the connection with the Governing Board (e.g. define the editorial policy)

Governance & Curation

The Editorial Board

Content curation is one of the assets of the Marketplace!

Some curation features to guarantee its success:
- Freshness and Accuracy of information
- Community driven
- Curation made by an editorial board integrating diverse skills (both technical and disciplinary)
- Links with other resources (e.g. WikiData)

Curation plays definitely a great role in the **sustainability** of the Marketplace

Governance & Curation

A global view

Community Forum

Feedback

Users & Curators

Governing Board

Propositions

Editorial Board

Operation

Usage
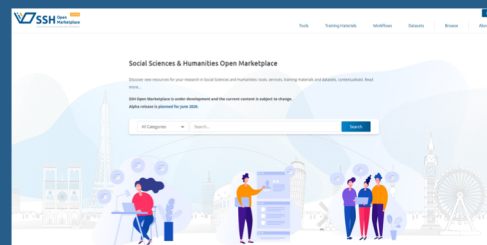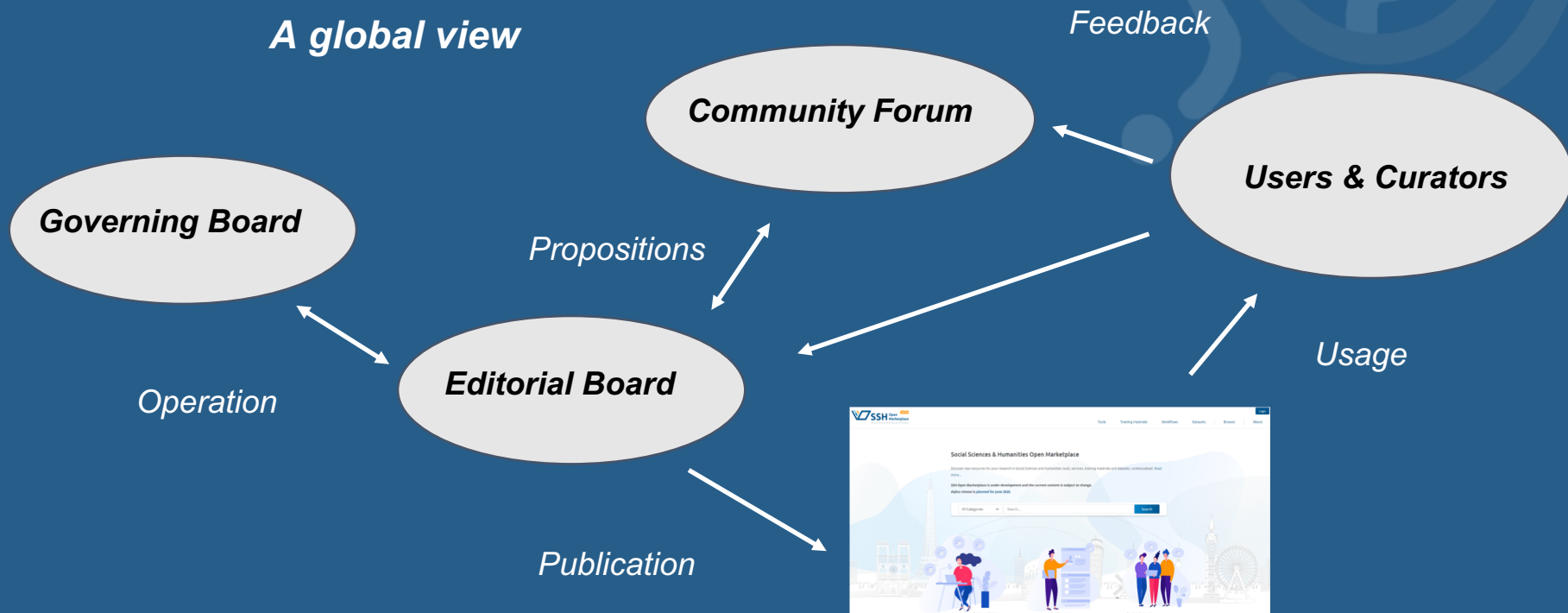
Publication

SSHOC Marketplace

Save your questions for a moment - Let's start the name your challenge session! (20 min)

Name your challenge!

- This webinar is a public consultation for the DARIAH community. Therefore:
  - We want an open and honest discussion about the success and risk factors for the SSH Open Marketplace (some ideas for such risks next slide).
  - We will focus and re-adjust our upcoming work considering your feedback.
  - But: Please consider that we presented today the alpha release instance and not the final product.
- However: IF the service is lacking essential functions, UX or content that would hamper the usefulness for you personally, please let us know!
- 2 minutes of thinking or use of the chat and then we discuss the challenges in the group.
- We hope for a vivid discussion!

Examples for challenge areas (success & risk factors)

➜ **Mission statement or unique selling proposition unclear** > *We don't hope!*

➜ **Lack of critical amount of content or sources** > *Of course!*

➜ Insufficient instruments for aggregation, curation or governance

➜ **Visibility, community-embedment or reputation of the service** - *Is there a too much?*

➜ Strange design of the user interface

➜ Lack of browsing facets or even other browsing functions

➜ Lack of metadata fields, e.g. details on a tool

Wrap up (5min)

- Final questions - go to menti.com

- If you haven't done it yet: think about registering as a tester

- Do not hesitate to get in touch with DARIAH & partners involved in the SSHOC day-to-day activities!

SSHOC
social sciences & humanities open cloud

https://www.sshopencloud.eu/ssh-open-marketplace

Thank you for your attention!

Slides and a recording will be sent to all registered participants.

Join our community

https://www.sshopencloud.eu

@SSHOpenCloud

info@sshopencloud.eu

/in/sshopencloud