Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782   Start Date of Project: 01/01/2019   Duration: 40 months

Deliverable 6.4 Report on awareness webinars

Dissemination Level	PU
Due Date of Deliverable	28/02/2022 (M38)
Actual Submission Date	15/03/2022
Work Package	WP 6 - Fostering Communities, Empowering Users, & Building Expertise
Task	Task 6.2 Fostering Communities: Engaging New & Existing Users
Type	Deliverable
Approval Status	Approved by EC - 27 April 2022
Version	V1.0
Number of Pages	p.1 – p.84

Abstract: The report summarises the ten awareness-raising webinars, organised between March 2020 and April 2021 to and foster communication and inform the academic community in social sciences and humanities of the developments within the SSHOC project.

History

Version	Date	Reason	Revised by
0.1	25/01/2022	Initial version	Ana Inkret, Irena Vipavc Brvar
0.2	31/01/2022	Internal review	Rosie Allison, Judith Wehmeyer
0.2	03/02/2022	Peer review	Hina Zahid
0.3	14/02/2022	Addressed comments, internal review	Ana Inkret, Astrid Verheusen
0.4	14/02/2022	Addressed comments, final formatting	Ana Inkret
1.0	15/02/2022	Final version for submission	Ana Inkret

Author List

Organisation	Name	Contact Information
CESSDA/UL-ADP	Ana Inkret	ana.inkret@fdv.uni-lj.si
CESSDA/UL-ADP	Irena Vipavc Brvar	irena.vipavc@fdv.uni-lj.si

List of Contributors

Organisation	Name	Contact Information
DARIAH-EU	Laure Barbot	laure.barbot@dariah.eu
CLARIN ERIC	Iulianna van der Lek	iulianna@clarin.eu
CLARIN/UL-FF	Kristina Pahor de Maiti	kristina.pahordemaiti@ff.uni-lj.si
CESSDA/GESIS	Veronika Keck	veronika.keck@gesis.org

Executive Summary

The awareness-raising webinars were part of the community engaging activities of task 6.2 *Fostering Communities: Engaging New & Existing Users*. The goal was to inform the community about the development of SSHOC services and tools, consult with the community members, and showcase practical use cases. Webinars were organised in close cooperation with other work packages within the SSHOC project.

Overall, ten awareness webinars took place:

1. Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository
2. How to improve the quality of your repository? SSHOC and certification of repositories
3. SSH Open Marketplace: Public Consultation for the DARIAH Community
4. SSHOC Online Information Sessions: Open-Source Vocabulary Hosting and Management Platforms - First session: Wikibase
5. SSHOC Online Information Sessions: Open-Source Vocabulary Hosting and Management Platforms - Second session: SKOSMOS
6. DARIAH Community Requirements for a Dataverse Repository
7. SSHOC Online Information Sessions: Open-Source Vocabulary Hosting and Management Platforms - Third session: CESSDA Vocabulary Service
8. Building trustworthy repositories: Introduction to CoreTrustSeal certification
9. Best Practices for Preserving Oral Archives
10. The Multilingual Corpus of Survey Questionnaires

The concise and targeted webinars were an effective tool for communication with specific community groups. In the portfolio of SSHOC awareness-raising events, the webinars are distinguished by their focus on specific subjects and target audiences and were in some instances an introduction for topics that were later explored in workshops or stakeholder events.

Abbreviations and Acronyms

EOSC	European Open Science Cloud
GA	Grant Agreement
SSHOC	Social Science and Humanities Open Cloud
SSH	Social Science and Humanities
WP	Work package
WP2	Work package 2 - Communication, Dissemination, and Impact
WP3	Work package 3 - Lifting Technologies and Services into the SSH Cloud
WP4	Work package 4 - Innovations in Data Production
WP5	Work package 5 - Innovations in Data Access
WP6	Work package 6 - Fostering Communities, Empowering Users, & Building Expertise
WP7	Work package 7 - Creating the SSH Open Marketplace
T	Task
T6.2	Task 6.2 Fostering Communities: Engaging New & Existing Users

Table of Contents

1. Introduction

The awareness-raising webinars were part of the community-engaging activities of task 6.2 *Fostering Communities: Engaging New & Existing Users*. The aim of the webinars was to keep the community up to date on the development of specific services and tools as well as to showcase practical use cases within the SSHOC project. Complementing the awareness workshops[1] and stakeholder series events[2], the organisation of webinars contributed to the engagement of target communities and to the wider WP6 - *Fostering Communities, Empowering Users and Building Expertise* goals of broadening the SSHOC network of user communities and empowering them with knowledge to utilise the SSHOC services, tools and data throughout the research lifecycle and according to the FAIR Principles.

The role of awareness webinars had previously been defined by *Deliverable 6.1 Community Engagement Strategy* (Torma et al. 2019). The first organisational steps were documented in *Milestone 38 Launch Community engagement strategy* (Vipavc Brvar 2019).

2. Organisation

The awareness webinars were planned as short informative events, meant to share information and consult with the community. As the specialised SSHOC tools and services required relevant feedback from their future users, the webinars catered to specific communities and stakeholder groups[3] rather than reaching across stakeholder groups like the discussions and workshops of the stakeholder series.

The organisation of the webinars depended closely on the SSHOC project partners, just as the other event formats within the task. The work began with the identification of potential products, services or other developments that would be of interest to members of the SSH community. The task partners closely followed the progress of work in other work packages, together with WP2 - *Communication, Dissemination, and Impact*, and communicated regularly with the task leaders. All WPs were offered the opportunity to present their work through the webinars (and other awareness-raising events). The organisation was initiated either by the partners of T6.2 who wished to raise awareness of certain aspects or developments of the project, or by project partners who wanted to share updates of their work and receive feedback from a specific community.

The content of the webinars was under the jurisdiction of the project partners who were involved in the development of the tools or services, while the dissemination activities were supported by partners in

---

[1] See *Deliverable 6.3 Report on Awareness Workshops* (Vipavc Brvar & Inkret 2021) for more information.
[2] See *Deliverable 6.5 Report on Stakeholder Series events* (Vipavc Brvar & Inkret 2021b) for more information.
[3] It should however be noted that this did not impact the broad dissemination of news and materials of the webinars.

WP2. Partners of T6.2 acted as a facilitator of engagement with the community by organising the dissemination together with WP2, supporting the planning of the content, providing the technical infrastructure, recording and documenting the events, and finally by publishing the event materials in collaboration with WP2 and preparing reports that are part of this deliverable (see annexes).

Most webinars were self-standing events. Only two were organised as part of established community meetings: at the DARIAH annual event[4] (see section 3.8) or as part of the CLARIN Café series[5] (see section 3.9).

Webinars are inherently digital, online events which meant that the organisation was minimally affected by the pandemic of COVID-19 when compared to other awareness events. The webinars did however face competition and the so-called "Zoom fatigue" when the ban on travelling and face-to-face events brought a surge of online events. The sheer numbers of online events and other restrictions of work during the lockdowns could have impacted the number of participants. In response, the webinars were kept concise, lasting from 1 hour to a maximum of 1.5 hours. They were strictly tailored to specific target communities and thus kept highly relevant for the participants. This sometimes resulted in a series of webinars on the same topic, such as the CoreTrustSeal webinars (see sections 3.1 and 3.8 below) that were adapted for the respective audiences, or the Vocabulary Platforms series (see sections 3.4, 3.5, and 3.7 below), which broke the broad topic into three short events that were easier to join and follow than a single exhaustive event. Finally, the webinars seemed to give way to awareness workshops and conference sessions where the participants could be engaged even further, and the topics could be explored and discussed in depth. Another reason for that is the level of maturity that the SSHOC work reached by mid-2021. The brief and informative mode of the webinars inevitably gave way to the more interactive and immersive experience of the workshops.

3. Summary of the webinars

A total of 10 webinars took place between March 2020 and April 2021. In total, about 400 people participated in the events.[6] The distribution of participants per SSHOC stakeholder categories is:

- Research libraries and archives: 141
- Universities and research performing institutions: 80
- Researchers: 51
- Research and e-infrastructures, EOSC thematic clusters: 50
- Private sector and industry: 8

---

[4] DARIAH Annual event, https://annualevent.dariah.eu/archive/, accessed 25 January 2021.
[5] CLARIN Café, https://www.clarin.eu/content/clarin-cafe, accessed 25 January 2021.
[6] Precise numbers could not be documented (see Annexes 4 and 6 for details).

- Civil society and citizen scientists: 2
- Uncategorised: 67

These numbers do not accurately reflect the reach of the webinars, however, as the event materials were published and disseminated within the SSHOC community and continue to be circulated.

The sections below note the key elements of the webinars while full reports are annexed at the end of the deliverable.

3.1 Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository

Date	18 March 2020
Topic	Tailoring of the Dataverse repository service for SSH institutions in Europe
Links to materials	
Announcement	https://sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse
Presentations	https://doi.org/10.5281/zenodo.3714988
Video recording	https://youtu.be/f7-r-80M-Fk
Q&A	https://sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse
Blog post	https://sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse
Audience	
Targeted audience	Staff of CESSDA service providers
Total participants	46

By stakeholder category	Research libraries and archives 46
Event report	*Annex 1*

3.2 How to improve the quality of your repository? SSHOC and certification of repositories

Date	23 April 2020
Topics	Certification of digital repositories and SSHOC support for repositories seeking certification
Links to materials	
Announcement	https://sshopencloud.eu/sshoc-webinar-repositories-quality-certification
Presentations	https://doi.org/10.5281/zenodo.3774396
Video recording	https://youtu.be/Ev0KRXkH4jo
Blog post	https://sshopencloud.eu/news/sshoc-webinar-certification-data-repositories
Audience	
Targeted audience	Staff of data repositories
Total participants	61
By stakeholder category	Research libraries and archives 27 Research and e-infrastructures, EOSC thematic clusters 18

	Universities and research performing institutions 11 Researchers 5
Event report	*Annex 2*

3.3 SSH Open Marketplace: Public Consultation for the DARIAH Community

Date	3 July 2020
Topic	Alpha release of the SSH Open Marketplace
Links to materials	
Announcement	https://www.sshopencloud.eu/events/ssh-open-marketplace-public-consultation-dariah-community
Presentations	https://doi.org/10.5281/zenodo.3935345
Video recording	https://youtu.be/IQSdQbaA6oY
Blog post	https://sshopencloud.eu/news/ssh-open-marketplace-notes-our-public-consultation-dariah
Audience	
Targeted audience	DARIAH community
Total participants	50
By stakeholder category	Universities & Research Performing Organisations 25 Researchers 11 Research libraries and archives 9 Research and e-infrastructures, EOSC thematic clusters 3 Private Sector and Industry 2

	Other 2
Event report	*Annex 3*

3.4 SSHOC Online Information Sessions: Open-Source Vocabulary Hosting and Management Platforms - First session: Wikibase

Date	3 September 2020
Topics	Wikibase software
Links to materials	
Announcement	https://www.sshopencloud.eu/sshoc-online-information-sessions-open-source-vocabulary-hosting-and-management-platforms-first
Series overview	https://www.clarin.eu/event/2020/sshoc-online-information-sessions-open-source-vocabulary-hosting-and-management
Presentations	https://doi.org/10.5281/zenodo.4153783 https://doi.org/10.5281/zenodo.4153771
Video recording	https://youtu.be/ZKaXhi6MOnw
Blog post	https://sshopencloud.eu/news/takeaways-sshoc-webinar-series-open-source-vocabulary-hosting-and-management-platforms
Audience	
Targeted audience	Academic or industry professionals who operate, integrate, and manage vocabulary platforms
Total participants	41

By stakeholder category[7]	Universities and research performing institutions 35.6% Research libraries and archives 30.5% Research and e-infrastructures, EOSC thematic clusters 20.3% Researchers, civil society and citizen scientists, private sector and industry 13.6%
Event report	*Annex 4*

3.5 SSHOC Online Information Sessions: Open-Source Vocabulary Hosting and Management Platforms - Second session: SKOSMOS

Date	14 September 2020
Topics	SKOSMOS tool
Links to materials	
Announcement	https://www.sshopencloud.eu/sshoc-online-information-sessions-open-source-vocabulary-hosting-and-management-platforms-second
Series overview	https://www.clarin.eu/event/2020/sshoc-online-information-sessions-open-source-vocabulary-hosting-and-management
Presentations	https://doi.org/10.5281/zenodo.4154123 https://doi.org/10.5281/zenodo.4154016 https://doi.org/10.5281/zenodo.4154147
Video recording	https://youtu.be/zwZsI4JTnHU

---

[7] Data on participants in the SSHOC Online Information Sessions was aggregated. Numbers per stakeholder categories for all three sessions are: research libraries and archives 33, universities and research performing organisations 29, research and e-infrastructures, EOSC thematic clusters 13, uncategorised 10, private sector and industry 5, researchers 5, civil society and citizen scientists 1.

	https://youtu.be/PcYvDXJX6vY https://youtu.be/RXscwBc5UV0
Blog post	https://sshopencloud.eu/news/takeaways-sshoc-webinar-series-open-source-vocabulary-hosting-and-management-platforms
Audience	
Targeted audience	Academic or industry professionals who operate, integrate and manage vocabulary platforms
Total participants	61
By stakeholder category	Universities and research performing institutions 35.6% Research libraries and archives 30.5% Research and e-infrastructures, EOSC thematic clusters 20.3% Researchers, civil society and citizen scientists, private sector and industry 13.6%
Event report	*Annex 4*

3.6 DARIAH Community Requirements for a Dataverse Repository

Date	28 September 2020
Topics	Tailoring of the Dataverse repository service for DARIAH and wider SSH community
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-webinar-dariah-community-requirements-dataverse-repository
Presentations	https://doi.org/10.5281/zenodo.4054847

Video recording	https://youtu.be/wiPg6yaI8Rk
Blog post	https://www.sshopencloud.eu/news/webinar-notes-dataverse-development-sshoc
Audience	
Targeted audience	Researchers and institutes across the SSH community
Total participants	33
By stakeholder category	Research libraries and archives 12 Universities and academic institutions 6 Researchers 3 Research and e-infrastructures, EOSC thematic clusters 3
Event report	*Annex 1*

3.7 SSHOC Online Information Sessions: Open-Source Vocabulary Hosting and Management Platforms - Third session: CESSDA Vocabulary Service

Date	30 September 2020
Topics	Vocabulary service of Consortium of European Social Science Data Archives (CESSDA)
Links to materials	
Announcement	https://www.sshopencloud.eu/sshoc-online-information-sessions-open-source-vocabulary-hosting-and-management-platforms-third
Series overview	https://www.clarin.eu/event/2020/sshoc-online-information-sessions-open-source-vocabulary-hosting-and-management

Presentations	https://doi.org/10.5281/zenodo.4154428 https://doi.org/10.5281/zenodo.4154426
Video recording	https://youtu.be/EFcr8drCY8Q https://youtu.be/DhWMqJThxY8 https://youtu.be/iwDuM1PBulA
Blog post	https://sshopencloud.eu/news/takeaways-sshoc-webinar-series-open-source-vocabulary-hosting-and-management-platforms
Audience	
Targeted audience	Academic or industry professionals who operate, integrate and manage vocabulary platforms
Total participants	57
By stakeholder category	Universities and research performing institutions 35.6% Research libraries and archives 30.5% Research and e-infrastructures, EOSC thematic clusters 20.3% Researchers, civil society and citizen scientists, private sector and industry 13.6%
Event report	*Annex 4*

3.8 Building trustworthy repositories: Introduction to CoreTrustSeal certification

Date	14 October 2020
Topics	CoreTrustSeal certification requirements and procedures
Links to materials	

Announcement	https://sshopencloud.eu/building-trustworthy-repositories-introduction-coretrustseal-certification
Main event homepage	https://dariah-ae-2020.sciencesconf.org/
Presentations	https://doi.org/10.5281/zenodo.4233672 https://doi.org/10.5281/zenodo.4233726 https://doi.org/10.5281/zenodo.4233873 https://doi.org/10.5281/zenodo.4233994
Video recording	https://youtu.be/8SUA8XbM2RA
Audience	
Targeted audience	Data repository managers
Total participants	20
By stakeholder category	Researchers 10 Research libraries and repositories 6 Research and e-infrastructures, EOSC thematic clusters 4
Event report	*Annex 2*

3.9 Best Practices for Preserving Oral Archives

Date	24 February 2021
Topics	Preservation of oral sources
Links to materials	

Announcement	https://sshopencloud.eu/events/sshoc-webinar-best-practices-preserving-oral-archives
Presentations	https://doi.org/10.5281/zenodo.4647718
Video recording	https://youtube.com/playlist?list=PLpuwwe1MpnqIC-3c-KFflAUdKvvKuvg82
Audience	
Targeted audience	Researchers, lecturers, students, and topic experts
Total participants	45
By stakeholder category	Researchers 17 Research and e-Infrastructure, EOSC thematic clusters 9 University and research performing organisations 9 Research libraries and archives 8 Private sector and industry 1 Civil society and citizen scientists 1
Event report	*Annex 5*

3.10 The Multilingual Corpus of Survey Questionnaires

Date	6 April 2021
Topics	Characteristics and showcase applications of the Multilingual Corpus of Survey Questionnaires
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-webinar-multilingual-corpus-survey-questionnaires

Presentations	https://doi.org/10.5281/zenodo.4696180
Blog post	https://sshopencloud.eu/news/multilingual-corpus-survey-questionnaires-mcsq-how-social-scientists-can-benefit-corpus
Audience	
Targeted audience	Academics and practitioners from the fields of Sociology, Communication Sciences, Lexicology, Linguistics and Translation Studies, interested in international comparable surveys, survey research and methods, and translation of questionnaires
Total participants	55
Event report	*Annex 6*

4. Outcomes and Conclusions

The short, concise, and informative webinars were an effective way of communicating with the future users of SSHOC services and tools. During the influx of online events, the webinars proved to be a preferred format for organisers and participants alike. The organisers assured that the participants were actively involved in each of the events, that their feedback during the events was documented, and that their input contributed to future developments of tools and services.

The webinars are part of a rich portfolio of awareness and engagement events, where they are distinguished by their focus on specific subjects and target audiences, whereas the events of the stakeholder series addressed broader topics and succeeded in bringing together different stakeholder groups across scientific disciplines. The webinars were attended by highly engaged target groups with a keen interest in the presented topics.

5. References

Irena Vipavc Brvar, & Ana Inkret. (2021). D6.3 Report on Awareness Workshops. Zenodo. https://doi.org/10.5281/zenodo.5913518

Irena Vipavc Brvar, & Ana Inkret. (2021b). D6.5 Report on Stakeholder Series events. Zenodo. https://doi.org/10.5281/zenodo.5638651

Torma, Martina, Kalaitzi, Vasso, Vipavc Brvar, Irena, Fišer, Darja, Pahor de Maiti, Kristina, Petitfils, Clara, Durco, Matej, Grant, Friedel, & Willems, Marieke. (2019). SSHOC D6.1 Community Engagement Strategy (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3592244

6. List of Annexes

Annex 1: Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository and DARIAH Community Requirements for a Dataverse Repository - Webinar Report

Annex 2: How to improve the quality of your repository? SSHOC and certification of repositories and Building trustworthy repositories: Introduction to CoreTrustSeal certification - Report of the Webinars

Annex 3: SSH Open Marketplace: Public Consultation for the DARIAH Community - Webinar Report

Annex 4: Report on the SSHOC Webinar Series on Open-source Vocabulary Hosting and Management Platforms

Annex 5: Best Practices for Preserving Oral Archives - Webinar Report

Annex 6: The Multilingual Corpus of Survey Questionnaires - Webinar Report

ANNEX 1

DISCUSSION MEETING ABOUT REQUIREMENTS OF CESSDA SERVICE PROVIDERS FOR A DATAVERSE REPOSITORY AND DARIAH COMMUNITY REQUIREMENTS FOR A DATAVERSE REPOSITORY

WEBINAR REPORT

By Ana Inkret, CESSDA/UL-ADP

Background

The two webinars covered by this report – Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository, that took place on 18 March 2020, and DARIAH Community Requirements for a Dataverse Repository, that took place on 28 September 2020 – were organised in cooperation of T5.2 *Hosting and sharing data repositories* and T6.2 as an awareness-raising activity and a communal discussion of the adapted functionalities.

Webinars presented the developing data repository service, coordinated by T5.2, and gathered feedback from CESSDA and DARIAH communities. Webinars were organised by CESSDA/ADP and LIBER.

Discussion Meeting About Requirements of CESSDA Service Providers for a Dataverse Repository

Webinar Overview & Format

Aim. Task 5.2 is developing a data repository service on EOSC for SSH institutions. Service is based on the open source, community driven platform Dataverse. Task 5.2 is adapting the software to the needs and requirements of European research infrastructures. Besides the presentation of the current functionality and the features under development, the goal of the webinar was to collect input from the CESSDA Service Providers, focusing on the essential requirements for a data repository service, preferences regarding installation, organisation, and extra functionalities, as well as necessary training.

Speakers. Webinar was delivered by Marion Wittenberg, service manager of DataverseNL, Laura Huis in 't Veld, functional manager, and Vyacheslav Tykhonov, data scientist, all of Data Archiving and Networked Services (DANS), partner in T5.2.

Organisers. Webinar was organised by LIBER and CESSDA/UL-ADP, partners in T6.2.

Participants. Webinar was intended for staff of CESSDA service providers. 46 participants in total attended the webinar. 36 of these were representatives of member organisations, while 6 were associated with the CESSDA partners. Participants were part of the research libraries and archives stakeholder group.

Brief summary of the event structure. Marion Wittenberg started the webinar with a short introduction of the SHOC project and the objectives of task 5.2. In the main session, select functionalities were presented. Each presentation was followed by Q&A. Webinar lasted for 90 minutes.

Presentations & Discussions: Key Points

Introductory session: Introduction to SSHOC project, task 5.2 and Dataverse
Speaker. Marion Wittenberg
Main points. Session introduced the SSHOC project, the role of Task 5.2, and the Dataverse software, including a demonstration of the interface. Dataverse has several advantages over other repository platforms, such as adaptability, the fact that it is suitable for all kinds of data, has a persistent identifier on both study and file level, and allows for a range of user permissions.
Links to materials:
- Presentation
- Video

Main session: SSHOC Dataverse: overview of functionality, configuring Dataverse for CESSDA, Weblate introduction
Speakers. Vyacheslav Tykhonov, Laura Huis in 't Veld
Main points. To be accepted as a service in EOSC, software must be at least level 8 on CESSDA's Software Maturity Model, meeting high technical demands. SSHOC is developing Weblate service, a tool for multilingual translations for the user interface, metadata schema and Solr. External controlled vocabularies support will promote interoperability. Data previewers for a wide range of formats will also be developed, as well as a data processing tool to migrate NESSTAR DDI to Dataverse. To configure Dataverse for CESSDA, CESSDA Mandatory Metadata fields must also be made available in Dataverse. Session included time for questions (see attachment).
Link to materials:
- Presentation

- [Video](#)

Dariah Community Requirements for a Dataverse Repository

Webinar Overview & Format

Aim. The goal of the webinar was to share ideas and requirements of SSH researchers and research institutes across the community, not limited to the DARIAH community. Team presented current and developing functionalities and collected feedback from the audience.

Speakers. Webinar was delivered by [Marion Wittenberg](#), Data Archiving and Networked Services (DANS), and [Peter Kiraly](#), software developer and researcher at Georg-August-Universität Göttingen, partners in T5.2.

Organisers. Webinar was organised by LIBER and CESSDA/UL-ADP, partners in T6.2.

Participants. Webinar was intended, but not limited to the DARIAH community. 33 participants in total attended the webinar. According to the survey that took place during the webinar, majority (12 out of 25 respondents) belonged to the research libraries and archives stakeholder group, with a minority representing universities and academic institutions (6), researchers (3) and research infrastructures (3) groups.

Presentations & Discussions: Key Points

Introductory session: Introduction to SSHOC project and task 5.2
Speaker. Marion Wittenberg
Main points. After a short introduction of the speakers and a Mentimeter survey, SSHOC project and Dataverse software were introduced. Dataverse as an open-source repository software is supported and developed by a large and diverse community. Task 5.2 is developing a mature software pipeline to adapt it to the European RI.
Link to materials:
- [Presentation](#)
- [Video](#)

Main session: Research Data Management, Dataverse Demo
Speakers. Péter Király
Main points. Archiving and sharing data with Dataverse contributes to reproducibility and long-term usability of data. Main elements of Dataverse software were explained, presented and demonstrated in

the Harvard Dataverse and Texas Data Repository Dataverse. Demonstration included instructions for how to edit data and provide rights. Session ended with Q&A.

Link to materials:

- [Presentation](#)
- [Video](#)

Outcomes & Feedback

Webinars were an opportunity to reach out to the community of potential users and software adopters. A survey before (for the first webinar) or during (for the second) the event gathered information on previous experiences with Dataverse, required functionalities, and planned use and adjustment of the service. Most of the participants had little knowledge of Dataverse, but they were interested in testing out the functionalities developed within SSHOC. In addition, the majority of the participants also expressed the need for the translation of the user interface to their national language.

Both webinars included an intensive Q&A session. Questions and answers were edited and published on the event pages.[8] Videos were also made public and were viewed 198 times on SSHOC YouTube account as of 1 April 2021.

Further webinars on the Dataverse functionalities are planned, tailored for specific communities or to highlight certain features.

Post-event survey for the Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository

Overall, how would you rate the webinar? (1 – poor, 5 – excellent)

5

4

5

4

Where did you learn about the webinar?

mailinglist

mailinglist

---

[8] See Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository, https://sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse (accessed 14 february 2022), and DARIAH Community Requirements for a Dataverse Repository, https://sshopencloud.eu/events/sshoc-webinar-dariah-community-requirements-dataverse-repository (accessed 14 February 2022).

Colleagues
mailinglist

How well organised was the webinar (in terms of time management, length, venue)?
5
4
5
5

What did you hope to gain from the webinar?
Answers to specific questions, and any other relevant information
some overview about dataverse and SSHOC project
learn more about dataverse
information on how 5.2 affects CESSDA and the metadata schemas of CESSDA and how they are interlinked and if there is something the CESSDA MDO needs to do concerning the dataverse metadata and the CESSDA metadata.

Did the webinar meet your expectations?
4
3
3
4

Do you see this webinar having a positive impact on your work and how?
Yes
yes
yes, networking
Yes, it was good to know that the project takes the CMM into account.

What did you like most about the webinar?
Very professionally executed
the presentations and the "openness"
interactions
the webinar was very informative and well organized despite the problems that came with the Corona-lockdowns. But what impressed me most was the speed with which the material (including the recording) of the webinar were made open for public. Great job

What did you miss or could be improved at the webinar?
too many people and it was not possible to better discuss some points
less details and more general information on the software
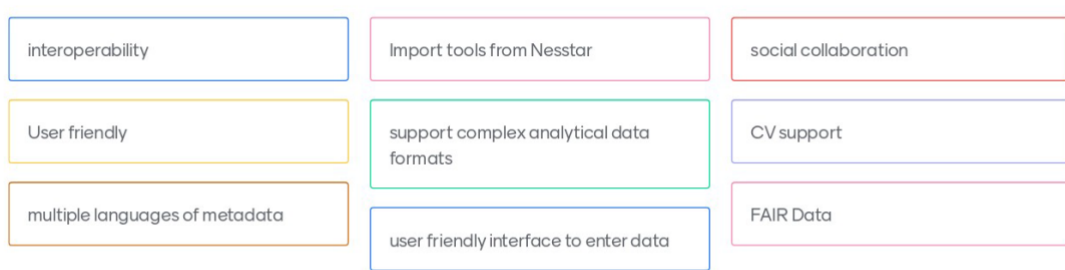
Mentimeter Survey from the DARIAH Community Requirements for a Dataverse Repository Webinar
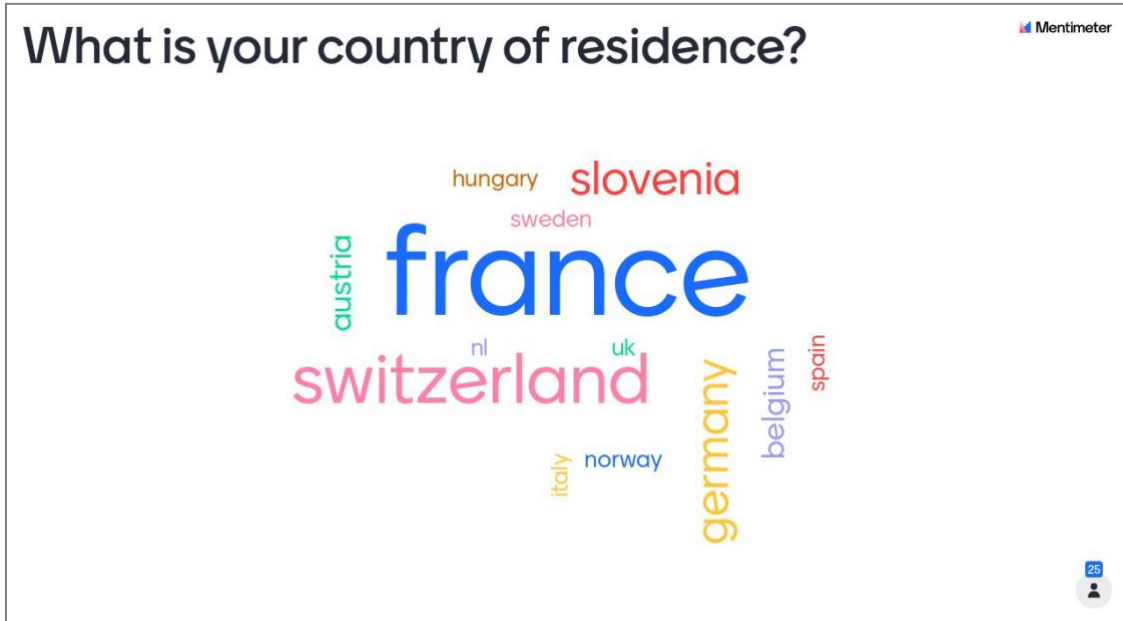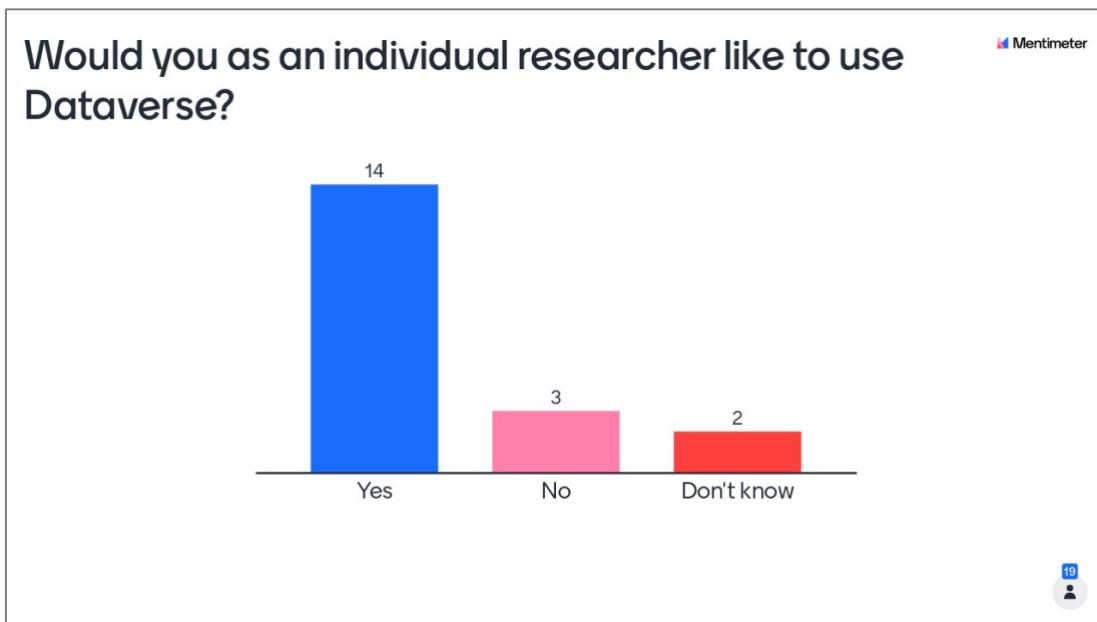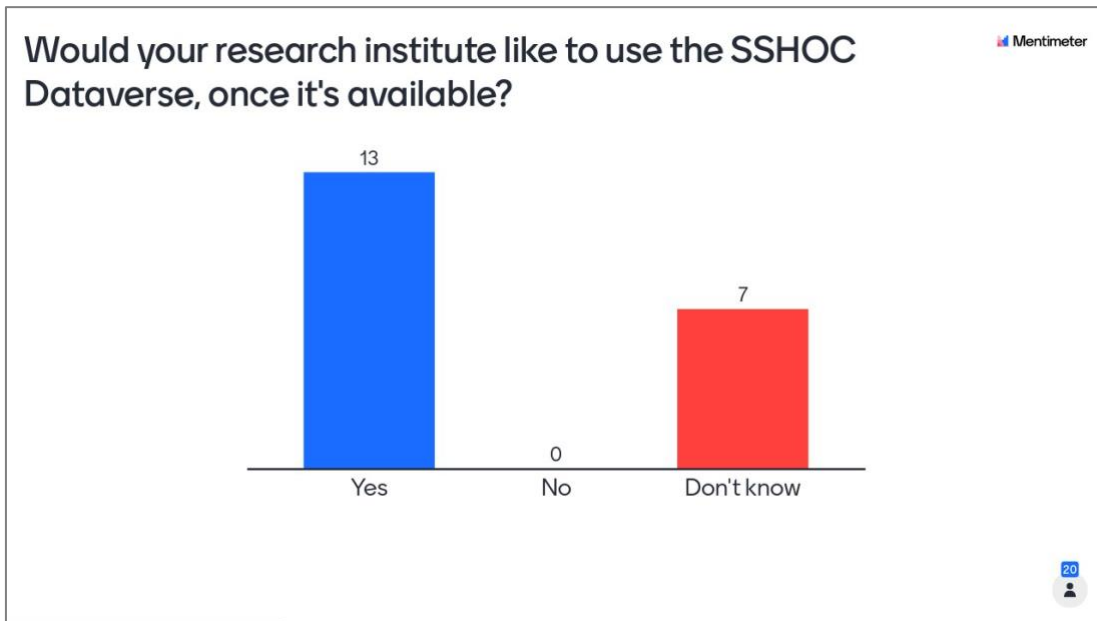
Please indicate your research domain or area of your expertise:

Would you as an individual researcher like to use Dataverse?

Would you be interested to test the functionality we are developing within SSHOC.

Is it well defined what are research data in your particular domain?
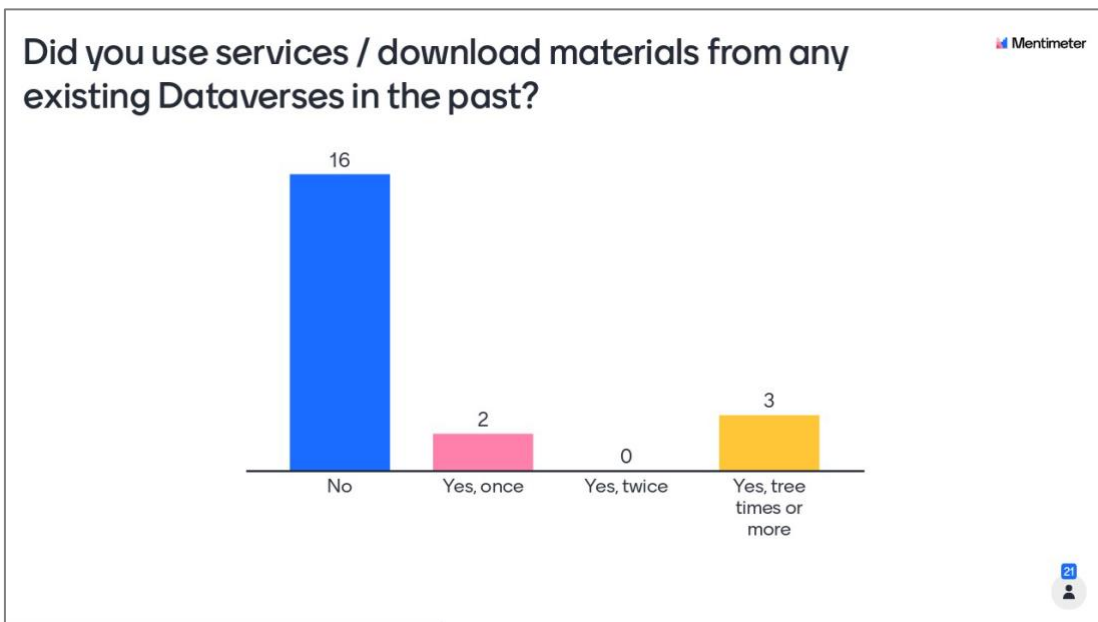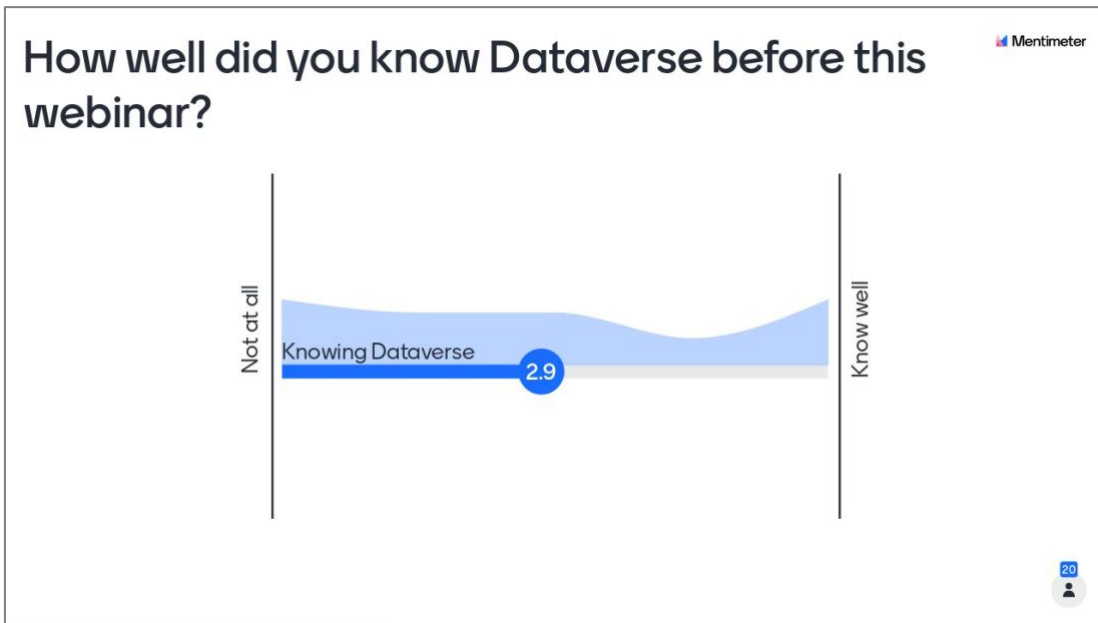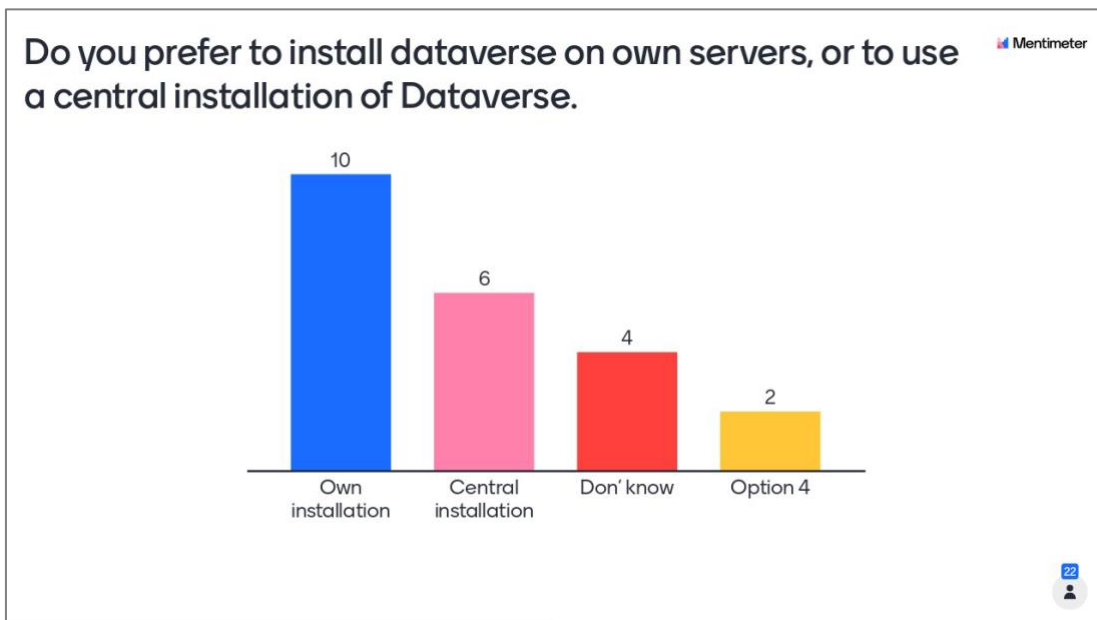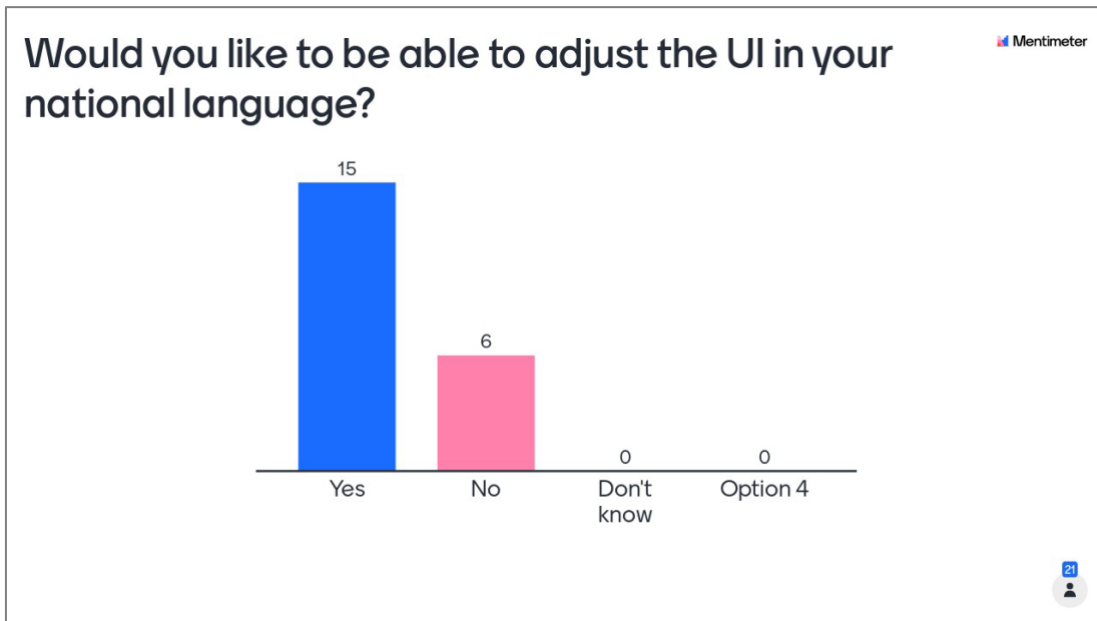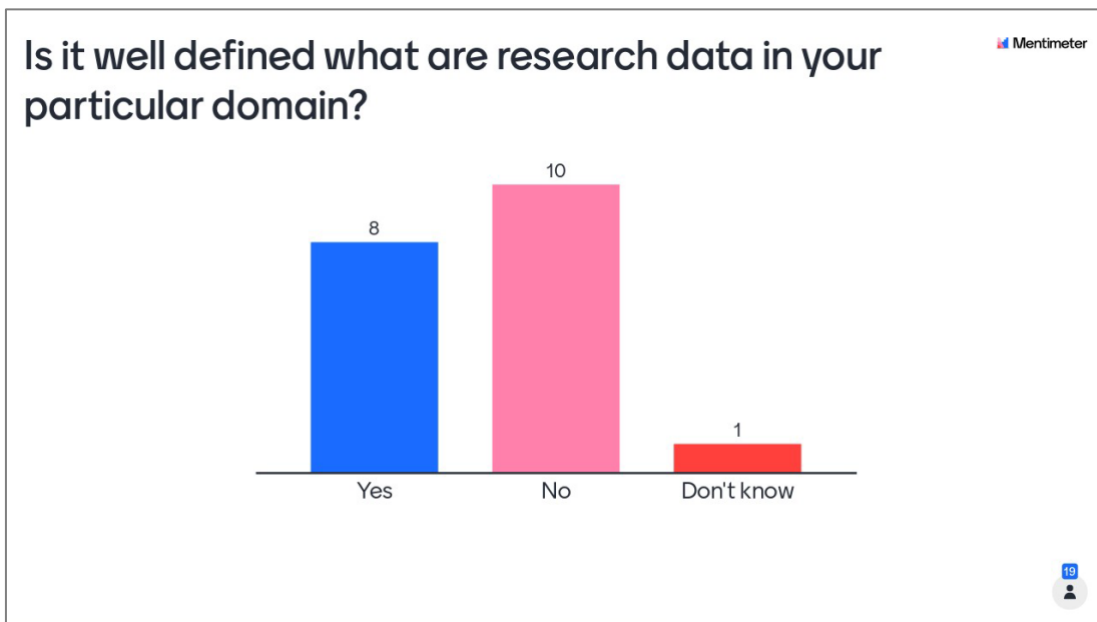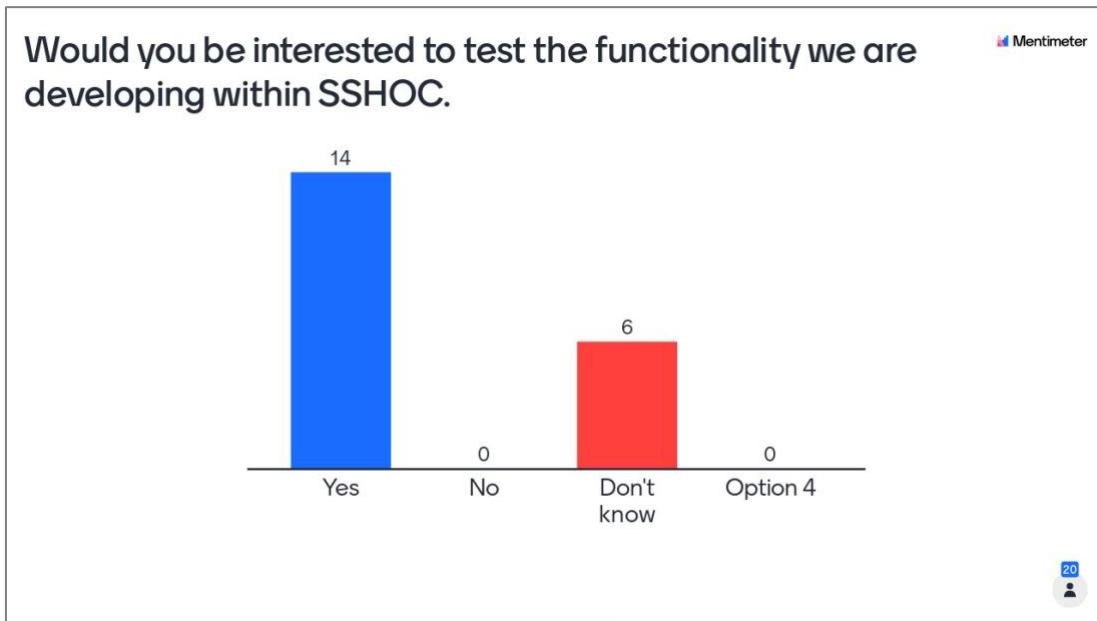
If yes: could you give us some examples?

Mentimeter

There is no gudefor researchers how to manage data	analytical datasets generated during materialtechnical research on cultural objects	XML, RDF, turtle

How does your typical research data look like?

Mentimeter

Not clear ...	blue	structured survey / raw text / coded text / audio files / jpgs
Large data, quantitative, controlled access, spreadsheets	proprietary datasets	text, digitised documents from archives, survey/statistical data
Option 4	Small data, encoded text (TEI, etc.)	
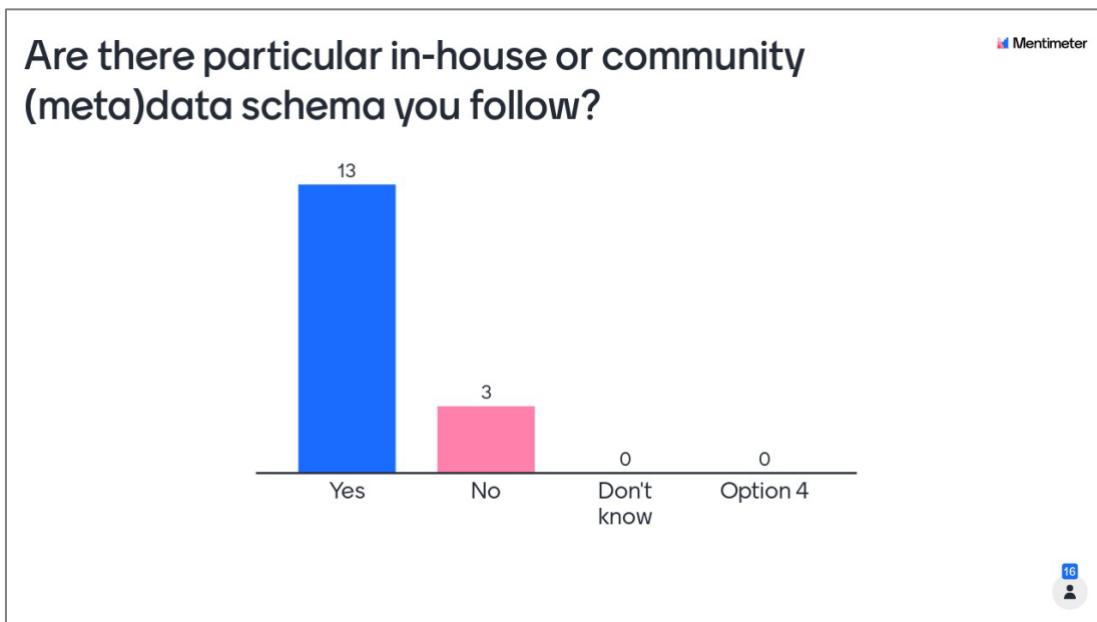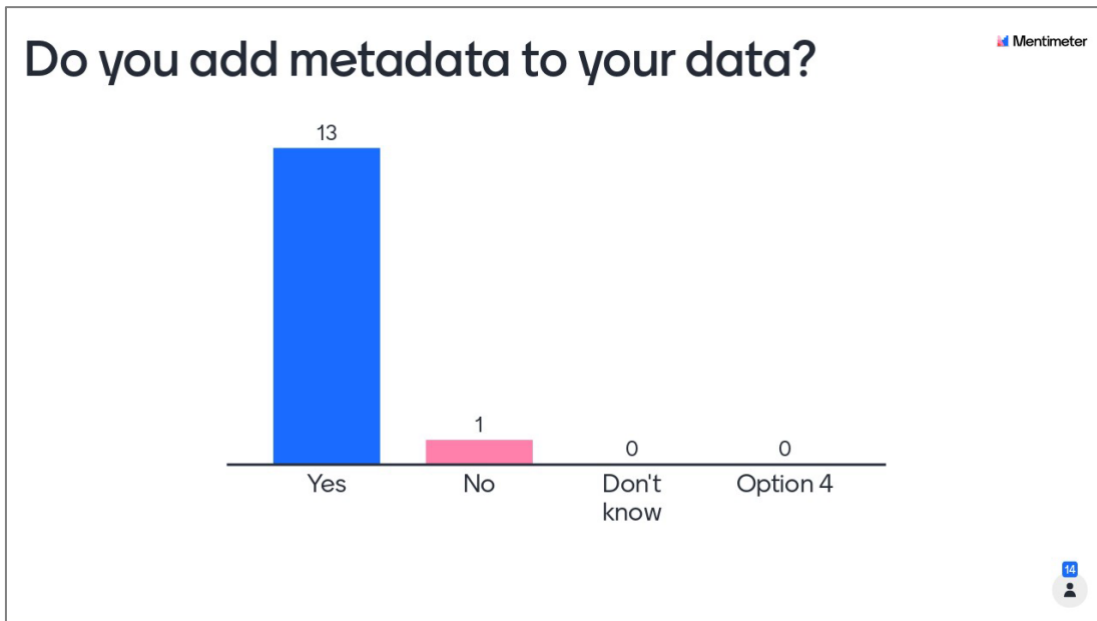
If yes: could you give us some examples?

Mentimeter

DDI 2.5	CV + DDI	most of the data that I have collected are archived by DANS, GESIS etc
DDI Controlled VocabulariesAATPICO TesaurusTGN		

4

Do you usually rely on third-party data sources (e.g. libraries, archives, museums)?

Mentimeter

Yes	No	Don't know	Option 4
3	3	2	0

8
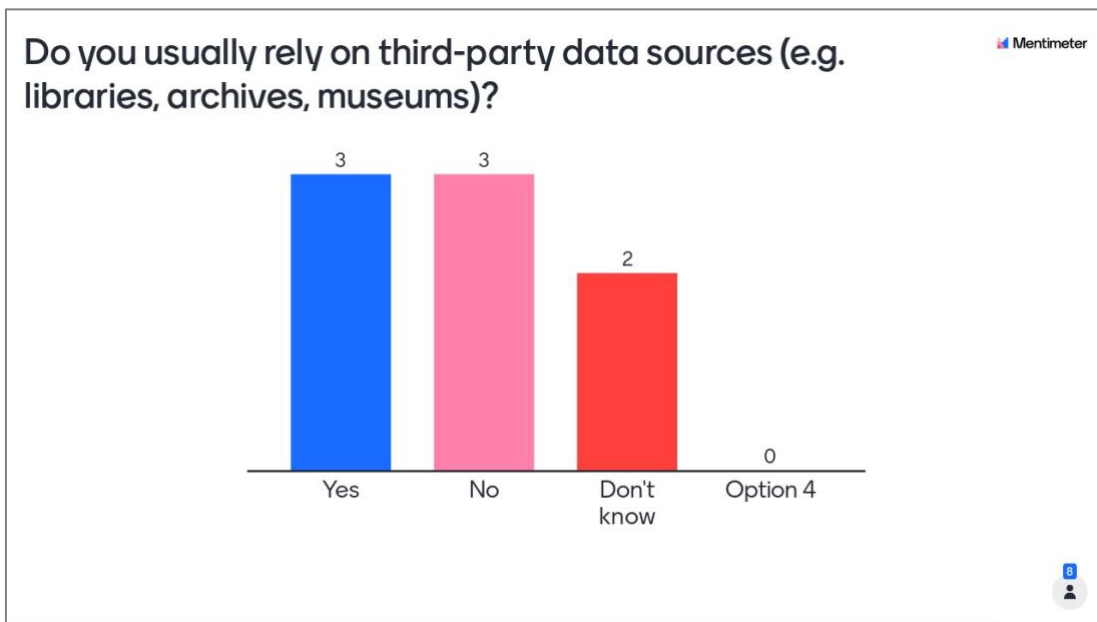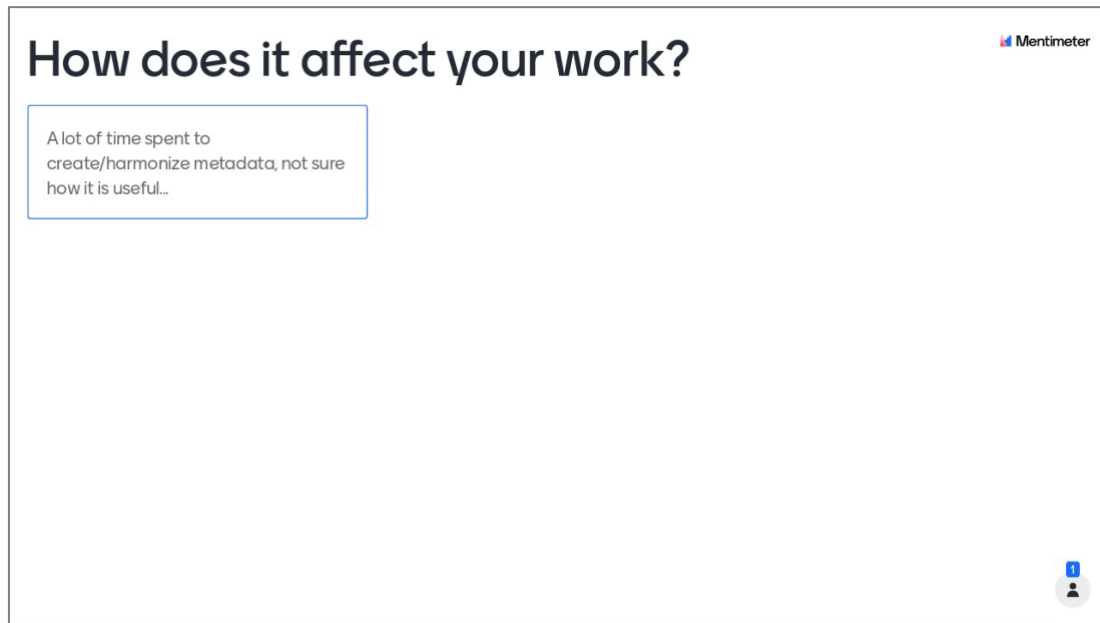
Post-event survey for DARIAH Community Requirements for a Dataverse Repository Webinar

Overall, how would you rate the webinar?
3
5
3
4
4
4

Where did you learn about the webinar?
Social Media
Event
SSHOC website
mailinglist
Colleagues
mailinglist
Colleagues

How well organised was the workshop/webinar (in terms of time management, length, venue)?
3
5

3
4
4
3
4

What did you hope to gain from the workshop/webinar?
Getting to know Dataverse
Goals, current state, and plans for the SSHOC Dataverse project
detailed knowledge on use of dataverse as data depositor as well as in an end-user capacity
Basics about Dataverse
info
insights into development and discussion of humanities repositories, against a backdrop of an apparent
lack of domain-specific solutions in the humanities
To learn more things about Dataverse

Did the workshop/webinar meet your expectations?
3
4
3
3
4
3
4

Do you see this workshop/webinar having a positive impact on your work and how?
I have a better idea of what is Dataverse.
Yes, I now have a better understanding of the SSHOC Dataverse project plans and goals.
a little
not yet...
yes
I will continue to watch this initiative, but at present this workshop won't have a particular impact on my
work
I have learned some new things about Dataverse, since I will use it at my work

What did you like most about the workshop/webinar?
It was not too long.
Two presenters working well together :-)
on-screen demo
Some demonstrations
questions

extensive answers to questions

Nothing special, I liked the webinar as a whole

What did you miss or could be improved at the workshop/webinar?

I was expecting more technical details but also high-level considerations on global and European research data management strategies and challenges: what is the current landscape, what are the challenges, what are the main directions, etc. But these were not the focus of the webinar, but it was still very interesting. Thanks!

Not sure.

"on-screen demo: not always clear what the demonstrator did (needs an online equivalent of a laser pointer!)

Also, it would be helpful to have a wider set of references to supporting materials/manuals/use cases etc. "

Some more interaction and live discussion

dont know

I already knew Dataverse a bit, so much of the demo was dispensable forme. Some of the mentimeter questions at the end were hard to answer (to broad to answer shortly)

Nothing it was ok

ANNEX 2

HOW TO IMPROVE THE QUALITY OF YOUR REPOSITORY? SSHOC AND CERTIFICATION OF REPOSITORIES AND BUILDING TRUSTWORTHY REPOSITORIES: INTRODUCTION TO CORETRUSTSEAL CERTIFICATION

REPORT OF THE WEBINARS

By Ana Inkret, CESSDA/UL-ADP

Background

This report concerns two SSHOC project webinars on the certification of data repositories and the support provided by SSHOC project in obtaining the CoreTrustSeal certification. The first webinar, "How to improve the quality of your repository? SSHOC and certification of repositories" took place on 23 April 2020. It was followed by another webinar, "Building trustworthy repositories: Introduction to CoreTrustSeal certification" on 14 October 2020, specifically tailored for DARIAH community.

Both webinars are part of the SSHOC project engagement and awareness-raising effort, presenting the value of repository certification and the certification support available within the project to the wider community of SSH data service providers.

How to improve the quality of your repository? SSHOC and certification of repositories

Webinar Overview & Format

The webinar focused on the certification of digital repositories and how a repository can apply for the CoreTrustSeal certification. The webinar also touched upon how SSHOC will be able support repositories

seeking certification. The aim of the webinar was to present the benefits of certification, the process of obtaining the CoreTrustSeal, and to introduce the SSHOC assistance during the certification process.

The content of the webinar was organised by task 8.2 Trust & Quality Assurance. Technical support was provided by LIBER (T6.2). Webinar was hosted by Mari Kleemola (Finnish Social Science Data Archive). The main speaker was René van Horik (Data Archiving and Networked Services). Hervé l'Hours (CESSDA/UKDS), Emiliano Degl'Innocenti (E-RIHS/CNR), and Tuomas J. Alaterä (CESSDA/FSD) were part of the discussion panel.

A total of 61 people registered for the webinar, representing 46 different organisations from 25 countries around the world.
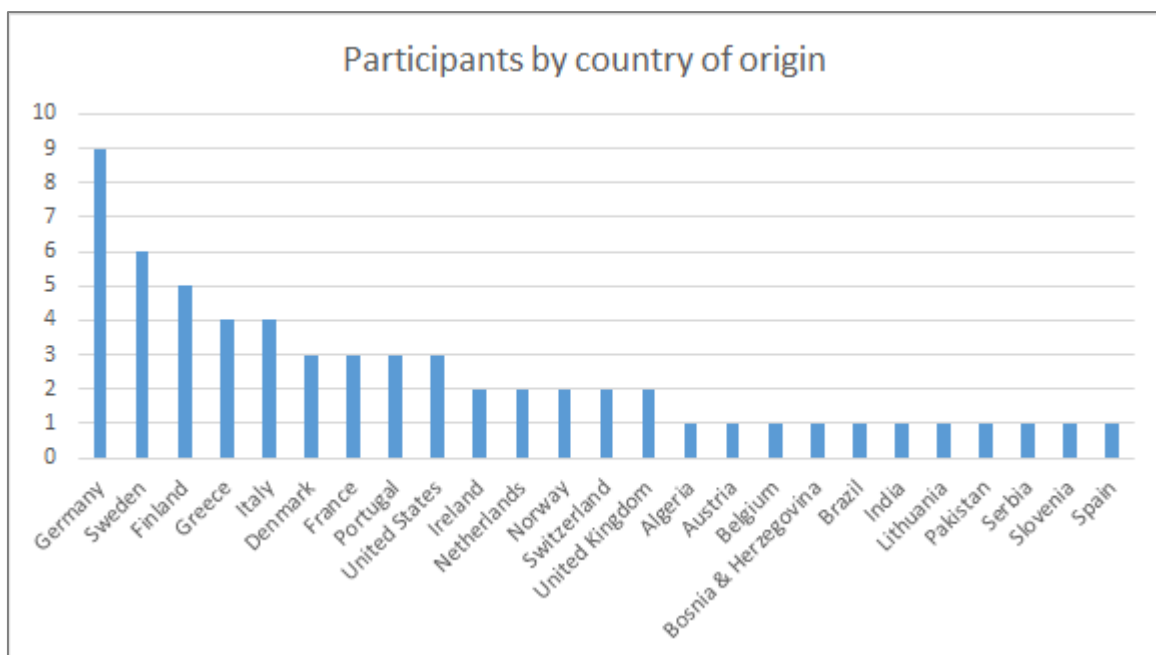
Participants included researchers (5 in total) as well as data service staff at universities and research infrastructures (11 participants), research and e-infrastructures, EOSC thematic clusters (18 participants), and research libraries and archives (which constituted the majority with 27 participants).
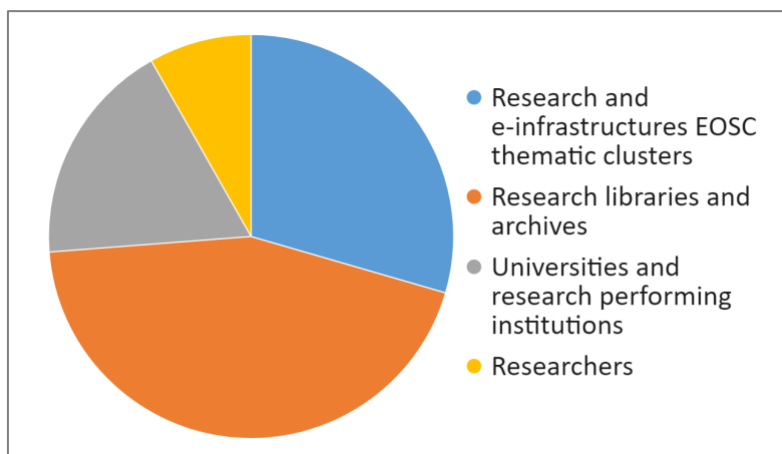
Participants by stakeholder categories

Webinar started with an introduction by Mari Kleemola. René van Horik then presented the CoreTrustSeal certification. The final part of the webinar was allocated to the Q&A and discussion with the panel. The webinar lasted one hour in total.

Presentations & Discussions: Key Points

SSHOC and the CoreTrustSeal Certification of Repositories
Speaker. René van Horik (DANS)
Main points. Scientific data should be managed, curated, and archived to preserve the initial investment in collecting them. Sustainability of repositories challenges different areas: organisational, technical, financial, and legal. Certification can contribute to ensuring the reliability and durability of data repositories and hence the potential for sharing data over a long period of time. The certification process also provides a valuable opportunity to enhance processes and activities.

CoreTrustSeal is a community-based organisation that offers certification for data repositories, promoting trust in data repositories. The trustworthiness is assessed using 16 requirements in three categories: the organisational infrastructure of the repository (funding, mission and continuity plan), the management of digital objects (documentation, integrity, quality), and technology issues (software, systems, data security). Obtaining CoreTrustSeal certification consists of four stages: submission of an application, review of the requirements, revision of the application and approval of the certification by the Board.

SSHOC project assists the repositories in the certification process by explaining the procedure to acquire the CoreTrustSeal, providing guidance in filling in the application, and in some cases, review and feedback on repository's self-assessment before they submit their application to CoreTrustSeal.

Questions and answers

Speakers: Mari Kleemola (FSD), Hervé l'Hours (CESSDA/UKDS), Emiliano Degl'Innocenti (E-RIHS/CNR), Tuomas J. Alaterä (FSD), René van Horik (DANS)

Main points: Panel members have wide experience as CoreTrustSeal board members, reviewers of CoreTrustSeal applications, and as representatives of repositories that have acquired a CoreTrustSeal certification. Questions from the audience concerned the certification process and requirements for applying, including the appraisal of restricted access to sensitive data. Emiliano Degl'Innocenti introduced the E-RIHS approach to fostering a culture of FAIR principles and trustworthiness in the multidisciplinary community of heritage science. Tuomas Alaterä presented the activities of the Nordic EOSC project.

Links to materials

- **Slides:** https://doi.org/10.5281/zenodo.3774396
- **Video:** https://www.youtube.com/watch?v=Ev0KRXkH4jo

Building trustworthy repositories: Introduction to CoreTrustSeal certification

Webinar Overview & Format

The webinar presented the key characteristics of sustainable digital research infrastructure. It was intended for data repository managers planning to seek CoreTrustSeal (CTS) certification or wanting to understand how their data is managed in the context of CoreTrustSeal requirements.

The webinar was part of the workshop programme of the DARIAH Virtual Annual Event 2020: Scholarly Primitives. It was organised by René van Horik (DANS) and Tomasz Parkoła (Poznań Supercomputing and Networking Center), representatives of T8.2. Birger Jerlehag (Swedish National Data Service), Daan Broeder (KNAW Humanities Cluster/CLARIN ERIC), René van Horik and Tomasz Parkoła joined the event as speakers.

A total of 20 participants attended the webinar. Researchers were the majority (10 participants), 6 participants identified themselves as representatives of research libraries and repositories, while 4 belonged to the e-infrastructures and thematic clusters.
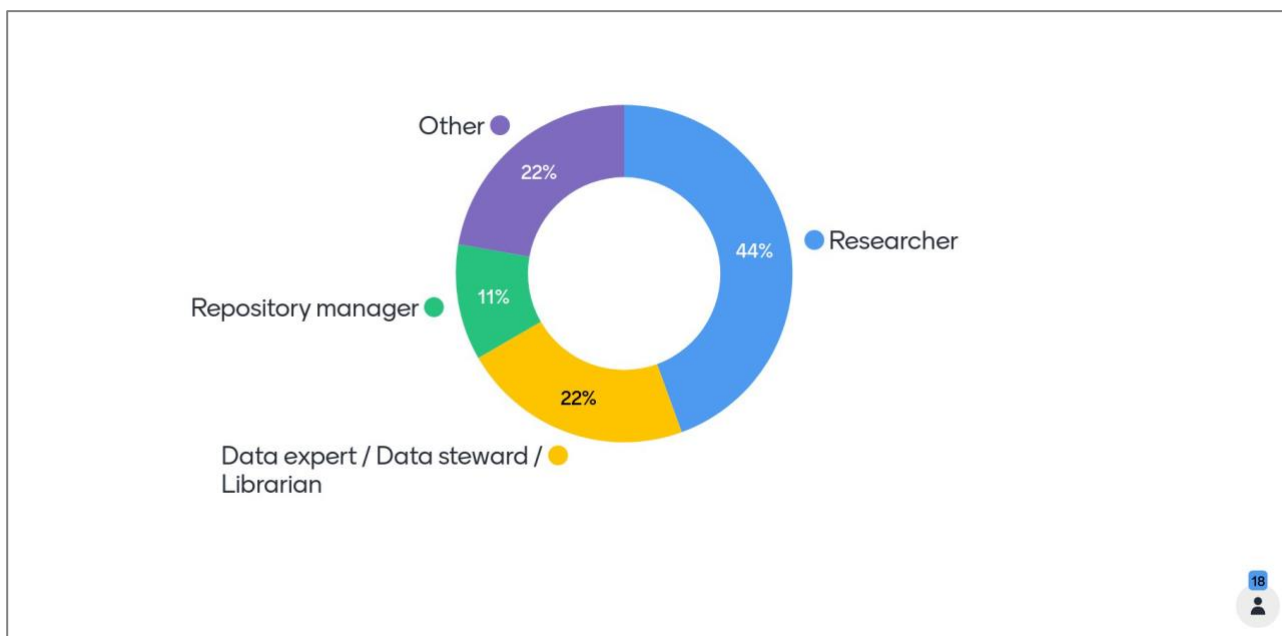
Outtake from the Mentimeter results: Which role suits you best?

After an introduction on CoreTrustSeal within the SSHOC project by Tomasz Parkoła, Rene van Horik introduced the CoreTrustSeal. Birger Jerlehag then presented the relation between certification and FAIR data objects. The presentations concluded with two use cases: Dann Broeder presented the CLARIN perspective, and Birger Jerlehag discussed the CESSDA approach to CTS.

Presentations & Discussions: Key Points

CoreTrustSeal: Short Introduction
Speaker: René van Horik (DANS)
Main points: Validating publication, preserving and sharing the data are main incentives for certification of data repositories. Participants learnt about the requirements of the CTS and the process of application. Two additional sources of information on CTS were recommended: the previous SSHOC webinar and a short video by DANS.
Link to slides: https://doi.org/10.5281/zenodo.4233672

CTS repositories and FAIR data objects
Speaker: Birger Jerlehag (Swedish National Data Service)
Main points. The main objectives of Trusted Digital Repositories is to preserve and disseminate well documented and accurate research data, formatted and packaged to suit it's designated community - i.e. FAIR Data. The most important element of a FAIR data object is the persistent identifier that consists of a unique identifier and the service that locates the resource even when its location has changed. Compliance with CTS requirements also means compliance with FAIR principles. The core of repository

organisation is the OAIS model, where CTS can be compared with the management and FAIR with the interaction with the customer.

Link to slides: https://doi.org/10.5281/zenodo.4233726

CoreTrustSeal, a CLARIN perspective
Speaker: Daan Broeder (KNAW Humanities Cluster/CLARIN ERIC)
Main points. CLARIN - Common Language Resources and Technology Infrastructure supports the sharing, use and sustainability of language data and tools for research in the humanities and social sciences. CLARIN B-centres started the certification to improve quality of the centre contributions to CLARIN infrastructure, decrease interoperability problems, and prove the contributions from national projects to EU infrastructure. Making internal processes explicit was also useful for centres themselves. CTS was found very useful, but partial CTS would be welcome, adding the "Access" topic and certification of services.
Link to slides: https://doi.org/10.5281/zenodo.4233873

Use Case 2: CTS approach in CESSDA
Speaker: Birger Jerlehag (Swedish National Data Service)
Main points. CESSDA - Consortium of European Social Science Data Archives - service providers demonstrate their reliability by adhering to the principles of the Open Archival Information System reference model and any agreed CESSDA ERIC requirements for operating trusted repositories (CTS). Service providers are supported by the CESSDA Trust Group that guides in acquiring and maintaining compliance with CESSDA obligations and the requirements of the CTS, monitors and reviews compliance, and maintains an overview of the trust landscape including certification standards and the emergence of the FAIR data principles and the requirements of the EOSC.
Link to slides: https://doi.org/10.5281/zenodo.4233994

Link to video: https://youtu.be/8SUA8XbM2RA

Outcomes & Feedback

Webinars were an introduction into the benefits of certification, the requirements of the CoreTrustSeal certificate, use cases, and the support offered by SSHOC. The recordings and materials from the events continue to be a useful resource for users. A blog post on the event was published after the first webinar, and the frequently asked questions of the participants are also available online. The participants were invited to join the SSHOC Trust Support mailing group to continue communication.

A post-event survey was only carried out for the first webinar. The responses were very favourable. Over 80% of participants rated the event at the highest rates. They confirmed it provided "essential information", they valued "good time management, good moderation, friendly atmosphere, taking up all the questions", "concise information, to the point, large time for questions", and "aside from a very

interesting Q&A session, to be informed regarding mailing list and dedicated email support address". Participants "will be able to use things from this webinar in practice".

The work in task 8.2 continued with individual consultations and a Certification Support Webinar for the 14 repositories that SSHOC supports in the certification process.

Post-event survey, 23 April 2020

How well organised was the webinar (in terms of time management, length, venue)?

17 odgovorov
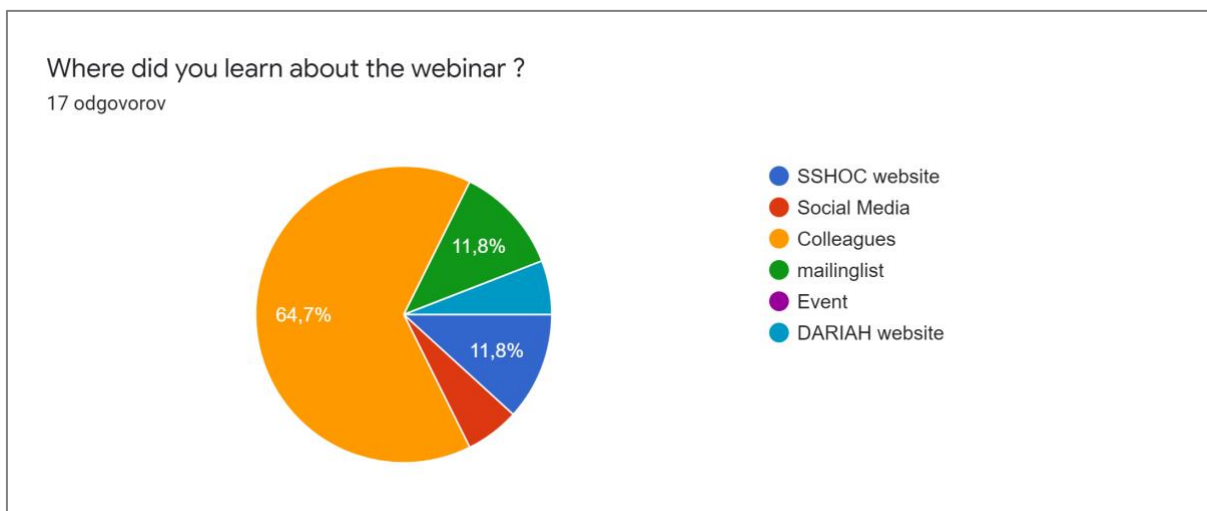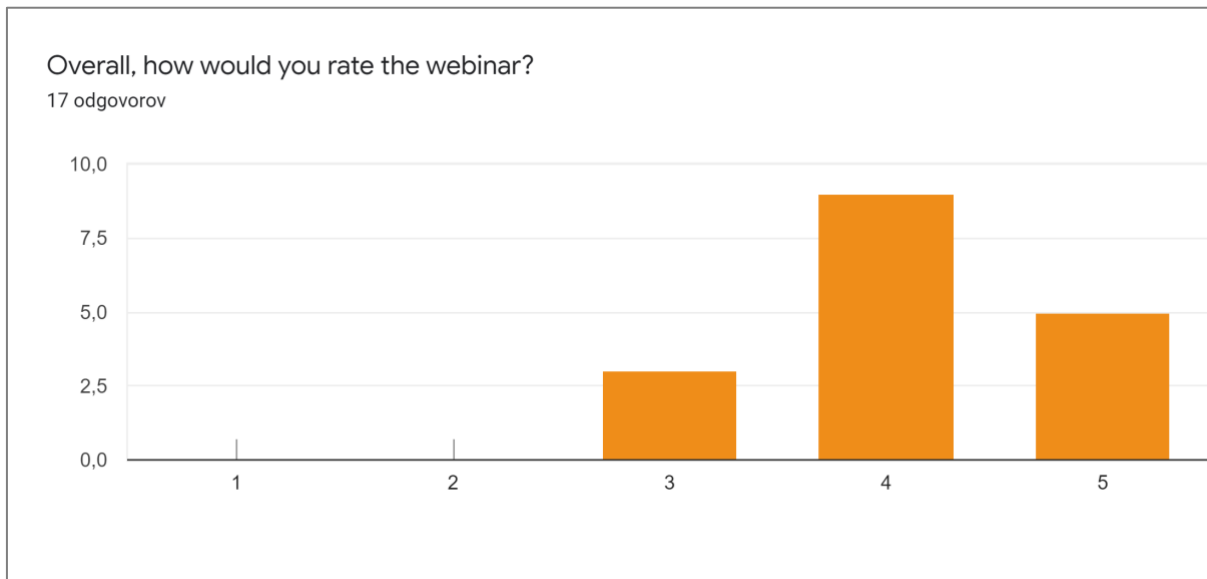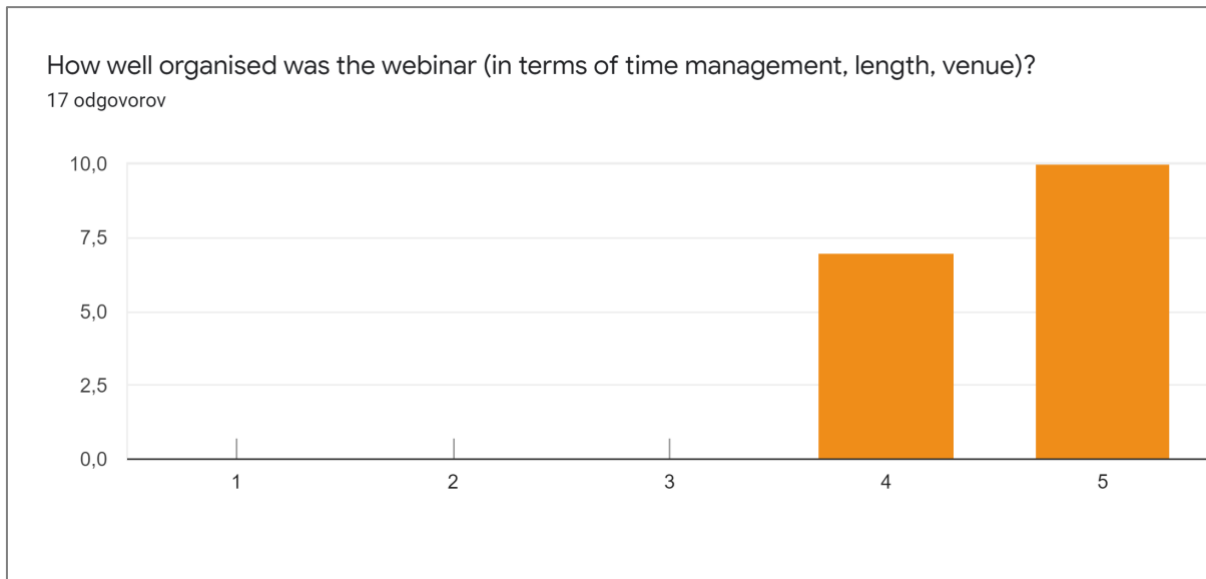
What did you hope to gain from the webinar?

To understand the process of creating a repository and getting it certified

Inputs about CTS

more insights about what CTS certification requires

Knowledge

Information about the CoreTrustSeal

An overview

General information, as expected

deep look into the certification categories, maybe by going through examples

First ideas about repository certification

Additional information about the certification of repositories as well as practical examples of certified repositories and their experiences in achieving that.

As introductory webinar it was nice. Next time a bit longer on CTS :)

learn more about certification, get an overview of Core Trust Seal

To hear more about CoreTrustSeal

To learn more on how CTS works

Get guidance on specific questions that may be challenging (i.e. adapting our response to what is expected in terms of granularity, evidences …)

overview

More information concerning TrustSeal procedures
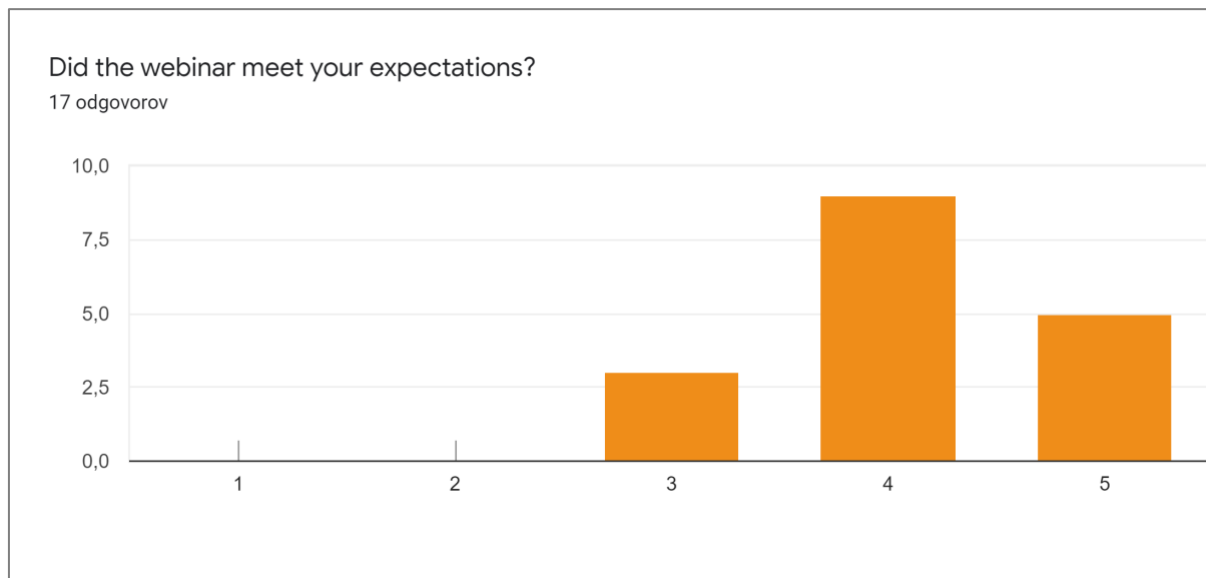
Do you see this webinar having a positive impact on your work and how?

Yes

Exchange of experience, inputs

yes, interesting overview of the topic

Yes. Essential Information about certification repositories.

It could have a positive impact into incorporation more mechanism about FAIRness of our data in our repositories

Align our work with this standard

Important information to read and prospect further

I did get a first overview, that does help…

not yet

It was a good introduction.

always good to be reminded about continuous improvement on local workflows - to get re-certification

don't know yet. But of course it won't have a negative impact!

We will be able to use things from this webinar in practice, Data Center Serbia for Social Sciences is still small archive and we need to improve our organisational and technical infrastructure according to CoreTrustSeal requirements.

I noticed a few colleagues joining in, so we all have someone to discuss with.

Aside from very interesting Q&A session, to be informed regarding mailing list and dedicated email support address

Standard compliance

Yes, because I have got the information requested

What did you like most about the webinar?

The presentation

very well moderated

short and concise

Information about certification repositories.

The solid and clear information it provided

Overview and Q&A

Concise information, to the point, large time for questions

Good time management, good moderation, friendly atmosphere, taking up all the questions

the informal atmosphere in combination with the expertise of the speakers

Interaction between organisers and participants.

Q&A Session

many reminders via e-mail (thank you!)

All details about CoreTrustSeal

The presentations

Getting direct responses from experts

Q&A

Interaction/ complementary of presenters

What did you miss or could be improved at the webinar?

Nothing

The sound broke off some time

Sound Quality

maybe some more concrete examples

Nothing. ALL ok.

Concrete examples of repositories and evaluations

Overview of variety of respositories that are certified already, examples reagrding problematic aspects in certification processes

show a test case scenario of an example repository that acquires CoreTrustSeal certification (although, of course, this would take a longer webinar -- maybe optional at the end?)

Presentation skills could be improved; less reading during talking and more engaging with the audience, if possible.

perhaps focusing on few reque. more in detail how it works

two or three model cases would have been good for illustration

Nothing for now

You who work with CTS all seem to have an understanding of how it works, and the threshold for being able to join in this understanding is high. How could it be lowered and made easier to join the community?

A supporting slide or two for Tuomas' and Emiliano's short presentation maybe

Links to documentation

it was OK

Can you suggest themes for other online SSHOC certification events?

The basics of creating a repository, the process of planning

FAIR principles and repositories

CoreTrustSeal for FAIR data

Analysis of 2-3 repositories test-cases, showing what is in compliance with SSHOC expectations or what should be improved

in-depth illustration of the certification process and requirements

Practical examples of getting certified; sharing of experiences from people involved in this process.

technical / archival - long term preservation part of CTS

More concrete cases that show step by step how it is done. Maybe a webinar made in this way could be divided in the three main areas. Or else it could be too long. Thanks for an interesting webinar!

-

CoreTrustSeal and FAIR

You could develop different units of TrustSeal requirements in different webinars

Mentimeter survey, 14 October 2020
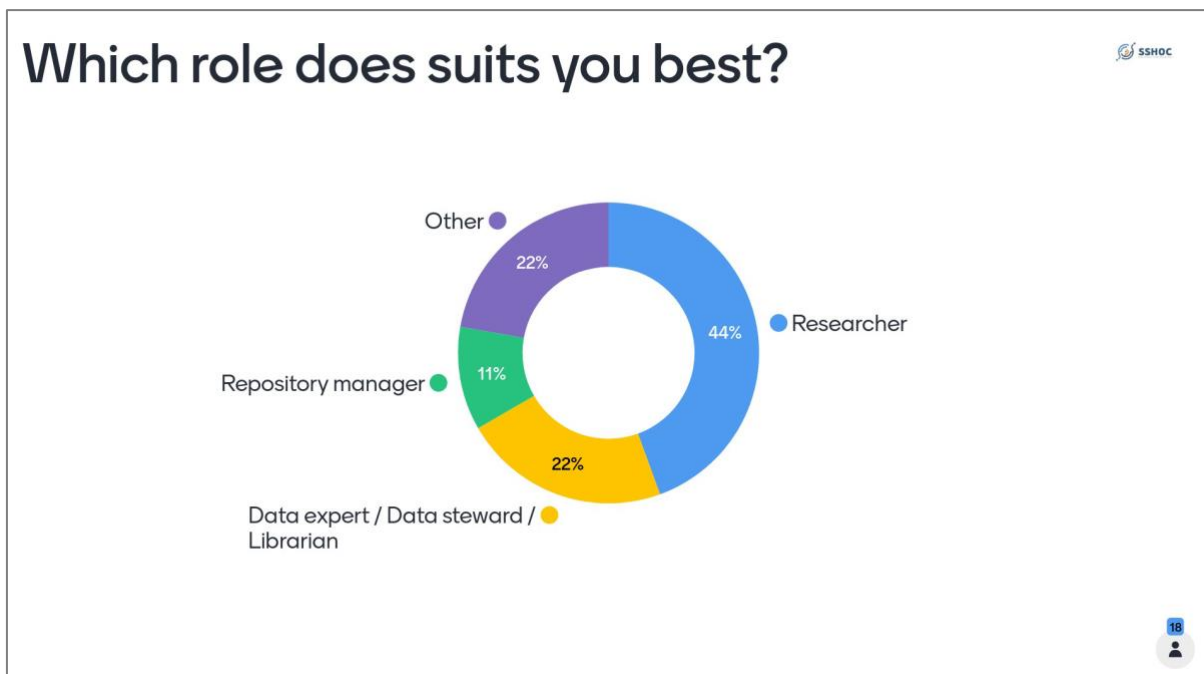
Why are you attending this workshop?

4 — I was sent here

7 — I am curious but do not expect any concrete follow-up

8 — I am very interested and can imagine a concrete follow-up

19

Were you aware of the existence of CoreTrustSeal before the invitation to this workshop?

3 — Yes, I know it very well

9 — Yes, I have heard of it

6 — No, I am not familiar with CoreTrustSeal

18

When did you learn about the concept of repository certification?

9 — Less than a year ago
6 — Between three and one year ago
2 — More than three years ago
1 — Not applicable
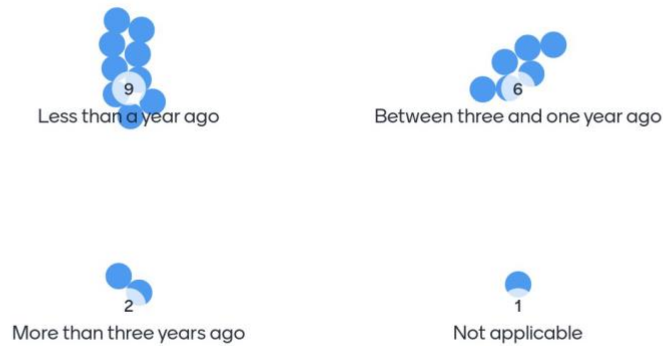
What are the three keywords that would be important for you in the context of repository certification?

reliability, ease of use, fair data, quality, inexpensive, sustainable, valid, fair, accessibility, persistence, sustainability, interoperability, added value, policy, visibility, trust, quality assessment, preservation, ease of deployment, digital preservation, trustworthiness, community

Do you think CoreTrustSeal is relevant to the work you do?
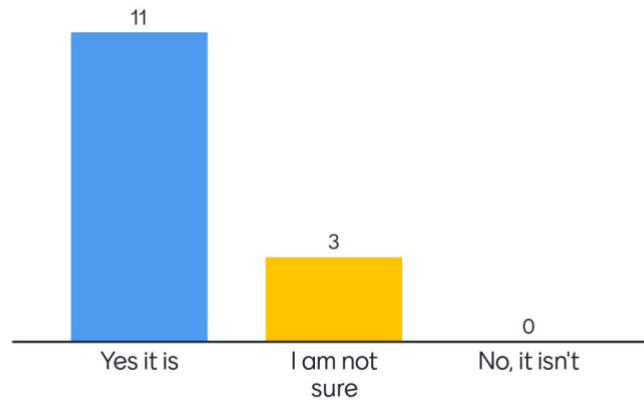
Why do you think the CoreTrustSeal could be beneficial for your organisation?

Prestige	it will show we offer quality	networking
increeased trustworthiness	Meeting funder requirements	FAIR requirements
demostrate trustworthiness	Prestige	As a seal of trust

Why do you think the CoreTrustSeal could be beneficial for your organisation?

increase trust	reliability

Would you participate in webinars on specific groups of CTS requirements (with detailed explanation and examples)?
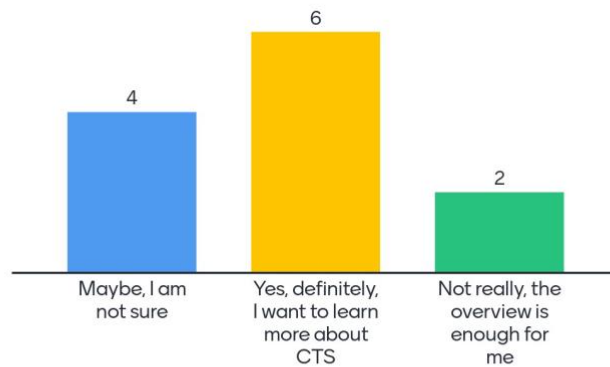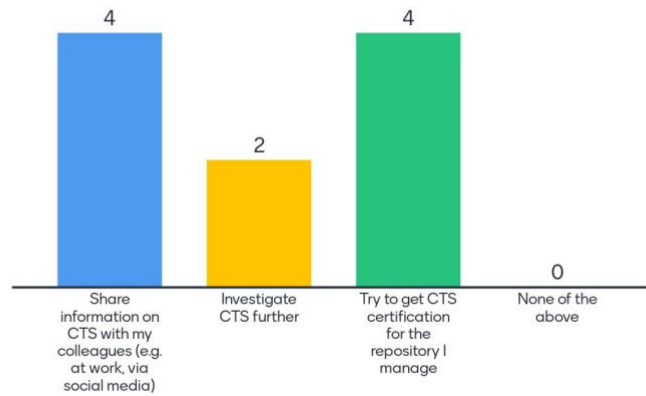
4	6	2
Maybe, I am not sure	Yes, definitely, I want to learn more about CTS	Not really, the overview is enough for me
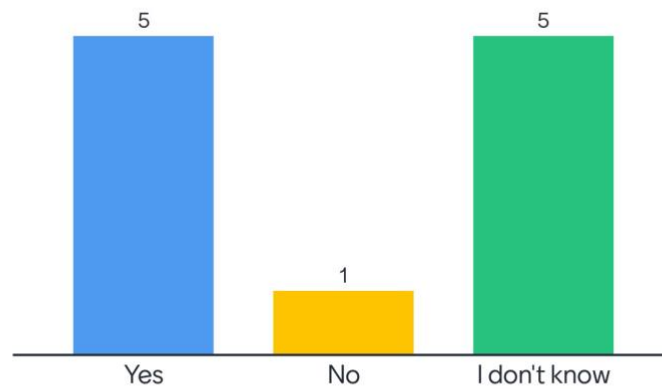
ANNEX 3

SSH Open Marketplace: Public Consultation for the DARIAH Community Webinar Report

By Laure Barbot, DARIAH-EU

Background

The [SSH Open Marketplace: Public Consultation for the DARIAH Community](#) webinar took place online on 3 July 2020. It was organised as a user consultation and engagement event to present the alpha release of the SSH Open Marketplace and collect feedback and requirements from the DARIAH community.

Workshop Overview & Format

The aim of the workshop was to give DARIAHns an early glimpse of the alpha release of the SSH Open Marketplace developed within the [SSHOC project](#). The goal of the session was to present this discovery portal connected to the EOSC ecosystem, to collectively think about the concept, and to present the different ways the DARIAH community can get involved.

The content of the workshop was planned and delivered by Laure Barbot, Frank Fischer (both DARIAH-EU), Stefan Buddenbohm (DARIAH/UGOE), Klaus Illmayer (DARIAH/OEAW), Clara Petitfils and Nicolas Larrousse (both CNRS). Technical support was provided by TRUST-IT and LIBER.

Over 50 representatives - mostly researchers and librarians coming from universities and research performing organisations - of the DARIAH community joined this consultation, members of DARIAH Working Groups, such as the Research Data Management Working Group, the Bibliographical Data Working Group, the Women Writers in History or the Community Engagement Working Groups, partners in DARIAH-affiliated projects and initiatives.
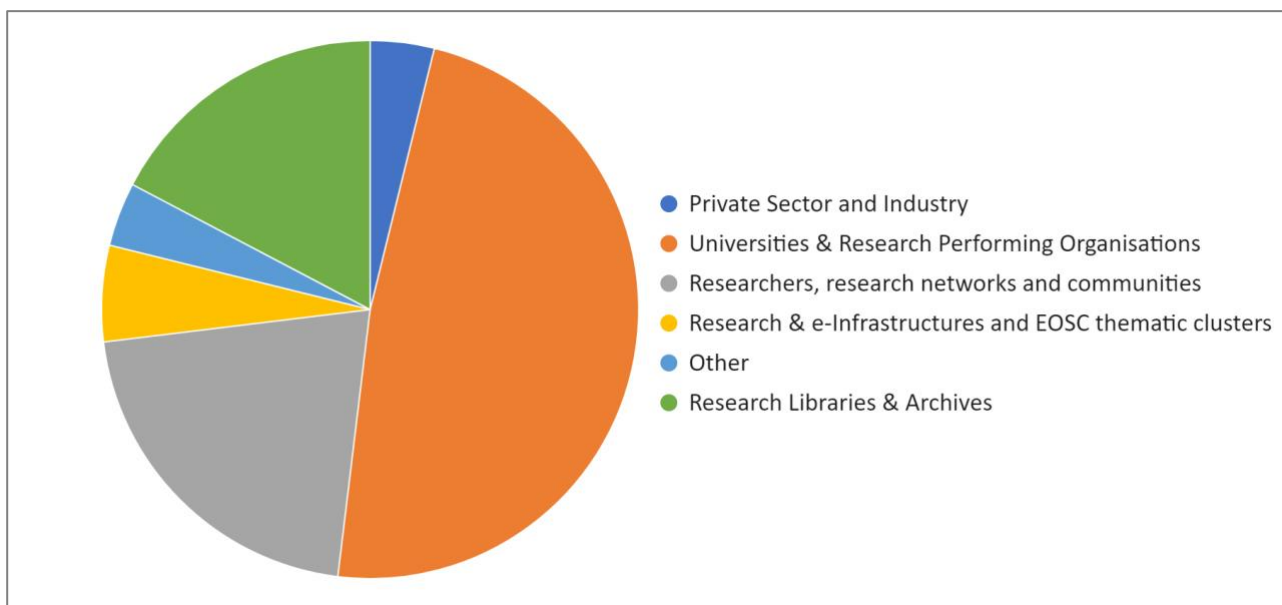
Participants by stakeholder categories

The webinar opened with four short presentations and a demonstration before opening to a rich discussion with the audience.

Presentations & Discussions: Key Points

Introduction - The context of development of the SSH Open Marketplace

Speaker. Frank Fischer, DARIAH-EU Director

Main points. Frank Fischer highlighted that the Marketplace is one of the four pillars of the DARIAH Strategic Plan 2019-2026, and explained how the development was conducted in collaboration with other ERICs and partners within the SSHOC project.

Alpha version of the SSH Open Marketplace

Speaker. Stefan Buddenbohm, from the University of Göttingen

Main points. Stefan Buddenbohm presented the newly released alpha version of the SSH Open Marketplace. Attendees of the webinar explored the portal through screenshots and a living walkthrough.

Data sources and technical background

Speaker. Klaus Illmayer, ACDH-CH

Main points. Klaus Illmayer explained the content of the SSH Open Marketplace and elaborated on the data ingestion workflows for sources to populate the SSH Open Marketplace and other technical aspects such as the data model and technical background.

Curation & governance

Speakers. Clara Petitfils and Nicolas Larrousse from CNRS/Huma-Num

Main points. Clara Petitfils and Nicolas Larrousse presented the curation and governance strategies developed for the SSH Open Marketplace. They showed how (future) contributors and moderators of the portal could shape the community around the Marketplace, and how significant is the curation of content, especially in terms of future sustainability.

Discussion session

The discussions that followed these presentations and demonstrations were very rich and the DARIAH community gave feedback, shared concerns and made suggestions. A summary of the key points can be found below.

Don't make it just a catalogue!: Participants underlined that the SSH Open Marketplace shouldn't be "just another catalogue". They suggested that future developments should be more focused on the "market" aspect of the portal, positioning thus the Marketplace as a social place where users can exchange practices and methods and where social interactions can be fostered, in forms of community efforts, tagging or curation for example.

Build on existing communities: One of the key messages was that curation is very important. As one of the core aspects to sustain the Marketplace in the long run is to foster the motivation of curators (contributors and moderators), different incentives were discussed as to how to build this community of editors. Participants stressed the importance of relying on existing community-building experiences within DARIAH by involving DARIAH Working Groups members in the curation process for example.

Credits and not competition: Giving credits to contributors and moderators of the SSH Open Marketplace was one of the clearest demands and comments coming from the participants. One way to ensure contributions of different research communities is to demonstrate the added value of having accurate and carefully curated entries in the SSH Open Marketplace. Furthermore, for research-oriented communities like DARIAH, the possibility to reuse the aggregated and well curated content of the Marketplace could be a potential motivation for contributions. In the long run, if researchers could get academic credit for publishing their workflows in the SSH Open Marketplace, that could be a great step for the DH community.

Interoperability: A topic of concern for participants was the interoperability of the Marketplace with other platforms. The presenters explained the interoperability framework that is guiding the development: for example, wherever it is possible, Linked Open Data IDs – such as Wikidata IDs for tools – are assigned to the Marketplace entries. The API of the SSH Open Marketplace will also be available to allow reuse of the features or the content collected and curated in the portal.

The session ended with a poll to collect answers of the participants regarding the ways the SSH Open Marketplace could support their work.

In what ways could the SSH Open Marketplace support your work?

- getting new research ideas, access to tools and services
- to find friendly tools for my research and for my teaching
- context of resources
- be up to date
- Get to know what is available
- Facilitating the exchange of what works and what doesn't work well - so I don't repeat mistakes etc
- identify also smaller communities or individual sources
- Providing info about our tools
- To understand workflows and links between tools
- access to tools and services, learning matrerials
- share tools and code
- Understand what tools would be good for my research
- Know what other researchers use
- identify current developments
- Perhaps what tools didn't work?
- broaden the disciplinary scope beyond SSH?
- connect to tool developers and users of those tools, user communities
- visualizations

Except from the Mentimeter survey

The webinar concluded by describing the SSH Open Marketplace as a "broker" to get to the desired tool, resource or workflow for a research topic, or as an entry point to discover resources not known before and useful to achieve a task.

Links to materials
- **Slides:** https://doi.org/10.5281/zenodo.3935344
- **Video:** https://www.youtube.com/watch?v=IQSdQbaA6oY
- **Blog post:** https://www.dariah.eu/2020/08/07/dariah-public-consultation-of-the-ssh-open-marketplace/

Outcomes & Feedback

This workshop was part of the consultation series organised by different SSHOC WorkPackages in order to gather future users' and testers' feedback on the alpha version of the SSH Open Marketplace, prepare the beta version (Dec. 2020) and support the participatory design of the portal. The feedback from the attendants of the workshop was implemented in the beta release. More details on this consultation series can be found in the MS43 report.

ANNEX 4

REPORT ON THE SSHOC WEBINAR SERIES ON OPEN-SOURCE VOCABULARY HOSTING AND MANAGEMENT PLATFORMS

By Iulianna van der Lek, Project coordinator at CLARIN ERIC

Background

This report concerns a [SSHOC Webinar Series on Open-Source Vocabulary Hosting and Management Platforms](#) consisting of three online information sessions organised in September 2020 by CLARIN as part of the SSHOC tasks 3.1 Multilingual Terminology and 6.2.

Apart from the sessions being self-standing events raising-awareness about specific vocabulary management and publication platforms, these sessions as a whole also worked as an informative basis for the follow-up virtual workshop [SSHOC *Considerations for the Vocabulary Platforms*](#) which was organised on 6 November 2020.

Sessions Overview & Format

One of the main goals of SSHOC Task 3.1 *Multilingual Terminology,* as described in the workplan/DoA of WP3, is to find a suitable vocabulary server and publication platform in order to maximise the accessibility and improve discovery by non-native speakers. Following up the evaluation of the vocabulary publication and editing platforms provided in *MS08 Report Choice of Vocabulary Publication platforms for SSHOC,* and the vocabulary harmonisation initiative recently undertaken by CLARIN, three online information sessions about open-source vocabulary management and publication platforms have been organised.

The main goal of these information sessions was to raise awareness in the SSHOC community about different open-source vocabulary management and publication platforms and invite experts to share their insights and practical experience. Therefore, each info session included an introduction to the particular platform and a presentation of at least one use-case. Each session concluded with a lively Q&A part. All the information sessions were organised virtually via Zoom throughout the month of September 2020:

- 3 September 2020, 15:00 - 16:30 CEST *An introduction to Wikibase*

  Speakers: Georgina Burnett and Jens Ohlig (Wikimedia Deutschland), Barbara Fischer and Sarah Hartmann (German National Library)

- 14 September 2020, 14:00 - 16:00 CEST *An introduction to Skosmos*

  Speakers: Mikko Lappalainen (National Library of Finland), Darren Bell (UKDS), Klaus Illmayer and Matej Durco (OEAW)

- 30 September 2020, 14:00 - 16:00 CEST *An introduction to CESSDA Vocabulary Service*

  Speakers: Taina Jääskeläinen (Finnish Social Science Data Archive, Tampere University), Johan Fihn Marberg (Swedish National Data Service)

The full programme overview for the sessions is available on the CLARIN website.

Organisers

The event has been organised by CLARIN as part of their tasks in SSHOC (T3.1 and T6.2).

- Iulianna van der Lek, Project Coordinator of the Vocabulary Initiative at CLARIN ERIC

- Monica Monachini, CLARIN - IT National Coordinator and SSHOC T3.1

- Dieter Van Uytvanck, Technical Director at CLARIN ERIC

- Daan Broeder, Project Manager at CLARIN ERIC

- Darja Fišer, CLARIN/UL-FF, SSHOC T6.2

Participants

There were 159 attendees at the webinars: 41 at the 1st session, 61 at the second session and 57 at the third session.[9] Additionally, there were 26 views of the recordings: 5 views of the 1st session recording,

---

[9] Overall, there were 96 unique registrations. The final sum of participants for all three sessions is, however, higher, because certain individuals participated at more than just one info session.

11 views of 2nd session recordings, and 10 views of the 3rd session recording.[10] In terms of geographical range, the info sessions had a great coverage with the vast majority of participants (85%) from Europe. Overall, 23 countries have been represented at the sessions.

The programme of the info sessions was designed for academic or industry professionals who operate, integrate and manage vocabulary platforms, such as librarians, information specialists, and knowledge management engineers. The majority of the attendees (60%) represented the categories "Research Library and/or Archive" and "University and/or Research Performing Institution", followed by "Research and E-infrastructures" with a representation of 20%. The remaining 20% were represented by "Researchers", "Civil society and/or Citizen Scientist" and "Private Sector and Industry Players".

Which of the below categories would you identify yourself with?
59 responses

- Research Library and/or Archive
- University and/or Research Performing Organisation
- Research and e-Infrastructure, EOSC thematic clusters
- Researcher
- Civil society and/or Citizen Scientist
- Private Sector and Industry Player

20.3%
35.6%
30.5%
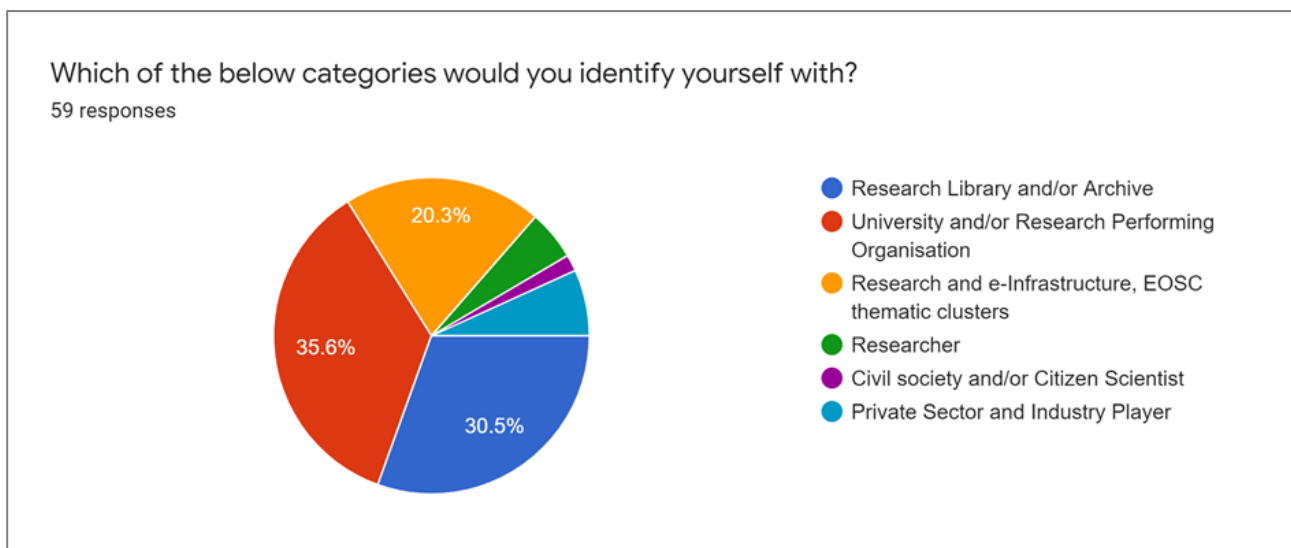
Breakdown of the stakeholder categories

Event summary

For each session, a speaker from the product development team was invited to present a general overview of their product, and one or two presenters from an organisation that has implemented the product were asked to share their experience. The focus was on practical use cases and exchange of best practices regarding the implementation and use of the vocabulary management platforms, namely Wikibase, Skosmos and the CESSDA Vocabulary Service. At the end of each session, one of the CLARIN team members moderated the Q&A part.

---

[10] It should also be noted that the number of the recording views was extracted on 16 November 2020 and is subject to change.

Presentations & Discussions: Key Points

First session: Introduction to Wikibase

This session presented Wikibase, which is the underlying software for Wikidata. It is suitable for structured data, and it can be used to create and manage catalogues, authority files, and controlled vocabularies. Users can create and manage their knowledge base, create links between things and concepts, and collaborate, thus contributing to the Semantic Web. The flexible data model and JavaScript-based UI allows users to easily access and update their data.

The first presentation was given by Georgina Burnett and Jen Ohlig, Partner Relationship Managers at Wikimedia Deutschland. They gave an overview of Wikidata, a project launched in 2012. This platform is the largest example of a Wikibase instance and it currently has more than 89 million data items and over 45, 0000 active users. Wikidata has been released under CC0 public domain and it is useful both for humans and machines.

The data model of Wikibase is flexible and it supports both diversity and multilingualism. Users can easily access their data and export it in a number of formats, such as JSON, RDF/XML, N3, and YAML. Queries are done using SPARQL endpoints. In the near future, Wikimedia will enable federation of WIkibase instances by allowing other Wikibases to reuse the ontology of Wikidata. This way users will be able to set up a Wikibase instance much faster. See the 2020 roadmap to learn more about the future developments at Wikibase.

To learn how the implementation of Wikibase works in practice, two experts were invited from the German National Library to share their experience.

Use case: The German National Library's work on the Integrated Authority File

Barbara Fischer and Sarah Hartmann from the German National Library presented how they used Wikibase for the Integrated Authority File (GND), a service facilitating the collaborative use and administration of authority data. They explained how authority data present and describe entities, i.e. persons, corporate bodies, topics and works relating to cultural and academic collections. Because each entity has a unique and stable identifier (GND ID), it is possible to link the authority data sets with external data resources. This improves the users' overall search experience and results in a machine-readable data network.

Libraries in particular use authority data to catalogue publications. The GND project aims at making the authority data not only more easily accessible but also interoperable. The tests performed have revealed that importing and exporting the data from Wikibase can be difficult and it requires more advanced skills.

Links to materials	**Slides:** https://doi.org/10.5281/zenodo.4153783 https://doi.org/10.5281/zenodo.4153771
	Video recording: https://youtu.be/ZKaXhi6MOnw

Second session: Introduction to Skosmos

Miko Lappalainen, development manager at the Library of Finland and the head of the Finto ontology service, was invited to give an introduction to Skosmos, its current status and future developments.

Skosmos is a web-based tool, developed by the National Library of Finland, that offers advanced search and browsing opinions for vocabulary concepts, support for multilingual vocabularies, and features to link vocabularies. Based on the community feedback, the platform is fast and easy to use. A demo of the platform is available here.

Skosmos is used by the following types of organisations:

- Organisations publishing their vocabs as linked open data.

- Information professionals doing indexing work, such as libraries.

- Vocabulary developers for visualising and searching vocabularies.

- Discovery systems providing subject based information retrieval.

Here are some organisations that use Skosmos to publish their vocabularies: Norwegian National Library, the EU Publications Office, ACDH-CH vocabulary service, UNESCO, FAO (AGROVOC), EU Space Agency (ESA).

The presenter announced that a new version of the platform would be released in the fourth quarter of 2020 and that new features are under development, such as a map widget for vocabularies with geographical information, improved search functionalities for classifications and automatic indexing

extensions. Long term developments will focus on accessibility, better support for classifications of concepts and SKOS XL support, as well as an overall usability.

While the platform provides advanced search and browsing options for vocabulary concepts, Skosmos does not provide a Vocabulary Editor. Therefore, experts from UK Data Service and ACDH Vocabulary Service were invited to demonstrate how they use Skosmos in combination with other tools to edit their vocabularies.

Use case: ELSST Thesaurus: Implementation and integration with VocBench

Darren Bell from the UK Data Service (UKDS) shared their experience with integrating Skosmos with VocBench for the management of the ELSST thesaurus. UKDS is a member of CESSDA (Consortium of European Social Science Data Archives) which is a partner in SSHOC.

The ELSST thesaurus, hosted at UK Data Archive, contains about 3300 concepts and it supports discovery by assigning them to entities, allowing for indexing and grouping.

The architecture of ELSST has recently been redesigned to support a single data model based on SKOS, a more sophisticated ontology editor, and separate the editing and publishing platforms to suit different target audiences.

After evaluating both commercial and open-source products, UKDS chose Skosmos and VocBench3 to manage their thesauri. The presenter remarked that while Skosmos is easy to deploy and it supports interoperability, VocBench requires knowledge of RDF and ontological practice. The revamped ELSST will be launched by CESSDA in November. In addition, UKDA will launch an updated version of the HASSET Thesaurus. Finally, the audience learnt that UKDS plans to migrate their Controlled Vocabularies to VocaBench and Skosmos in 2021.

Use case: ACDH Vocabulary Service: Workflows and editing options in Skosmos

The second use case was presented by Matej Durco and Klaus Illmayer from the Austrian Academy of Science (OEAW). ACDH-CH uses Skosmos as a vocabulary service to manage their vocabularies. The service is also used by the DARIAH community to host vocabularies developed in Arts & Humanities. Both DARIAH-EU and ACDH-CH are partners in SSHOC and contribute to the SSHOC Open Marketplace.

After evaluating a number of vocabulary platforms, ACDH-CH selected Skomos to publish their vocabularies because it fully supports the SKOS data mode and Triple Stores, it has a well-structured and adaptable presentation layer, and a well-maintained API and readable source code, hosted on GitHub. However, the missing vocabulary editor and import/export functionalities still pose some challenges when dealing with multiple users, conflicts and versions of vocabularies. The conclusion of this use case was that it is not possible to guarantee quality without a dedicated vocabulary manager.

Links to materials	**Slides:** https://doi.org/10.5281/zenodo.4154123
	https://doi.org/10.5281/zenodo.4154016
	https://doi.org/10.5281/zenodo.4154147
	Video recording: https://youtu.be/zwZsI4JTnHU
	https://youtu.be/PcYvDXJX6vY
	https://youtu.be/RXscwBc5UV0

Third session: Introduction to CESSDA Vocabulary Service

In this information session, the participants learnt about the CESSDA Vocabulary Service, which enables users to discover, browse, and download controlled vocabularies in a variety of languages. The service is provided by the Consortium of European Social Science Data Archives (CESSDA).

The majority of the source (English) vocabularies included in the service have been created by the DDI Alliance. The Data Documentation Initiative (DDI) is an international standard for describing data produced by surveys and other observational methods in the social, behavioural, economic, and health sciences.

Use Case: How CESSDA Vocab Service is used by archives for translation

The first use case presented by Taina Jaaskelainen from the Finnish Social Science Data Archive (FSSDA) demonstrated how the CESSDA Vocabulary Service is used by archives to translate vocabularies that are included in the CESSDA metadata model into their selected language. For this purpose, a specific semi-automatic XML-based workflow was developed to produce the English versions of the Finnish metadata descriptions. The translated vocabularies would then be used in metadata published in data catalogues to harmonise metadata across organisations.

Use Case: How CESSDA Vocab Service is used by the Swedish National Data Service: metadata editor and Service API

The second use case, presented by Johan Fihn Marberg from the Swedish National Data Service, complemented the first use case by emphasising once again that Controlled Vocabularies can be used to harmonise the metadata, and increase its overall quality and findability. The presenter showed how vocabularies can be implemented in a metadata editor, thus preventing people producing metadata records from using non-vocabulary terms in Controlled Vocabulary elements. The conclusion was that a vocabulary management tool should:

- have multilingual capabilities (both on the level of term and description)

- offer a hierarchical structure and the option to add multilingual synonyms, as well as terms specific to a research domain within a broader controlled vocabulary.

It was also discussed how the CESSDA Vocabulary Service API could be used to handle vocabulary versioning more efficiently.

Links to materials	**Slides:** https://doi.org/10.5281/zenodo.4154428 https://doi.org/10.5281/zenodo.4154428 https://doi.org/10.5281/zenodo.4154426
	Video recording: https://youtu.be/EFcr8drCY8Q https://youtu.be/DhWMqJThxY8 https://youtu.be/iwDuM1PBulA

Outcomes & Feedback

Outcomes

An analysis of the presentations and especially of the Q & A sessions confirms the preliminary requirements for vocabulary platforms that T3.1. have identified in their Milestone report, such as SKOS/RDF support, advanced user management, advanced browsing and search functionalities, collaborative editing interface, linking/mapping and alignment features for vocabularies, track changes and versioning, support for multilingualism and translation.

In addition to the requirements listed above, the webinar series provided some new insights into editing workflows, automatic data provenance tracking, vocabulary curation strategies and quality management in general. Some participants were interested in the integration of vocabulary platforms with Computer-Aided Translation (CAT) tools to enable easy export, conversion and import of the vocabulary databases into translation memories that translators working in SSH could use for translation. Finally, community support and a well-defined governance model are essential in ensuring that the open-source vocabulary platforms are sustainable.

Based on the feedback we have received via the evaluation form and directly from the participants in the Q&A session, we can conclude that the SSH community appreciated this initiative. The insights collected during the information sessions together with the outcome of the follow-up virtual workshop organised on 6 November 2020[11] will be used to reassess the initial requirements for vocabulary platforms. Last but not least, we will further investigate how the harmonisation of SSHOC vocabularies can be added to SSHOC outputs.

Feedback

Almost all respondents to the post-event survey indicated that the session(s) were excellent/very good in terms of the content and organisation, and that the events matched their expectations.[12] All of the respondents confirmed that the sessions will have a positive impact on their work. Near half of the respondents stated that the sessions helped them raise awareness about the existing solutions which could be useful for them sometime in the future, while the other half said that the knowledge gained will be directly applicable in their professional (teaching, research, data management) tasks. A couple of respondents remarked that some more time could be dedicated to the presentation of technical aspects of the platforms (which were in the sessions mainly addressed during the Q&A part), while one respondent also commented that the connection between SSHOC and the platforms could be explained in more detail.

In terms of organisation, the survey data showed that respondents learned about the event mainly through institutional/project newsletters (e.g., CLARIN, CESSDA, SSHOC newsletter) or by personal invitations. As written above, the majority praised the format of the webinar series coupling theoretical and practical aspects, however, some missed greater emphasis on the technical aspects of platforms (the answers to these kinds of questions were, however, provided, but mainly during the Q&A part of the sessions), while the others suggested a different delivery format with pre-recorded presentations and only live Q&A part. In addition, one respondent highlighted the need to adapt the presentation to the online format stating that: "/.../ it would be good if presenters remembered that listening to a digital

---

[11] For more information on the SSHOC Considerations for the Vocabulary Platforms Workshop, see section 3.3 of *D6.3 Final report on the outcome of the awareness raising workshops* (Vipavc Brvar, Inkret 2021, p. 10).

[12] Feedback for all three sessions has been collected with one survey.

event requires more focus, so that it's important to speak a little slower and not rush presentations" (respondent No. 8, Q7).

Post-event survey

Which vocabulary information session did you attend?
Introduction to CESSDA Vocabulary Service
Introduction to CESSDA Vocabulary Service
Introduction to Wikimedia;Introduction to Skosmos
Introduction to CESSDA Vocabulary Service
Introduction to CESSDA Vocabulary Service
Introduction to Wikimedia;Introduction to CESSDA Vocabulary Service
Introduction to Skosmos
Introduction to Wikimedia;Introduction to Skosmos
Introduction to Wikimedia;Introduction to Skosmos; Introduction to CESSDA Vocabulary Service

Overall, how would you rate the events?
Very Good
Very Good
Excellent
Good
Very Good
Good
Very Good
Very Good
Excellent

Where did you learn about the event?
more on the Service, examples to use, still in development
I already knew the CESSDA Vocabulary Service, but I found its application very interesting in the Swedish case
probably CLARIN newsflash
SSHOC web site
At work.
Invitation
SSHOC website
Through a CESSDA newsletter
Colleague information

Did the event meet your expectations?

Matched expectations
Matched expectations
Exceeded expectations
Less then expected
Matched expectations
Matched expectations
Matched expectations
Matched expectations
Matched expectations

Do you see this event having a positive impact on your work and how?

Compact information that will be used for my work, but also could be presented to students etc.

allowed me to better understand how to apply the tool

Having to set up knowledge bases from time to time, expanding my knowledge about vocabulary management was very helpful.

Not immediately or directly. Improved my awareness of the available services.

Yes. I obtained useful informations.

I now have a better overview of vocabulary solutions

Yes, it introduced me to tools I will have to manipulate

It gave me more understanding to the systems behind the tools I use or will use. I can't say yet whether it will have an impact on my work, as I work with the content in the systems, rather than with the technical aspects.

Yes

What did you like most about the event?

was interesting

the application and use of the tool

theoretical and practical aspects combined

Knowing more about available services and how they are adopted/applied.

Discussion.

Coverage of both the system and use cases.

Skos vocabulary presentation

It was informative.

What did you miss or could be done better at the event?

technical answers were kind of slow - like uncertainty was present.

more technical issues

Nothing

Everything was OK.

Technical info, like the underlying database system, was not always presented and only available in the Q&A

Maybe a little more context at the beginning : for instance, why is this topic relevant for SSHOC members ?

Both of the webinars I attended had speakers who talked really fast; it would be good if presenters remembered that listening to a digital event requires more focus, so that it's important to speak a little slower and not rush presentations.

Is there anything else you'd like to share about your experience (your suggestions/needs for expansions on the topic, etc.)?

No

No

all fine

No

Not at this moment.

After several zoom events I get the idea the following would be optimal: play a prerecorded presentation (nice timewise and presenters can practice and tune as much as they like), the presenter can directly answer questions in the chat and have a short plenary live Q&A. At the end of the session give each presentation a breakout room for 1 to 1 discussions.

I would have liked the possibility to attend a more technical demonstration at the end.

I would also have liked a little bit of additional reading for those of us who may not be fluent in the technical side of things." I understand that these were mainly informative webinars; for educational/training purposes, a workshop model of webinars, with a test environment, would be fantastic.

Go on

How well organized was the event (in terms of time management, length, venue/webinar system)?

Well organized

Excellently organized

Excellently organized

Excellently organized

Well organized

Excellently organized

Well organized

Well organized

Excellently organized

ANNEX 5

BEST PRACTICES FOR PRESERVING ORAL ARCHIVES WEBINAR REPORT

By Kristina Pahor de Maiti, CLARIN/UL-FF

Background

This report concerns the [SSHOC Webinar on Best Practices for Preserving Oral Archives](#), which was delivered in February 2021 as part of the SSHOC T6.2 awareness-raising events.

The webinar was a self-standing event aimed at raising awareness in the research community about the specifics of preserving oral archives. As such, it formed a coherent whole with other [SSHOC training events,](#) i.e. covering the processing and managing of spoken and sensitive data, as well as supporting the SSHOC objective of maximising reuse of knowledge and data by empowering the research community to become aware of best practices in the preservation of oral sources. By focusing on oral sources and targeting oral history and related communities, this webinar is related to the work performed in SSHOC WP3, WP4 (T4.4 Voice recorded interviews and audio analysis) and WP5 (5.4 Remote Access to Sensitive Data).

Sessions Overview & Format

Aim & Format

Oral sources are valuable and fragile objects compared to other historical sources such as artefacts and written documents, and thus rely on a specialised infrastructure for their preservation. The aim of the webinar was to share with the participants best practices for preserving such oral sources by presenting the preservation guidelines and two particular use cases, as well as offer them an opportunity to discuss these practices with the expert presenters. The webinar consisted of an introduction to CLARIN and the SSHOC project, continued with two presentations, and concluded with an informative discussion.

Organisers

The event was organised by CLARIN and SSHOC WP3 and WP6 members in the framework of the [CLARIN Café](#) series, which offers researchers, lecturers, students and topic experts an interactive space for discussion.

Speakers

The webinar was hosted by Francesca Frontini, (SSHOC WP3 and a member of the CLARIN ERIC board of directors) and Monica Monachini (SSHOC WP3 and the national coordinator of the Italian CLARIN consortium), who provided a CLARIN and SSHOC introduction at the beginning of the webinar, while the content in focus was delivered by Silvia Calamai (Associate Professor of glottology and general linguistics at the University of Siena and a member of the scientific board of CLARIN-IT), Véronique Ginouvès (head of the CLARIN-FR centre *Phonothèque MMSH – Maison méditerranéenne des sciences de l'homme*) and Claire Cialone-Gregoire (an archivist who edited a special issue of the *Sonorités* journal).

Participants

There were 45 attendees at the webinar[13], majority (approx. 90%) from the EU countries, around 5% from non-EU countries and remaining from the countries outside Europe. Overall, 20 countries were represented at the webinar. In terms of stakeholder groups coverage, the attendees (95% equally distributed among the categories) mainly identified with the categories "Researchers", "Research and E-infrastructures", "University and/or Research Performing Institution" and "Research Library and/or Archive". Remaining participants belonged to the categories "Civil society and/or Citizen Scientist" and "Private Sector and Industry Player".

Presentations & Discussions: Key Points

First session: Introduction to CLARIN and SSHOC

The webinar was introduced by Francesca Frontini and Monica Monachini. Francesca Frontini started with a brief introduction on the technical and knowledge sharing infrastructure of CLARIN and presented it in the context of oral history research. Followed on to this, Monica Monachini presented the SSHOC project and its goals, such as the creation of an EU-wide SSH Open Marketplace with tools and service conforming to Open Science and FAIR Principles,and situated the event with regard to SSHOC's ongoing efforts to provide up-to-date content to inform and educate the scientific communities.

For more see the presentation slides and the recording 1 and recording 2.

Second session: Italian Vademecum

This section was delivered by Silvia Calamai, who presented the Italian Vademecum – a document with best practices and guidelines for the preservation of oral sources, which are significantly more fragile

---

[13] It should be noted that the actual reach is higher since it should include the subsequent views of the webinar recordings on the SSHOC and CLARIN YouTube channel.

than other archival materials. She first described the main problematic aspects of dealing with oral data in research.

First, researchers were historically less careful with the handling of analogue carriers like tape recorders than they were with written documents, which resulted in many recorded oral interviews becoming lost (due to re-recording for instance). Second, oral sources are located at the crossroads of different fields of knowledge, being of interest not only to oral historians but also to linguists, sociologists, dialectologists, library scientists, and archivists, which can be an issue in the sense that each research field has different assumptions about how to best handle such data. Third, oral sources are not only of interest to academia but also to museums as well as non-professional users who for instance might be passionate about traditional music and therefore interested in having such sources easily accessible.

The Italian Vademecum arose as a response to such needs. It is a document made up of three parts that (1) describe a set of best practices for the creation of shared vocabularies describing the oral archives, (2) provide guidelines for the long-term conservation of oral archives, taking into account their peculiar fragility, and (3) offer information about the use and re-use of the sources, with guidelines on where to deposit them and how to make them accessible to others. It is noteworthy that the Vademecum was also the first document in history where Italian humanities institutions started working on a shared vocabulary for preserving oral sources. The final version of the Vademecum is set to be released in summer 2021.

For more information, see the presentation slides and the recording.

Third session: Two use cases from the Sonorités journal

The next presentation focused on the last two issues of the open access journal *Sonorités*, and was given by Véronique Ginouvès and Claire Cialone-Gregoire.

Véronique Ginouvès began the talk by introducing *Sonorités*, a journal that is published by the French Association for Sound, Oral, and Audiovisual Archives (AFAS), which explores the use of oral sources both in humanities and social sciences as well as in the wider cultural context. The presentation focused on the current, i. e. 48th issue of the journal, which addresses the preservation and curation practices in the ethnomusical collections of two museums in France. To showcase the collections, the presenter played a recording of *Ai rescontra ma mio*, a popular French burlesque song recorded on cylinder in 1900.

In the second part of the talk, Claire Cialone-Gregoire presented the special issue of the journal dedicated to the late Michel Seurat, a sociologist who worked in Lebanon and researched political aspects of Islam in the context of the Lebanese Civil War. The special issues focuses on oral interviews carried out in Lebanon and Syria by Seurat between 1981 and 1985, which among others include two survivor testimonies on the Hama massacre that took place in 1982. Such recordings from times of strife and

turmoil are immensely powerful historical artefacts that can also strongly affect, on an emotional level, the archivist curating them.

For more see the presentation slides and the recording.

Fourth session: Discussion

The presentations were followed by a lively discussion with a number of questions from researchers currently working on the publication of their own archives. For instance, a corpus linguist currently working with a multidisciplinary team made up of mathematicians, architects and historians to create the AFOr archive: a collection of different resources such as scanned oral documents, interviews, posters works and other promotional materials self-produced by Italian artisans in the 1960s. He remarked how hard it is to build a network of collaborators from different disciplines working on a single project. He therefore found the experiences of the Italian Vademecum, whose creation involved precisely such a network-building effort between researchers of different backgrounds, especially illuminating for his own endeavours. Lastly, a post-doctoral researcher from the University of Bolzano said she was inspired by the Café presentations to deposit her own spoken corpora in the ILC4CLARIN repository, which offers a platform for long-term preservation and re-use.

For more see the presentation slides and the recording.

Outcomes & Feedback

The webinar brought together experts from SSHOC and CLARIN and offered the participants up-to-date content on best practices and guidelines for the preservation of oral sources while situating those efforts in the work performed by the CLARIN infrastructure and the SSHOC project. By raising awareness about the specifics involved in the oral data processing, it helped solidify the understanding of the important aspects that need to be considered in order to maximise the reuse of existing sources among different scientific domains and non-academic stakeholders. Last but not least, since the webinar covered work and resources that focus on other languages than English, i.e. Italian and French, it also represents an important contribution in the view of promoting research in various natural languages.

Given the considerable number of participants and a lively discussion that evolved at the end, we can conclude that the webinar had addressed a timely topic in an attractive way. The post-event survey, despite the reminders sent, only returned two answers, but from those it is apparent that the event was successful (cf. Q8). According to one respondent, the event was also very nice due to its "informal and informative" format (cf. Q11), but there could be even more time dedicated to discussion (cf. Q10). All this points to an adaptive scientific community that is eager to learn, and as such represents an important target audience for SSHOC-related activities even in the future.

Post-event survey

1. What is your background?
Technical developer
(Computational) linguist

2. Where did you learn about the event?
CLARIN Newsflash
AISV Conference

3. Why did you decide to attend the event?
To learn about digital tools and methods relevant for oral history
To learn about CLARIN

4. Have you already used CLARIN resources for research?
No
Yes

5. If you have answered "yes" to Question 4, could you please point us to the results of your work (papers, datasets, blog posts, etc)?
I used BAS web services to phonologically and phonetically segment and annotated speech data

6. Do you plan to use CLARIN resources for research in the future?
Yes
Yes

7. What kind of digital tool would be the most useful for your research?
TEI compliant named entity recognition and linking as web service.

8. Overall, how would you rate the event?
Very good
Excellent

9. What is your impression of the length of the event?
It was optimal
It was optimal

10. What is your opinion about the format of the event?
Should be more interactive and more focused on discussions

11. What did you like the most about the event?
It was informal and informative.

13. What kind of impact will this event have on your work?

It was the first step to

14. What kind of follow-up would you like to see?

I'd like to have practical suggestions on oral data archiving

ANNEX 6

The Multilingual Corpus of Survey Questionnaires

Webinar Report

By Veronika Keck, GESIS – Leibniz Institute for the Social Sciences

Background

This report concerns a webinar organised by the Research and Expertise Centre for Survey Methodology at Universitat Pompeu Fabra (SSHOC task 4.2 Preparing tools for the use of Computer Assisted Translation) and supported by SSHOC T6.2.

The event titled "The Multilingual Corpus of Survey Questionnaires" was held on the 6th of April 2021. The announcement and event information is available here and the presentation has been uploaded to the dedicated SSHOC Community on Zenodo.

Organisers

From the SSHOC project task 6.2 Veronika Keck (GESIS), Irena Vipavc, and Ana Inkret (UL-ADP) were involved in the organisation of announcement, post-event survey, and reporting. For the presentation, the speaker was Danielly Sorato (UPF). The webinar was technically supported by the UPF team.

Workshop Overview & Format

The virtual workshop was streamed via ZOOM and was designed as a one-hour event. It started with an introduction by the speaker. This was followed by a 30-minutes long presentation and a 30-minutes Q&A part at the end.

Participants

Around 55 participants attended the webinar. Unfortunately, only little information is available regarding the stakeholder categories (see results from the evaluation below) and no information is available regarding the organisation types participants belong to due to the fact that no registration was required for the event. The webinar was included in the Research and Expertise Centre for Survey Methodology

at Universitat Pompeu Fabra (RECSM) series of monthly webinars and was subject to an internal policy whose aim was to make the event as open and as participant-friendly as possible, resulting in less supervision over the attendance as customary for SSHOC events. The link to the event was publicly accessible on the RECSM website. Moreover, the webinar was not recorded. The responsible contact at UPF Diana Zavala -Rojas provided the following explanation via email on April 06, 2021:

"The idea is to not ask for registration nor recordings. In this pandemic year almost all events are being recorded and that practice is not necessarily followed comfortably by all institutions/research centres, etc. When the webinar series was started, at RECSM/UPF we agreed with this policy. The presentation can be shared and uploaded to Zenodo, we are just not asking people that this will be recorded."

The SSHOC Task 6.2 team and WP6 leader Vasso Kalaitzi emphasised that the "no registration, no recording" policy does not comply with SSHOC event guidelines. Moreover, the concern that having a Zoom link publicly available on a webpage might cause security issues was raised not only by the 6.2 teams but also by a participant of the webinar. Nevertheless, the UPF team proceeded with their internal policies.

Presentation: Key Points

The Multilingual Corpus of Survey Questionnaires (MCSQ) is the first publicly available database of survey questionnaires texts. The corpus is compiled from the European Social Survey (ESS), the European Values Study (EVS), and the Survey of Health, Ageing and Retirement in Europe (SHARE) questionnaires. The first version (entitled Ada Lovelace) was released in 2020.

The recently released Version 2.0 (entitled Mileva Marić-Einstein) includes the English source language and their translations into Catalan, Czech, French (with versions produced for France, Switzerland, Belgium, and Luxembourg), German (produced for Austria, Germany, Switzerland, and Luxembourg), Norwegian, Portuguese, Spanish and Russian (produced for Israel, Latvia, Lithuania, Russia, Ukraine, Estonia).

The corpus currently holds more than 3.5 million words and approximately 657.000 sentences. Nearly 80% of the corpus is sentence aligned, meaning that the source sentences in English are linked to their translations in different languages.

MCSQ is implemented as an open-source and open-access research resource and will represent an important resource for corpus linguists, computational linguists, statisticians, typologists, social scientists, as well as translation scholars and localizers. The MCSQ is a relevant digital artefact that allows users to quickly search and compare survey items in a myriad of languages.

The MCSQ provides linguistically aligned survey questions, answer scales, instructions, etc., meaning that segments of English source text have been matched with their translated version into different languages. Moreover, these survey questionnaires have valuable metadata associated with them, which is not present in PDF questionnaires, to identify the questions, their linguistic components, and the survey projects. The metadata include item type indication, country, language, year, etc.

The comparison between items/segments is fully automated. Therefore, the corpus is a valuable resource in the translation process for the new rounds or waves in international comparable surveys, e.g., by populating repeated questions with existing translations from previous rounds. Moreover, the corpus might be used for further development of computer-aided language tools and resources in the social sciences and humanities.

Q & A session

During the Q & A discussion the speaker received not only questions about the operationalisation of the corpus but also some ideas and suggestions for further development. The main questions and ideas are summarised below.

Can data from the corpus be exported for linguistic analyses in statistical tools?
The corpus has the functionality to export data in a CSV format. That format allows users to perform any linguistic analyses using statistical tools and functions, e.g., tidy data in R.
What was a segmentation rule for question batteries?
Question batteries are often used in questionnaires, the structure of such batteries includes a question part and several answer options. The corpus duplicates the question part for each answer option in the battery, but the answer option stays the same.
The corpus has logically created variable identifiers; however it would be interesting to know if you plan to connect those identifiers with existing QDDT variable identifiers for ESS items?
Although these identifiers are not part of the corpus because during the development stage this was not a corpus requirement, it would be possible to do mapping and add this information as metadata.
Are there any plans to manually correct errors in alignments?
There are no plans to correct errors in automatic alignments due to the high amount of corpus data.
Greatly appreciated was an idea to set word search as the default function because it makes it possible to exclude noise in the data. For searching the alignment, it was recommended to use the filter "target language". A suggestion to connect collocations, e.g., thick and ill, will be taken into consideration during the next improvement steps together with the list of functionalities pending for implementation.

Outcomes & Further Development

The MCSQ is an important digital resource that enables automatic linguistic analysis in survey text. Now, researchers interested in using the questionnaires and/or alignment data do not have to look for text in the PDFs, copy and paste, pre-process and align the text themselves since such treated texts are available in MCSQ (free of charge). Moreover, building resources such as the MCSQ is an important first step for developing other resources for downstream computational linguistic applications, such as domain-specific machine translation models, predictors of question complexity, etc.
The MCSQ and its interface were completely developed using open-source tools, no proprietary tools used in any of the steps. All the developed code comprising the corpus compiling steps, from pre-processing to storage and data annotation, is publicly available to the on the GitHub repository.
Goals for future work:
- New survey items
- New interface functionalities

- New linguistic annotations
- Testing the corpus in the context of translation procedures (translation memory)

Media coverage and promotion of SSHOC

Announcements were shared on the SSHOC website, mailing lists of the SSHOC Training Community, partner institutions, Basecamp, and on several Social Media channels. The presentation was uploaded to Zenodo. A blog post of the event was prepared afterwards and has been distributed through the SSHOC networks.

Results from the evaluation

14 participants filled in the general SSHOC evaluation form (see Appendix 1). Of them, 50% were from universities and research performing institutions, five people identified themselves as researchers, one participant came from research & e-infrastructure and EOSC thematic clusters, and one from the category Private sector & industry players (see graphic below). Other stakeholder categories were not chosen.
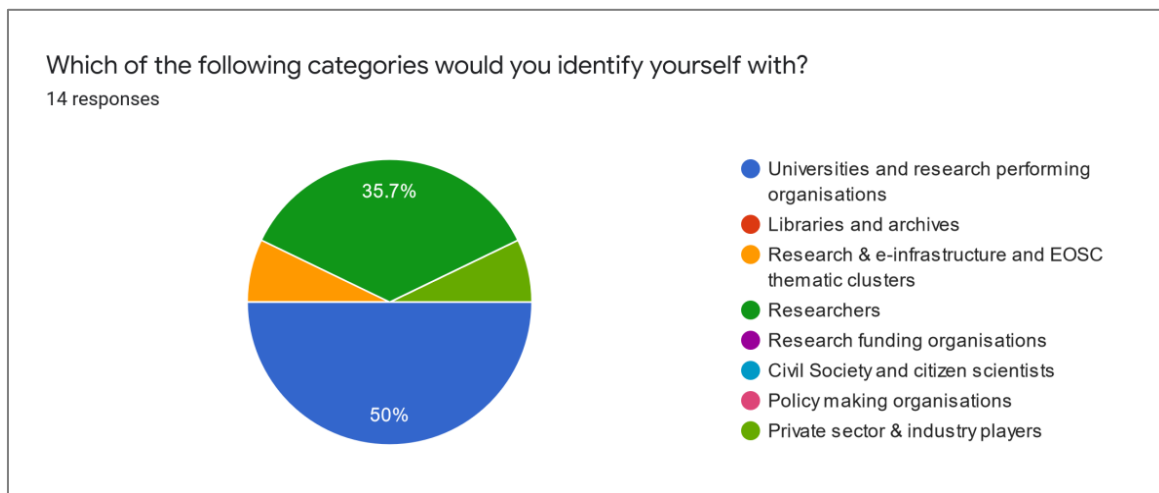
Overall, the event was positively evaluated. Around 57% of the participants evaluated the event as "very good" and 42,9% said it was "excellent" when rating the overall event. The organisation was evaluated as "excellently organised" by eight participants, as "very well organised" by five participants, and "well organised" by one respondent.

Feedback and rating with regards to whether or not the event met participants' expectations were positive. Six participants agreed that the workshop greatly exceeded their expectations, seven participants reported that the event exceeded their expectations, and one reported that it matched their expectations.

The most common positive feedback was given with regards to the clarity and brevity of the presentation and focus on the practical aspects of the corpus, followed by the informative character and clearness of the Q&A part of the event.

Responding to the question of what could have been better, participants indicated that "It would have been easier to follow the presentation if the presenter had immediately started with the goal of the project or giving an applied example". Participants wished for more examples on how the corpus can be used and hands-on experience, that is "interactive elements to engage the audience".

What was further positive is that all respondents who have filled out the post-event survey have agreed that the event had a positive impact on their work. Respondents reported that they found that "the corpus data could be relevant for their future work, i.e. could be used in future scholarly research" or "as the alignment tool that can be handy for quickly checking translations". They reported that the webinar "explained translation technology concepts to a layperson" very well and gave a general overview of how technology can be applied when looking for existing translations of survey questions.
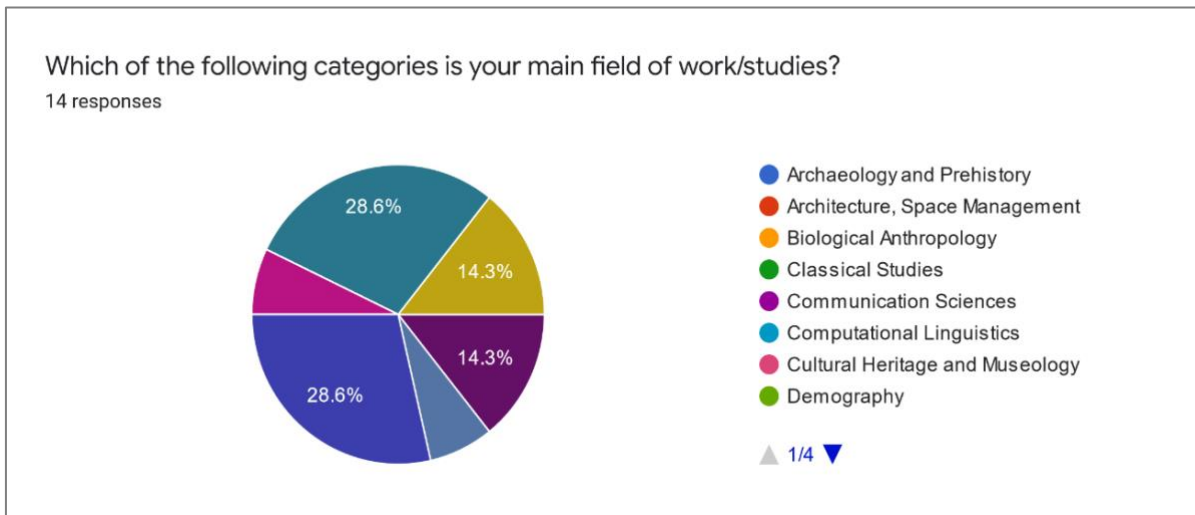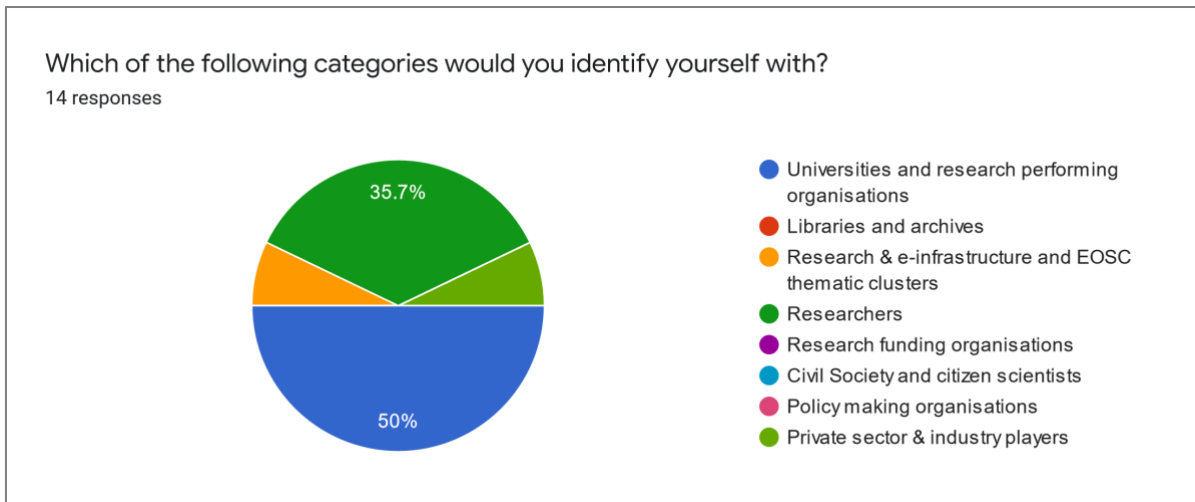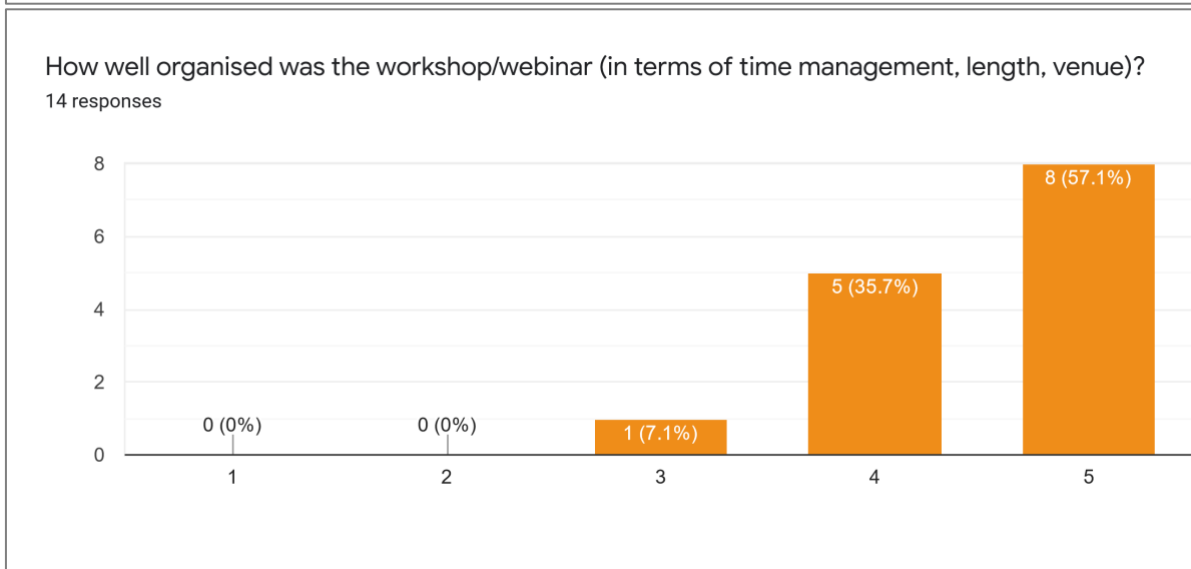
Suggestions for future webinars

In general, when organising a webinar about tools it would be better to have a practical part, hands-on experience part within a webinar. It is recommended to have more examples during the presentation, however, due to the time limit of one hour, it could be out of scope.

A possible solution is to organise a follow-up event, hands-on tutorial about the usage of the tool and its possible implementation in different research scenarios.

On the organisational part, it is recommended to communicate more about the existing SSHOC guidelines for the events and the need for participating organisations to comply with them. A possible solution is having this communication on a higher managerial or organisational level. Another solution is to communicate the SSHOC event guidelines during bigger project meetings.

Results from the post-event evaluation survey

Which of the following categories would you identify yourself with?
14 responses

- Universities and research performing organisations
- Libraries and archives
- Research & e-infrastructure and EOSC thematic clusters
- Researchers
- Research funding organisations
- Civil Society and citizen scientists
- Policy making organisations
- Private sector & industry players

Which of the following categories is your main field of work/studies?
14 responses

- Archaeology and Prehistory
- Architecture, Space Management
- Biological Anthropology
- Classical Studies
- Communication Sciences
- Computational Linguistics
- Cultural Heritage and Museology
- Demography

1/4 ▼

Overall, how would you rate the workshop/ webinar?

14 responses

Where did you learn about the workshop/webinar ?

14 responses

Legend:
- SSHOC website
- Social Media
- Colleagues
- Mailinglist
- Event
- I am a member of RECSM/UPF
- I am involved in SSHOC myself

How well organised was the workshop/webinar (in terms of time management, length, venue)?

14 responses

Did the workshop/webinar meet your expectations?

14 responses

What did you hope to gain from the workshop/webinar?
Helicopter view of the MCSQ (and how it is explained to researchers). Very helpful.
Understanding what the project is about and how to apply it
Insights into the corpus, how it is set up, how it looks, how it could be used
General insights into project
.
knowing the basics of the corpus
More information about the MCSQ, its functionalities and the surveys covered
what is already developed and what are future plans
an overview of MCSQ
Insights into MCSQ
Better understanding the concept and the practice of the corpus
An understanding on what was inside the webinar
Some news on the SSHOC project and the usefulness of this tool
Understanding what the project is about and how to apply it

Do you see this workshop/webinar having a positive impact on your work and how?
Yes - new ways of explaining translation technology concepts to a layperson
It is just good to know about the project
yes, my field is questionnaire translation, so this helps when looking for existing translations
Yes, awareness of the corpus data could be relevant for my future work, i.e. could be used in future scholarly research
.
by making researchers accessing the corpus
The MCSQ has some potential implication for my work, so yes the impact is positive
/
No, as I already knew about it
Can direct researchers to this resource

Yes
Maybe as the alignment tool can be handy for quickly checking translations, something I usually do by checking PDFs.
For now, neither positive nor negative
It is just good to know about the project

What did you like most about the workshop/webinar?
Clarity and brevity of the presentation.
Application und mechanism
Slide: Uses scenarios of the corpus
nicely structured presentation focusing also on the practical aspect, easy to follow
.
the discussion
Q and A
/
that the structure of the MCSQ and an example of a search were clearly presented
Presentation & discussion
Informative
Very well organised and presented!
Clearness
Application und mechanism

What did you miss or could be improved at the workshop/webinar?
Introduce an expert in the field as discussant, who prepares comments on the presentation in advance.
It would have been easier to follow the presentation, if the presenter had immediately started with the goal of the project or giving an applied example.
n/a
Presentation of an example of research with the data would have interesting
.
nothing comes to mind
n/a
would be good to have more examples on how corpora can be used
-
Some small interactive elements to engage the audience :-)
non
Nothing really in the workshop itself, although of course if the corpus were bigger that would be nice (and obviously, way more difficult).
More applications
It would have been easier to follow the presentation, if the presenter had immediately started with the goal of the project or giving an applied example.