

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

Deliverable 6.3 Final report on the outcome of the awareness raising workshops

Dissemination Level	PU
Due Date of Deliverable	31/12/2021 (M36)
Actual Submission Date	09/12/2021
Work Package	WP6 – Fostering Communities, Empowering Users, & Building Expertise
Task	Task 6.2 Fostering Communities: Engaging New & Existing Users
Type	Report
Approval Status	Waiting EC approval
Version	1.0
Number of Pages	p.1 – p.76

Abstract:

This report summarizes the organisation and outcomes of 9 awareness raising workshops, organised with the aim of promoting SSHOC services and tools, gathering feedback from various audiences and informing the work of the project partners.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



History

Version	Date	Reason	Revised by
0.1	05/11/2021	First draft	Ana Inkret (CESSDA/UL-ADP)
0.2	15/11/2021	Collecting and cleaning all annexed reports	Ana Inkret (CESSDA/UL-ADP)
0.3	17/10/2021	First internal version	Ana Inkret (CESSDA/UL-ADP), Rosie Allison (LIBER), Irena Vipavc Brvar (CESSDA/UL-ADP)
0.4	22/10/2021	Peer / internal review	Vasso Kalaitzi (DANS KNAW); Astrid Verheusen (LIBER)
0.5	29/11/2021	Second internal version with Addressed peer review comments send to coordination office	Martina Drascic Capar (CESSDA)
0.6	8/12/2021	Final annex added, internal review	Ana Inkret (CESSDA/UL-ADP), Rosie Allison (LIBER)
0.7	09/12/2021	Third internal version with final event added send to coordination office	Martina Drascic Capar (CESSDA)
1.0	09/12/2021	Final version; all comments addressed and send to project coordinator for review / publication	Ana Inkret, Irena Vipavc Brvar (CESSDA/UL-ADP)

Author List

Organisation	Name	Contact Information
CESSDA/UL-ADP	Ana Inkret	ana.inkret@fdv.uni-lj.si
LIBER	Rosie Allison	rosie.allison@libereurope.org
CESSDA/UL-ADP	Irena Vipavc Brvar	irena.vipavc@fdv.uni-lj.si

List of Contributors

Organisation	Name	Contact Information
CLARIN ERIC	Iulianna van der Lek	i.vanderlek@uu.nl
CESSDA/GESIS	Veronika Keck	Veronika.Keck@gesis.org
CESSDA/GESIS	Judith Wehmeyer	Judith.Wehmeyer@gesis.org

Executive Summary

The awareness raising workshops were part of the SSHOC stakeholder engagement and awareness raising activities, aimed towards fostering communities in the Social Sciences and Humanities. The aim of the workshops was to raise awareness on SSHOC services and tools in development, gather feedback from various participants (e.g., data producers, data users, data experts, researchers, research librarians, secure data facility professionals, policy makers and civil society), and to inform project work based on the received feedback.

Awareness workshops introduced SSHOC services and tools that were still under development and were therefore a way to maximise active contribution of the Social Sciences and Humanities (SSH) communities in the project. The content of the workshops was delivered by task partners and invited experts, while task 6.2 “Fostering Communities: Engaging New & Existing Users” provided communicational and technical support, assisted moderators in facilitating the discussions, and applied a standardized workflow to document the events. After the start of the COVID-19 pandemic, the workshops were adapted to take place online, while keeping the interactive elements and more extensive character of a workshop, as compared to webinars.

The awareness raising workshops comprised a total of **9 events**:

1. LR4SSHOC: LREC2020 workshop about Language Resources for the SSH Cloud
2. Agile Development of the SSH Open Marketplace: User Workshop, 30 June 2020
3. SSHOC Considerations for the Vocabulary Platforms, 6 November 2020
4. SSHOC Speech-to-text Workshop - Linking Social Survey and Linguistic Infrastructures Through Speech Interviews, 16 April 2021
5. SSHOC Archaeological Case Study Workshop - The Roman Theatre in Catania from Survey to Interactive 4D Visualisation, 25 May 2021
6. SSHOC Dataverse Translation Workshop, 2 June 2021
7. SSHOC Dataverse Translation follow-up event 1, 8 September 2021
8. SSHOC Workshop: Providing canonical training materials for secure data facility professionals, 21 September 2021
9. SSHOC Dataverse Translation follow-up event 2, 7 December 2021

The awareness workshops were a successful format for the introduction of SSHOC developments and interaction with the relevant audiences. The guiding principle of the workshops was to present the work that was still in progress, which meant that the organisers were able to receive valuable feedback from the participants that could be applied to future work. For the SSH communities, the awareness workshops were an opportunity to engage with the developers of innovative products, learn of the latest progress in their fields, and contribute to the development of SSHOC services and tools.

Abbreviations and Acronyms

EOSC	European Open Science Cloud
GA	Grant Agreement
LR4SSHOC	Language Resources for Social Science and Humanities Open Cloud
LREC2020	Language Resources and Evaluation Conference 2020
SSHOC	Social Science and Humanities Open Cloud
SSH	Social Science and Humanities
WP	Work package
WP2	Communication, Dissemination, and Impact
WP3	Lifting Technologies and Services into the SSH Cloud
WP4	Innovations in Data Production
WP5	Innovations in Data Access
WP6	Fostering Communities, Empowering Users, & Building Expertise
WP7	Creating the SSH Open Marketplace
T	Task

Table of Contents

1. Introduction	6
2. Organisation	6
3. Summary of the Workshops	7
3.1 LR4SSHOC: LREC2020 workshop about Language Resources for the SSH Cloud	8
3.2 Agile Development of the SSH Open Marketplace: User Workshop	9
3.3 SSHOC Considerations for the Vocabulary Platforms	10
3.4 SSHOC Speech-to-text Workshop - Linking Social Survey and Linguistic Infrastructures Through Speech Interviews	11
3.5 SSHOC Archaeological Case Study Workshop - The Roman Theatre in Catania from Survey to Interactive 4D Visualisation	12
3.6 SSHOC Dataverse Translation Workshop	12
3.7 SSHOC Dataverse Translation follow-up event 1	13
3.8 SSHOC Workshop: Providing canonical training materials for secure data facility professionals	14
3.9 SSHOC Dataverse Translation follow-up event 2	14
4. Outcomes and Conclusions	15
5. References	16
6. List of annexes	16
ANNEX 1 - Agile development of the SSH Open Marketplace: User Workshop: Workshop Report	17
ANNEX 2 - SSHOC Considerations for Vocabulary Platforms: Workshop Report	27
ANNEX 3 - SSHOC Speech-to-text Workshop - Linking Social Survey and Linguistic Infrastructures Through Speech Interviews: Workshop Report	43
ANNEX 4 - SSHOC archaeological case study Workshop - The Roman theatre in Catania from survey to interactive 4D visualisation: Workshop Report	49
ANNEX 5 - SSHOC Dataverse translation workshop: Workshop Report	62
ANNEX 6 - SSHOC Dataverse Translation Follow-up Events: Report of the Workshops	70
ANNEX 7 - SSHOC Workshop: Providing canonical training materials for secure data facility professionals: Workshop Report	74

1. Introduction

The awareness raising workshops were part of the SSHOC engagement and awareness raising activities, which took part along with awareness raising webinars¹ and the stakeholder series events² in the context of WP6 - Fostering Communities, Empowering Users and Building Expertise. The aim of the workshops was to promote SSHOC services and tools under development, gather feedback from various audiences (e.g., data producers, data users, data experts, researchers, research librarians, secure data facility professionals, policy makers and civil society), and to inform project work based on the received feedback. The workshops addressed different target audiences and disciplinary perspectives.³

The organisation of the workshops was based on foundations defined by previous deliverables and milestones. Key performance indicators for the workshops were outlined in *Deliverable 6.1 Community Engagement Strategy*.⁴ *Milestone 38 Launch Community Engagement Strategy*⁵ reported on the survey among the SSHOC project partners regarding the development of their products and their planned engagement activities which helped inform the organisation of task 6.2 work, including the workshops.

2. Organisation

Close collaboration with the project partners and alignment with WP leaders was key in the organisation of the workshops. For the project partners, the awareness workshops were an opportunity to promote their work-in-progress and discuss their approach and basic concepts with prospective user communities and intermediaries to gather valuable feedback that could be implemented in the ongoing work. Awareness workshops most often introduced services and tools that were still under development to specific communities of future users, which distinguished them from dissemination activities which focus on finished products, ready to use as they are, training activities that aimed at empowering users and building expertise, or stakeholder events that engaged larger cross-stakeholder participant in the discussion on broader matters raised by the project. The workshops were therefore a way to maximise community involvement and active contribution to the project.

¹ The webinars will be the subject of *Deliverable 6.4 Report on awareness webinars 1-6*.

² Irena Vipavc Brvar, & Ana Inkret. (2021). D6.5 Report on Stakeholder Series events. Zenodo. <https://doi.org/10.5281/zenodo.5638651>.

³ 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>.

⁴ Ibid.

⁵ Irena Vipavc Brvar. (2019). MS38 Launch Community engagement strategy (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.4584616>.

The content of the workshops was in the domain of those WPs or project tasks that found this approach helpful for their work, while task 6.2 provided technical support, facilitated the discussions, and applied a standardized workflow to promote and document the events. To organise the work during the project, each partner from task 6.2 was matched to a specific WP. This way it was possible to follow up with the development of services and tools to find the appropriate moment for the organisation of an awareness workshop or a webinar. Event checklist, templates, promotion workflow, evaluation forms and materials used at events were prepared in collaboration with WP2.

After the start of the COVID-19 pandemic, the workshops were adapted to take place online. The demand for geographical distribution of the workshops was made redundant as the online events were open and accessible to all. Special effort was made to ensure that the virtual setting fulfilled the purpose of the workshops as planned at the start of the project. The defining component of the workshops and one that was most influenced by the online format was the interaction and involvement of the participants. All the events kept the characteristics of a workshop where the participants were given the opportunity to actively contribute and exchange opinions, ideas, and comments. The potential downsides of online communication were mitigated by engaging the participants with the online real-time feedback tool Mentimeter or by dividing the participants in smaller groups to encourage discussion, using the breakout rooms functionality in the Zoom platform. Although organising online workshops was a challenge for the community building, the team took the opportunity to support the interaction of the project with a wide range of workshop participants even during the pandemic, facilitated their active involvement in the project and still fostered networking opportunities, albeit in online formats.

A workshop typically consisted of presentations, active involvement or discussion, and a conclusion with a summary of the main discussion points. These conclusions were then taken as input and action points for the further development of SSHOC products. The discussions were generally not recorded to encourage a livelier exchange and a more informal atmosphere of the working environment. Video recordings of the presentations were published on the SSHOC YouTube channel⁶ and the slides were published in the SSHOC project Zenodo community⁷ after each workshop. The post-event activities concluded with a blog post, prepared by the content organizers and/or task 6.2 partners and posted in the news section of the SSHOC website.

3. Summary of the Workshops

The awareness raising workshops comprised a total of 9 events, 3 more than originally planned in the GA or the engagement strategy. Their scope can be summarized in these points:

⁶ SSHOC YouTube channel: <https://www.youtube.com/channel/UCw-mY8v84yeHW2z4KG3ZLtA>, accessed 16 November 2021.

⁷ SSHOC Zenodo community: <https://zenodo.org/communities/sshoc>, accessed 16 November 2021.

- Disciplines: 3 workshops from the perspective of humanities, 1 from the perspective of social sciences, 5 cross-disciplinary or domain agnostic.
- SSHOC project work packages involved in the organisation: WP2, WP3, WP4, WP5, WP6, and WP7.
- Number of stakeholders reached:
 - Universities and research institutes: 77
 - Researchers: 64
 - Research and e-Infrastructures, EOSC thematic clusters: 56
 - Research libraries and archives: 57
 - Private sector and industry: 6
 - Research funders: 2
 - Policy makers: 3
 - Civil society and citizen scientists: 1
 - Other and uncategorized: 204⁸
 - Total: 470 participants

The first workshop on the list had to be cancelled due to the COVID-19 pandemic. Instead, the proceedings of this workshop were published as a series of peer-reviewed papers, which can be found online.⁹

The following section provides an overview of the awareness raising workshops which took place over the period from May 2020 to December 2021, providing a summary of the workshop content, attendance, and links to relevant materials.

3.1 LR4SSHOC: LREC2020 workshop about Language Resources for the SSH Cloud

Workshop at the 12th Language Resources and Evaluation Conference

Date and venue	Planned for 11 May 2020, the workshop was cancelled due to the epidemic. Extended proceedings were published in June 2020.
Topics	Language resources for SSH Open Cloud and language technologies in the context of SSH research
Links to materials	

⁸ The uncategorized participants stem from the workshop at ICTeSSH (section 3.2). Those who did not respond to the Mentimeter survey are included in the number.

⁹ Broeder, Daan, Eskevich, Maria, Monachini, Monica (2020). Workshop about Language Resources for the SSH Cloud. Proceedings. <https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/LR4SSHOCbook.pdf> (accessed 16 November 2021).

Announcement	https://www.sshopencloud.eu/lr4sshoc-lrec2020-workshop-about-language-resources-ssh-cloud
Main event homepage	https://lrec2020.lrec-conf.org/en/
Proceedings	https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/LR4SSHOCbook.pdf
Blog post	https://sshopencloud.eu/news/publication-proceedings-lr4sshoc-workshop
Audience	
Targeted audience	Researchers, universities and research institutes, research and e-infrastructures, EOSC thematic clusters, private sector and industry (specifically industrial researchers, digital language resources and technology providers, software developers, and SSH representatives)

3.2 Agile Development of the SSH Open Marketplace: User Workshop

Date and venue	30 June 2020, online
Topics	User feedback and engagement for the SSH Open Marketplace
Links to materials	
Announcement	https://www.sshopencloud.eu/events/agile-development-ssh-open-marketplace-user-workshop
Main event homepage	https://ictessh.uns.ac.rs/ictessh-2020/
Presentations	https://doi.org/10.5281/zenodo.3928148
Video recording	https://youtu.be/AZTrzHtiI3Q
Proceedings	https://doi.org/10.1051/itmconf/20203304001
Audience	
Targeted audience	Researchers, research libraries and archives, universities and research performing organisations, research and e-infrastructures, EOSC thematic clusters

Total participants	246
By stakeholder category	Universities and research institutes 17 Research libraries and archives 11 Research and e-Infrastructures, EOSC thematic clusters 9 Researchers 6 Private sector and industry 3 Research funders 1 Policy makers 1 Civil society and citizen scientists 1 Uncategorized 197
Event report	See <i>Annex 1</i>

3.3 SSHOC Considerations for the Vocabulary Platforms

Date and venue	06 November 2020, online
Topics	Vocabulary-related activities within the SSHOC project, alignment between different vocabulary platforms and interoperability challenges
Links to materials	
Announcement	https://www.sshopencloud.eu/events/sshoc-considerations-vocabulary-platforms
Presentations	See https://www.sshopencloud.eu/events/sshoc-considerations-vocabulary-platforms
Video recording	https://youtube.com/playlist?list=PLIKmS5dTMgw1EgPoFtzqa0rrcDVCE3N7Q
Blog post	https://sshopencloud.eu/news/workshop-notes-sshoc-requirements-vocabularies-and-vocabulary-management-platforms
Audience	
Targeted audience	Research libraries and archives, universities and research institutes, research and e-infrastructures, EOSC thematic clusters (specifically academic or industry professionals who operate, integrate, and manage vocabulary platforms)
Total participants	61

By stakeholder category	Universities and research institutes 20 Researchers 14 Research and e-Infrastructure, EOSC thematic clusters 13 Research libraries and archives 13 Private sector and industry 1
Event report	See <i>Annex 2</i>

3.4 SSHOC Speech-to-text Workshop - Linking Social Survey and Linguistic Infrastructures Through Speech Interviews

Date and venue	16 April 2021, online
Topics	Integration of social survey and linguistic infrastructures
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-speech-text-workshop-linking-social-survey-and-linguistic-infrastructures-through
Presentations	https://doi.org/10.5281/zenodo.4700104
Video recording	https://youtu.be/8OmXr0Fr7DY
Blog post	https://sshopencloud.eu/news/new-report-published-sshoc-speech-text-workshop-%E2%80%93-linking-social-survey-and-linguistic
Audience	
Targeted audience	Researchers, research libraries and archives, universities and research institutes, research and e-infrastructures
Total participants	53
By stakeholder category	University and Research Institutions 23 Researchers 23 Research Libraries and Archives 5 Other 3 Civil society and citizen science 1 Policy makers 1 Private sector and industry 1
Event report	See <i>Annex 3</i>

3.5 SSHOC Archaeological Case Study Workshop - The Roman Theatre in Catania from Survey to Interactive 4D Visualisation

Date and venue	25 May 2021, online
Topics	SSHOC supported technical developments for archaeological data
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-archaeological-case-study-workshop-roman-theatre-catania-survey-interactive-4d
Presentations	https://doi.org/10.5281/zenodo.5013566
Video recording	https://youtu.be/Nz0Q3HORgf4
Blog post	https://www.sshopencloud.eu/news/workshop-notes-sshoc-archaeological-case-study-roman-theatre-catania-survey-interactive-4d
Audience	
Targeted audience	Researchers (specifically archaeologists and heritage scientists)
Total participants	33
By stakeholder category	Researchers 14 Universities and research institutes 9 Other 4 Research libraries and archives 2 Research and e-infrastructures, EOSC thematic clusters 2 Private sector and industry 1
Event report	See <i>Annex 4</i>

3.6 SSHOC Dataverse Translation Workshop

Date and venue	2 June 2021, online
Topics	Translation of the user interface to national languages with the help of the Weblate tool, developed within the SSHOC project
Links to materials	

Announcement	https://sshopencloud.eu/events/sshoc-dataverse-translation-workshop
Presentations	http://doi.org/10.5281/zenodo.5017398
Video recording	https://youtu.be/Wilxney3ORU
Blog post	https://sshopencloud.eu/news/sshoc-workshop-notes-sshoc-dataverse-translation-workshop
Audience	
Targeted audience	Research libraries and archives, universities and research performing organisations, research and e-infrastructures
Total participants	23
By stakeholder category	Research libraries and archives 9 Researchers 5 Universities and research institutes 5 Research and e-infrastructures, EOSC thematic clusters 4
Event report	See <i>Annex 5</i>

3.7 SSHOC Dataverse Translation follow-up event 1

Date and venue	8 September 2021, online
Topics	Translation of the Dataverse user interface with the Weblate tool
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-dataverse-translation-follow-event
Blog post	https://sshopencloud.eu/news/workshop-notes-sshoc-dataverse-translation-follow-event
Audience	
Targeted audience	Research libraries and archives, universities and research performing organisations, research and e-infrastructures
Total participants	15
By stakeholder category	Research libraries and archives 8 Research and e-infrastructures and EOSC thematic clusters 3

	Researchers 2 Universities and research institutes 1 Research funder 1
Event report	See <i>Annex 6</i>

3.8 SSHOC Workshop: Providing canonical training materials for secure data facility professionals

Date and venue	21 September 2021, online
Topics	Canonical training materials for users of secure data facilities
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-workshop-providing-canonical-training-materials-secure-data-facility-professionals
Presentations	https://doi.org/10.5281/zenodo.5541587
Video recording	https://youtu.be/svx-n4LEh6M
Blog post	https://sshopencloud.eu/news/sshoc-workshop-notes-providing-canonical-training-materials-secure-data-facility-professionals
Audience	
Targeted audience	Research libraries and archives, research and e-infrastructures, EOSC thematic clusters
Total participants	39
By stakeholder category	Libraries and archives 13 Researchers 12 Research and e-infrastructures 8 Universities and research institutes 6
Event report	See <i>Annex 7</i>

3.9 SSHOC Dataverse Translation follow-up event 2

Date and venue	7 December 2021, online
-----------------------	-------------------------

Topics	Exchange of the users' translating experiences, issues and ideas
Links to materials	
Announcement	https://sshopencloud.eu/events/sshoc-dataverse-translation-follow-event-0
Blog post	https://www.sshopencloud.eu/news/sshoc-workshop-notes-dataverse-translation-follow-event#overlay-context=news
Audience	
Targeted audience	Research libraries and archives, universities and research performing organisations, research and e-infrastructures
Total participants	13
By stakeholder category	Research libraries and archives 9 Universities and research institutes 2 Policy makers 1 Research and e-infrastructures, EOSC thematic clusters 1
Event report	See <i>Annex 6</i>

4. Outcomes and Conclusions

The awareness workshops were a successful format for the introduction of SSHOC developments and interaction with the relevant communities. The guiding principle of the workshops was to present the SSHOC work that was still in progress, which meant that the organisers were able to receive valuable feedback from the participants which could be applied to future work in a bottom-up approach or contribution to development. Though not without its challenges, the benefits of informing and actively involving the communities of what was being developed within the project and receiving their input added to the bottom-up and community-led character of the project, ensuring that its products are relevant for the end users and adapted to their needs.

The organisers made sure to maximise the potential of online communication, so participants from all of Europe were able to take active part in the events. For the SSH communities, the awareness workshops were an opportunity to engage with the developers of innovative products, to learn about the latest progress in their fields, and to contribute to the development of SSHOC services and tools.

The workshops produced a rich portfolio of materials that will continue to be available online and might be of special interest to developers who are looking to engage the end-user in the planning and production phases of their products.

5. References

Broeder, Daan, Eskevich, Maria, Monachini, Monica. (2020). Workshop about Language Resources for the SSH Cloud. Proceedings. Accessible at <https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/LR4SSHOCbook.pdf>, last accessed 16 November 2021.

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>.

Vipavc Brvar, Irena. (2019). MS38 Launch Community engagement strategy (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.4584616>.

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

6. List of annexes

Annex 1: Agile development of the SSH Open Marketplace: User Workshop: Workshop Report

Annex 2: SSHOC Considerations for Vocabulary Platforms: Workshop Report

Annex 3: SSHOC Speech-to-text Workshop - Linking Social Survey and Linguistic Infrastructures Through Speech Interviews: Workshop Report

Annex 4: SSHOC archaeological case study Workshop - The Roman theatre in Catania from survey to interactive 4D visualisation: Workshop Report

Annex 5: SSHOC Dataverse translation workshop: Workshop Report

Annex 6: SSHOC Dataverse Translation Follow-up Events: Report of the Workshops

Annex 7: SSHOC Workshop: Providing canonical training materials for secure data facility professionals: Workshop Report

ANNEX 1

AGILE DEVELOPMENT OF THE SSH OPEN MARKETPLACE: USER WORKSHOP WORKSHOP REPORT

By Ana Inkret (CESSDA/UL-ADP)

Background

This report concerns the [Agile development of the SSH Open Marketplace User Workshop](#) that took place online on 30 June 2020. The workshop was part of the [ICTeSSH conference](#) (Information-communication technologies enhanced Social Sciences and Humanities), an annual international conference that explores the changing research ecosystem in SSH fields in the digital age.

Workshop Overview & Format

The aim of the workshop was to raise awareness of how the SSH Open Marketplace can help researchers in their daily activities and supplement the existing EOSC services, to collect feedback from the SSH research community on the functionality and content of the SSH Open Marketplace, and to begin recruitment of additional testers to work with the development team.

The workshop was organised by WP 2, 3, 6, and 7. It was moderated by Marieke Willems (Trust-IT) and featured Ron Dekker (CESSDA), Daan Broeder (CLARIN), Laure Barbot (DARIAH), Matej Durco (OEAW-ACDH), and four testers of the Open Marketplace: Agnieszka Szulińska (Institute of Polish Language Studies of the Polish Academy of Sciences), Vanessa Hanneschläger (Austrian Centre for Digital Humanities and Cultural Heritage), Maurizio Toscano (University of Granada), and Kasia Karpinska (ODISSEI).

246 participants took part in the workshop. According to the Mentimeter survey that was conducted during the workshop, most participants were from Europe (with the Netherlands, Italy, Germany and Slovenia being best represented), though some even joined from Turkey, South Africa and the United States. Of the 54 participants that responded to the question on stakeholder categories, 17 belonged to universities and research institutes, 11 to research libraries and archives, 9 to research and e-Infrastructures, EOSC thematic clusters, 6 were researchers, and 3 belonged to the private sector and industry. Research funders, policy makers, and civil society and citizen scientists each only had a single representative.

The two-hour workshop began with three short presentations on EOSC, research services in SSH, and the SSH Open Marketplace. The round table with the Marketplace testers in the second part of the workshop focused on the user-view of the SSH Open Marketplace. Workshop closed with a Mentimeter survey and a call to join the community of testers.

Presentations & Discussions: Key Points

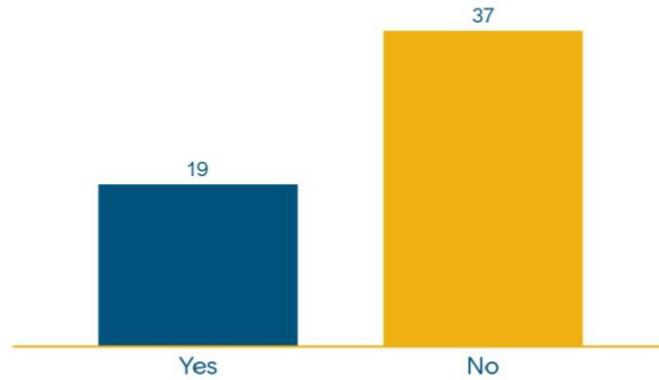
The content of the workshop was summarized in the official workshop proceedings.

Links to materials	Slides: https://doi.org/10.5281/zenodo.3928148
	Video recording: https://youtu.be/AZTrzHtl3Q
	Proceedings: Barbot, Laure; Biller, Tracey; Broeder, Daan; Dekker, Ron; Durco, Matej; Vipavc, Irena; Willems, Marieke. (2020). Agile development of the SSH Open Marketplace: User Workshop. ITM Web of Conferences. 33. 04001. https://doi.org/10.1051/itmconf/20203304001

Outcomes & Feedback

The workshop was informative and effective. The organizers were able to gain many useful observations and suggestions from the testers, particularly on UX design and navigation, content and curation, workflows and training materials, and the information users need about the Marketplace. In turn, the participants suggested how the Marketplace might support their work, which services and tools they would like to use, and how to motivate users to contribute to the curation of the Marketplace resources. The setup of the workshop was highly marked by the participants, receiving 4.3 on a scale of 1 to 5. The workshop was recorded and an article detailing the received feedback was published after the conference (see Links to materials).

Were you involved in previous SSHOC activities (workshop, webinar, etc.)?



56



How did you like the setup of the round table: 4 testers x 3 topics?



12



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



getting new resources

increase visibility

Serendipity

Get recommendations from research communities for tools, datasets, etc.

visibility of tools and services

with a proactive approach

New information on research

show results of our project

support researchers

23


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



Ease research

alignment with TRIPLE on the way to EOSC integration

Get rid of all other similar platforms and let me interact with only one!

providing a personalised experience

I work on the Triple project and I believe there will be close links to your SSH open marketplace

new research ideas

Updated knowledge

Ease search of tools

understand what is already done in the field

23


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



reduce time for searching tools/technologies

Data visualisation tools

open source tools

discipline-specific and general research tools/services

Data visualisation tools

23


What kind of services and tools would you like to see being incorporated in the SSH Marketplace?



14


Do you have ideas for incentives to motivate users to contribute to the curation of the marketplace' resources?



Citation on resources - doi?

highlight the possibility to promote own work

Credit and public acknowledge

Being cited

create user communities that cooperate on curation and on research

High quality service will create an ecosystem if it wants to be born

recognition, promotion of contributor

create a sense of community

reputation - certificate - prestige

17


Do you have ideas for incentives to motivate users to contribute to the curation of the marketplace' resources?



subscriptions to DH journals

option to share my curation efforts with others

Parallel discussion groups - develop communication on the areas

A dinner with the developer community

badges

A badge for 100 contribution done

Placement in official resources might emphasise this - links to scholarly societies?

Links to humanities societies who can provide links

17


What in your view could contribute to establishing trust in resources such as the Marketplace?



Open governance

clear information on governance

promotion by institutions (e.g. universities) and individual teaching staff and researchers

fff

test reports on tools by trusted users

finding what I am looking for easily

Established networks contributions

Transparency (on process) star system

create communities

13


What in your view could contribute to establishing trust in resources such as the Marketplace?



FAIR services, adherence to best practices and eg TRUST principles

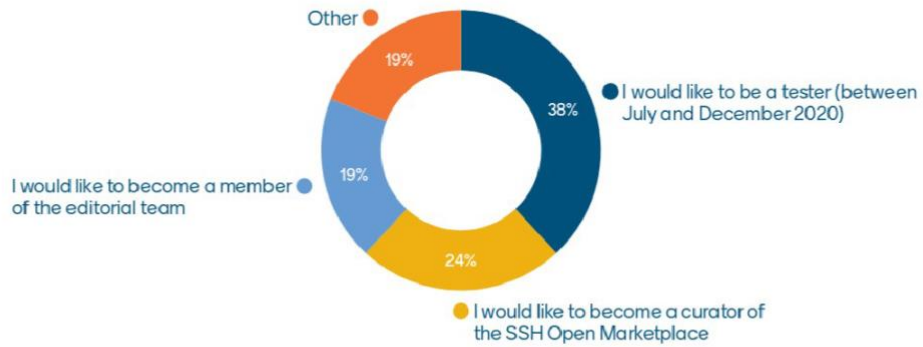
Access to users profiles

A difficult one - being open and transparent and not promoting 'advertised' products

enhance visibility: webinars, talks, social media

13


How would you expect to be involved in the community we aim at gathering on the marketplace?



ANNEX 2

SSHOC CONSIDERATIONS FOR VOCABULARY PLATFORMS

WORKSHOP REPORT

By *Lulianna van der Lek (CLARIN ERIC)*

Background

On November 6, [CLARIN](#) organised a virtual workshop to raise awareness with regard to vocabulary (platform) use in the research community and to discuss the *SSHOC Considerations for Vocabulary Platforms* as part of the SSHOC tasks 3.1 and 6.2.

The event wrapped up a [series of webinars](#) that CLARIN and SSHOC tasks 3.1 “Multilingual Terminology” and 6.2 started to organise in September as part of their initiative to collect, register, and harmonise SSH controlled vocabularies, thesauri and taxonomies, and to improve the service offer of vocabulary platforms and raise awareness about their use. This initiative emerged from the need to harmonise not only the vocabulary-related activities within the SSHOC work packages but also to unify the access to research contents and improve discovery in Social Sciences and Humanities.

Session Overview & Format

The virtual workshop, streamed via ZOOM, consisted of a total of eleven presentations (10-20 min long) and a discussion panel. Each session concluded with a lively Q&A part.

The [full programme overview](#) includes the links to presentations and recordings.

About the organisers

CLARIN set up the virtual workshop as part of their tasks in SSHOC (T3.1 and T6.2).

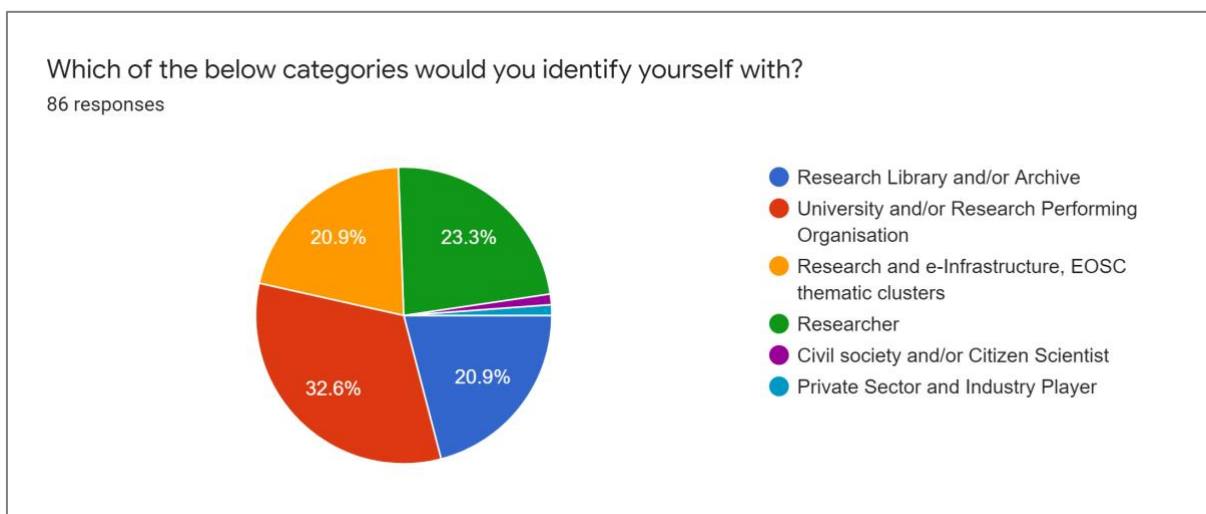
- Lulianna van der Lek, Project Coordinator of the Vocabulary Initiative at CLARIN ERIC
- Monica Monachini, CLARIN - IT National Coordinator and SSHOC T3.1.
- Dieter Van Uytvanck, Technical Director at CLARIN ERIC
- Daan Broeder, Project Manager at CLARIN ERIC

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

Participants

A total of 86 people signed up for the event, out of which about 60 participants attended. Most participants came from the Netherlands (17%), France (12.8%), Italy (11.6%) and Germany (9.3 %). It should be noted, however, that these numbers only account for the live viewers, and that the overall reach is higher since it also includes the views of the recordings.

About 32.6% of all the registered participants came from universities and research-performing entities, about 23.3% were researchers, and 20.9% worked for research and e-Infrastructure clusters or research libraries and archives. Only 1% of the participants represented the private sector.



Analysis of the applicants

Event summary

One of the main goals of SSHOC Task 3.1 *Multilingual Terminology*, as described in the DoA of WP3, is to find a suitable vocabulary server and publication platform to improve accessibility and discovery by non-native speakers. Therefore, the virtual workshop started with a presentation of the MS08 Report results: *Choice of vocabulary publication platforms for SSHOC*. The presentation was followed by an overview of the SSHOC WPs updates on vocabularies, after which CESSDA, CLARIN and DARIAH shared their experience with controlled vocabularies (CVs) and vocabulary platforms. The virtual workshop closed with a panel discussion where several experts from SSHOC and [TRIPLE](#) evaluated whether the CVs and vocabulary hosting platforms could be made more interoperable by following the [FAIR](#) principles.

Presentations & discussions: Key points

The following section outlines the main points of each presentation and the panel discussion.

Franciska de Jong, executive director at CLARIN ERIC, opened the workshop explaining the background of the vocabulary initiative. This initiative has emerged from the need to align the vocabulary activities across the SSHOC work packages, and it will help optimise the sharing of research data across various practices and domains.

De Jong further presented the work plan of the vocabulary initiative. In the early stages of the SSHOC project, WP3 collected the SSHOC requirements for the vocabulary registry in the milestone report MS8.0 *SSHOC considerations for vocabulary platforms*. In September, CLARIN launched a series of virtual events to raise awareness in the Social Sciences and Humanities (SSH) community about vocabulary hosting and publication platforms. Experts from SSHOC and other related Horizon-2020 projects, as well as end users, presented use cases and exchanged best practices. Future activities will consist of a re-evaluation of the first milestone, an inventory of the most known controlled vocabularies/taxonomies, and potential matching of vocabularies and their metadata.

Link to the recording:

Introduction ([recording](#))

Franciska de Jong, [Executive Director CLARIN ERIC](#)

WP3.1 MS8 report: Choice of vocabulary publication platforms for SSHOC

Monica Monachini, SSHOC WP3.1 leader and the CLARIN-IT National Coordinator, presented the results of the MS8.0 report: *Recommendations for Vocabulary platforms in SSHOC*. The main aim of WP3 is to lift technologies and services in the SSH Cloud by contributing to infrastructure components and content, and making the CLARIN technologies useful for the other SSH infrastructures. Within this context, the goal of WP3.1 is to provide resources and tools to improve discovery of the SSH research data and facilitate reusability. The team will translate the metadata in different languages and find a suitable vocabulary publication platform.

The milestone report MS8.0 provides a set of recommendations for vocabulary publication platforms for SSHOC. Through focused surveys, interviews with experts, the task team evaluated several platforms and produced a set of requirements that a vocabulary service should have, namely:

- Import and export of thesauri in SKOS / RDF format;
- Unified access to all vocabularies;

- Editing with collaborative functionalities;
- Alignment functions between vocabularies and external resource;
- Terminology management interface (hierarchical structure, semantic relationships translations, facets);
- Management of different roles and workflows;
- Management of versioning;
- API services;
- Friendly and intuitive interface, suitable for use by non-expert users;
- Flexible to be adapted to new needs and standards.

Three out of the eight platforms that have been evaluated seem to be the best candidates to host and publish the SSHOC vocabularies, namely ACDH-CH, CESSDA and CLARIN Vocabulary Services. However, none of them seem to fulfil all the requirements. Hence, more investigations are needed. The insights collected during the virtual sessions and the workshop will help refine the requirements and produce suitable recommendations.

The participants were interested to learn how WP3 envisions the collaboration between the SSHOC platform for CVs and the existing providers of vocabulary platforms.

Links to the presentation slides and recording:

Presentation of the MS08 results: *Choice of Vocabulary publication platforms for SSHOC (slides) (recording)*

Monica Monachini, [CLARIN - IT National Coordinator](#) and SSHOC WP3

SSHOC updates on vocabularies

In this one-hour session, five SSHOC WPs gave a short update about their work and experience with semantic artefacts, such as controlled vocabularies and taxonomies. Since some of the SSHOC WPs did not include specific controlled vocabulary tasks, the presenters shared their experience acquired in other Horizon 2020 projects, such as ARIADNE and DARIAH.

SSH Vocabulary survey

In a joint presentation, Clara Petitfils & Nicolas Larrousse (WP7-WP3) presented the results of the SSH vocabulary survey, and the resources needed for SSH Open Marketplace content description.

The survey was open from February until May 2020 and contained 16 questions about the use of SSH vocabularies, alignment, languages, availability, and maintenance. The results revealed the semantic artefacts that are often used by the community: the Data Documentation Initiative (DDI), Getty Art &

Architecture Thesaurus (AAT), CESSDA Controlled Vocabularies, ELLST, and Dublin Core. The respondents wished the vocabularies they used were matched with Getty and Wikidata.

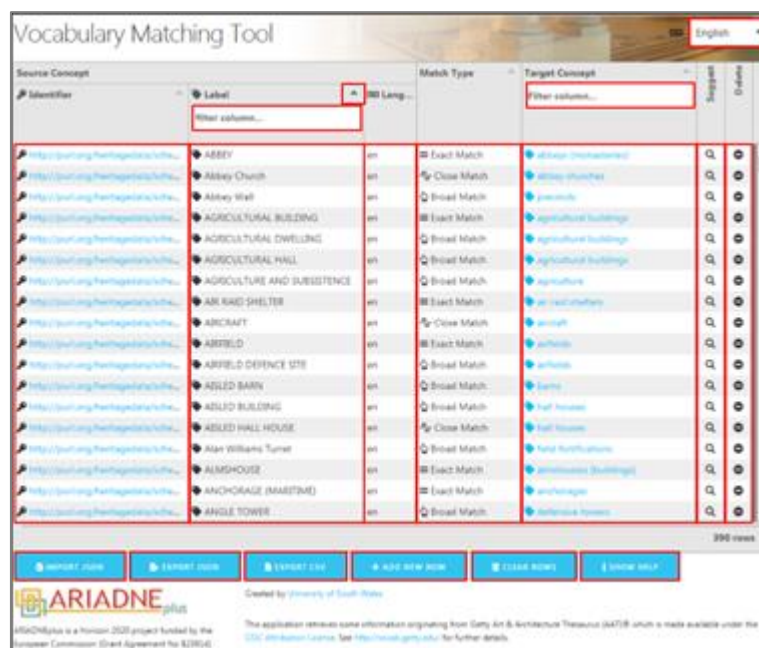
Nicolas Larousse pointed out that vocabularies are essential for the SSH Open Marketplace to describe the entries, improve search and retrieval, and foster discoverability. Furthermore, the external sources ingested into the marketplace will need to be aligned with the "local" vocabularies.

Vocabulary mapping tool for archaeology in ARIADNE plus

Holly Wright (WP5) from the [Archaeology Data Service](#) presented how they achieved vocabulary mapping in [ARIADNE](#), another Horizon-2020 project. With 1.9 million data records aggregated, the research project produced a total of 6 416 vocabulary mappings when aligning subject vocabularies to [Getty AAT](#).

The vocabulary mappings in ARIADNE cut out the linguistic barriers to cross-searching, such as language, spelling, homonyms, synonyms and level of specificity. Wright reiterated that it is important that the source datasets are produced with aggregation, cross-search and reuse in mind. She concluded that although there is a strong need to harmonise meaning, mapping everything to everything is nearly impossible.

To be able to match local subject terms and concepts to Getty AAT concepts, the research team developed a [vocabulary matching tool](#) in the first phase of the ARIADNE project.



Interface of the Vocabulary Matching Tool. Source: <https://vmt.ariadne.d4science.org/vmt/vmt-help.html>

The tool supports the SKOS mapping properties, and it has a multilingual user interface. Researchers can use it to search and browse through AAT vocabularies to make more informed decisions about the mappings. Users can export their data to JSON or delimited text (CSV) format that they can import in other applications. If users encounter terms/ concepts without URIs, they can add them manually to facilitate Linked Open Data source vocabularies.

Management of controlled vocabularies within the Aioli platform with the Opentheso tool

Isabelle Cao from CNRS-MAP presented how they create and manage controlled vocabularies in SSHOC WP4.6 within the [Aioli platform](#) using the [Opentheso](#) tool.

Aioli is a reality-based 3D annotation platform for the collaborative documentation of cultural heritage. It supports SKOS/RDF, and it is useful to reconstruct images automatically in 3D and annotate them by adding both custom and computer descriptors. The platform integrates several controlled vocabularies, which helps users consistently describe annotations.

The platform uses Opentheso, an ISO-compliant web-based thesaurus management tool, to create and manage controlled vocabularies. Data can be imported and exported in standard formats. The concept management interface is intuitive and supports hierarchical relationships, synonyms and translation. The thesaurus navigation allows for different views: tree, hierarchical, collection and index. Users can align concepts with other thesauri using semi-automatic alignment and pre-configured sources, such as Wikidata, Geonames etc. It is possible to add candidate concepts collaboratively, add notes and translate them. The system automatically adds the validated terms to the thesaurus. Finally, yet importantly, the platform offers features for user management and collaboration.

Opentheso has been implemented in the Aioli platform using a docker container. Users can manage thesauri directly in the Aioli platform using the Lexicon function. The published thesauri automatically synchronise with Opentheso. Future developments will include a search engine that will allow users to search for annotations within the project itself. Finally, there are plans to make the UI and the documentation available in English as well.

To learn more about the technical specifications of the Aioli platform, see deliverable [D4.16 Specification of the new feature of the Aioli platform](#).

BackBone Thesaurus, a model for thesauri interoperability

Because SSHOC T4.7 does not include tasks directly related to controlled vocabularies but ontologies, Chrissy Bekiari and Eleni Tsoulouha (WP4.7) shared their experience with the [BackBone Thesaurus \(BBT\)](#), a platform developed under Horizon-2020 project, DARIAH EU.

According to the [BBT Maintenance WG](#), the platform aims "to establish a coherent overarching thesaurus for the humanities, under which specialist thesauri and structured vocabularies used across scholarly communities can be aligned and form a thesaurus federation." The aligned vocabularies keep their autonomy, and the users can access them through global thesaurus. BBT follows the CIDOC-CRM classification, and users can access the federated thesauri through the [ACDH vocabulary repository service](#).

The platform supports high-level concepts, facets and hierarchies that do not exhaust the domain they classify. The BBT platform integrates [BBTalk](#), a multilingual online service for thesauri management and maintenance. The service developed by [FORTH-ICS](#) supports RDF, and it includes a thesaurus alignment tool, collaborative features and keeps track of versioning.

How Dataverse supports external vocabularies

Slava Tykhonov (WP5) from the [Data Archiving and Networked Services \(DANS\)](#) in the Netherlands presented how [Dataverse](#) supports controlled vocabularies with focus on semantic and technical interoperability, as defined by [EOSC Interoperability Framework v1.0](#).

Dataverse is a product developed by Harvard University and used in SSHOC T5.2. A recent self-assessment analysis performed by Merce Crosas¹⁰ of the implementation of the FAIR principles in Dataverse has revealed that while scoring high on Findable, Accessible and Reusable, the product is weak from an interoperability point of view. To tackle this, Dataverse aims to implement a FAIR metadata schema and connect the metadata to ontologies and controlled vocabularies.

Using [GRID \(Global Research Identifier Database\)](#) in SKOS, the platform provides a convenient depositor web interface to link the metadata of the datasets stored in the Dataverse network to external controlled vocabularies. The development team also plans to investigate if it is possible to disambiguate concepts using NLP tools and create the links between Dataverse and external vocabularies automatically.

The platform uses the [Skosmos API specification protocol](#) to ensure technical interoperability with other controlled vocabulary services (CESSDA). It connects to a Semantic Gateway application that enables users to query the vocabularies stored on different platforms (e.g. Skosmos). Users can link every field in Dataverse to the appropriate vocabulary following the FAIR principles. Finally, the platform supports multilinguality, and it allows researchers not only to enrich metadata but also export it to [Linked Open Data Cloud](#) to increase its findability or use it to train other Machine Learning models.

Links to the presentation slides and recordings of the SSHOC WPs updates on vocabularies session:

¹⁰ Merce Crosas, 2020, "FAIR principles and implementation in Dataverse": [FAIR-Dataverse-Tromso \(harvard.edu\)](#)

WP updates on vocabularies: ([slides](#))

- Vocabulary Survey ([recording](#)) Clara Petitfils & Nicolas Larrousse (WP7-WP3)
- Vocabulary mapping tool for archaeology in ARIADNE plus ([recording](#)) Holly Wright (WP5)
- Management of controlled vocabularies within the Aioli platform with the Opentheso tool ([recording](#)) [Isabelle Cao](#) (WP4)
- BackBone Thesaurus, a model for thesauri interoperability ([recording](#)) Chryssoula Bekiari (WP4)
- How Dataverse supports external vocabularies ([recording](#)) Slava Tykhonov (WP5)

Overview of the CLARIN & SSHOC webinar series

Julianna van der Lek, the CLARIN coordinator of the vocabulary initiative, gave an overview of the three vocabulary information sessions that took place in September: Introduction to Wikibase, CESSDA Vocabulary Service and Skosmos. A separate CLARIN-SSHOC report of these events summarises the sessions.¹¹

CESSDA: main requirements and best practices

Taina Jääskeläinen from the [Finnish Social Science Data Archive](#) gave an overview of the vocabularies available in the [CESSDA Vocabulary Service](#), the platform requirements and the tools they currently use.

The following controlled vocabularies are available within the service: subject-specific vocabularies, such as Thesaurus ELST and the CESSDA Topic Classification, vocabularies of research methods created by DDI Alliance, and vocabularies used for the CESSDA Data Catalogue.

The vocabulary service platform supports advanced user management, vocabulary management, translation and publication, concept hierarchy and synonyms. However, certain improvements still need to be implemented to ensure full functionality of the platform, among other things including concepts specific to a research domain within a broader controlled vocabulary which would support CESSDA organisations producing metadata from varied disciplines.

Links to the slides and the recording:

¹¹ WORKSHOP NOTES: SSHOC Requirements - Vocabularies and Vocabulary Management Platforms, <https://sshopencloud.eu/news/workshop-notes-sshoc-requirements-vocabularies-and-vocabulary-management-platforms> (accessed 9 December 2021).

CESSDA: main requirements and best practices
(slides) (recording)

Taina Jääskeläinen, [Finnish Social Science Data Archive](#)

CLARIN: main requirements and best practices

CLARIN has been managing vocabularies with [CLAVAS vocabulary server](#) since 2017. Currently, they are looking for a follow-up vocabulary management and publication platform. The requirements for the new platform are the following: it should be a sustainable open-source solution, SKOS compliant, with a GUI and a vocabulary editor, advanced browsing functionalities and fast look-up via API, and it should support persistent identifiers and multilinguality. Preferably, the solution should have preloaded and curated controlled vocabularies, such as ISO-693-3.

Different vocabulary platforms could be aligned through a compatible API, multidirectional synchronisation and double curation work.

The interoperability challenges between vocabularies could be tackled by reusing existing curated vocabularies and linking to authority files wherever possible. This approach would avoid reinventing the wheel over and over again.

Links to the presentation slides and the recording:

CLARIN: main requirements and best practices (slides)
(recording)

Dieter Van Uytvanck, [Technical Director at CLARIN ERIC](#)

DARIAH: main requirements and best practices

Matej Ďurčo (ACDH-CH) and Laure Barbot (SSHOC) shared the requirements and best practices for vocabularies and vocabulary platforms that they learnt from their DARIAH activities and how these could benefit the SSH Open Marketplace. DARIAH set up the WG Thesaurus Maintenance (BackBone Thesaurus), contributed to the vocabulary activities in [PARTHENOS](#), and have their central [vocabularies service](#).

According to M. Ďurčo, there should be a clear distinction between requirements for vocabularies and requirements for vocabulary hosting platforms.

Vocabularies should be published as Linked Open Data (based on SKOS data model) and provide comprehensive coverage of the domain through concept definitions and examples. Users should be able to reuse existing vocabularies or link them to other artefacts, thus ensuring semantic interoperability. Finally, vocabularies are useless if they are not available in authoring environments.

The vocabulary platforms, on the other hand, should support at least full SKOS data model, implement a curation workflow with full provenance and an API for easy access and look-up.

Matej Ďurčo shared best practices based on [TaDiRAH](#) (Taxonomy of Digital Research Activities in the Humanities) and the SSHOC Marketplace. The TaDiRAH vocabularies are available in various applications such as [SSK](#), SSHOC Marketplace and [DH Course Registry](#). The presenter revealed that it was challenging to integrate those vocabularies which did not have a URI and were not available via API.

In SSHOC, DARIAH has the task to develop the SSH Marketplace in WP7 that aims to be a discovery platform for tools, services and resources useful for research activities in Social Sciences and Humanities. Matej reiterated the controlled vocabularies are needed to describe and classify the items available via the platform and facilitate retrieval, browsing and interoperability. He mentioned that SSHOC D7.1 provides the first list of controlled vocabularies that could be part of the system specification of the marketplace, e.g. IANA mime, TaDiRAH, CESSDA Topic Classification, ISO 639-1, etc.

Custom-based properties in the data model and specific workflows for vocabulary creation and curation are under development. [PoolParty](#), a commercial taxonomy and vocabulary management server, is currently used for the hosting, management, and mapping of vocabularies. It seems that candidate concepts pose some challenges because of the communication between the ingest script, marketplace core and the Vocabulary Manager.

The presentation concluded with a snapshot of the vocabulary management workflow in ACDH-CH vocabulary service that uses Skomos to store and publish the vocabularies. Another challenge that we need to address is the following: What do we do when users want to manage the vocabularies directly within their applications, while they are stored and edited in third-party client applications?

Finally, yet importantly, Matej highlighted that it is essential to share the knowledge about vocabularies via dedicated training material, e.g. the [Controlled Vocabularies and SKOS](#) e-learning course available on DARIAH-CAMPUS.

Links to the presentation slides and the recording:

DARIAH: main requirements and best practices
(slides) (recording)

Matej Ďurčo, [ACDH-CH](#)

Panel: Improving the FAIRness of SSH Vocabularies

The theme of the panel was: "How can we make the SSH vocabularies FAIRer." The theme is in line with the SSHOC goal of helping SSH researchers integrate their work and results according to the [FAIR guiding principles](#) of scientific data management and stewardship.

[Daan Broeder](#), project manager at CLARIN ERIC, moderated the panel of discussion. The panellists consisted of the following experts: [Menzo Windhouwer](#) (KNAW), [Suzanne Dumouchel](#) (EOSC), [Matej Ďurčo](#) (ACDH-CH), [Melanie Bunel](#) (HumaNum).

The discussion revolved around the fact whether the FAIR principles could be applied to semantic artefacts as well. The moderator asked the panellists to reflect on the following topics:

- Would selecting and sharing a single vocabulary platform make the vocabularies FAIRer?
- What measures can further improve FAIRness of the SSH vocabularies?
- How do the editorial/curation processes influence vocabulary FAIRness?

Would selecting and sharing a single vocabulary platform make the vocabularies FAIRer?

Sharing a vocabulary platform can mean either sharing a single instance or sharing the code and creating federated instances.

While using a single platform could help overcome some of the current challenges that vocabulary platforms have, such as the lack of a uniform API to access vocabularies, the panellists agreed that it is something that it is not feasible to achieve. According to Matej Ďurčo, it would be challenging to unify the community and store all the vocabularies on one platform because there will always be different players, stakeholders, researchers who will prefer to be in control of their own platforms and instances.

Suzanne Dumouchel agreed with Matej Ďurčo stating that there should be at least a common place where all the existing SSH vocabularies are gathered and made more visible. The moderator suggested that we could share at least the same publication platform, but highlighted that the curation processes are quite different. He gave Skomos as an example.

Menzo Windhouwer proposed that if the SSH communities are not able to share the same vocabulary platform, we could at least try to reach a consensus on a small common API to be supported by the endpoints, for example, autocomplete to drill down into a vocabulary and getting information on a specific vocabulary item. Such a shared minimal API could help tackle interoperability challenges.

What measures can further improve FAIRness of the SSH vocabularies?

Findability

The first FAIR principle is to make data findable. The moderator encouraged the panellists to reflect on the findability aspects of vocabularies. Besides assigning a globally unique and persistent identifier, it is also important to describe the vocabularies in a coherent way. It does not mean that they need to be published in one instance because there are always ways to harvest the metadata and create a shared catalogue of all the available vocabularies to discover them.

Matej Ďurčo pointed out that researchers prefer to develop their vocabularies, but without following the current standards and guidelines for vocabulary development. The vocabulary service and the editing interface should guide the researchers in their work and support them to develop vocabularies in a consistent and standardised way. This will foster findability and reusability.

Accessibility

The second FAIR principle concerns data accessibility. This can be achieved via authentication and authorisation procedures.

Suzanne Dumouchel suggested that one way to make the vocabularies FAIRer could be by replacing the A for *Accessible* by A for *Adaptable* because we need to keep up with the changes and the evolving vocabularies. She also underlined the need for continuous cooperation with the research community in order to include new concepts., as well as to identify and define the SSH disciplines properly.

Interoperability

The third FAIR principle states that data need to interoperate with other applications or workflows for analysis, storage, and processing¹².

In the context of vocabularies, the moderator pointed out that there are tools to make the content of the vocabularies interoperable and support the mapping process between the vocabulary terms/concepts. It may be worthwhile to take a look at these tools when dealing with large vocabularies in the SSH community.

According to Suzanne Doumuchel, the SSH community could benefit from a project dedicated to SSH vocabularies and that CLARIN could play a role in it. Such a project would help all stakeholders reach decisions and agreements on SSH vocabularies (generic vocabularies and domain-specific vocabularies) and make them more interoperable (for example, using more Wikidata).

Matej Ďurčo reiterated that it is essential to distinguish between semantic and technical interoperability. Semantic interoperability could be achieved through an approach that counts for plurality, for example, letting researchers create their vocabularies and then use tools to link to match them.

Since it is time-consuming to map entire vocabularies, the moderator argued in favour of a pragmatic approach; for example, mapping only those parts of vocabularies that are relevant. He referred the audience to the [SEMAF EOOSC project](#) that proposes a flexible semantic mapping framework targeted at specific interoperability goals.

Menzo Windhouwer pointed out that the end users do not know what types of relationships knowledge engineers apply during vocabulary mappings. He proposed to set up a quality assessment of the mapping process that includes some provenance metadata with confidence metrics.

One member of the audience, [Andrea Scharnhorst](#), remarked that discussions around semantic interoperability have been going on for a long time and that there are similar projects taking place in parallel. Hence, we should avoid reinventing the wheel and encourage the SSHOC researchers to build on existing initiatives as much as possible. She gave as examples, the [Linked Open Vocabularies \(LOVs\)](#), the [BARTOC Vocabularies](#), and the [KOS Observatory for Social Sciences and Humanities](#), a project

¹² FAIR Principles, <https://www.go-fair.org/fair-principles/>

developed by DANS. The latter includes about 125 KOSs from the Social Sciences and Humanities (SSH), and the Life Sciences (LS) that researchers have mapped by applying a thorough methodology.

Sustainability

Menzo Windhouwer asked how the vocabularies could be maintained over time, for example, after the funding received for a research project had ended. He pointed out that it would be laborious to keep the resources up to date by using volunteers because they need specific knowledge engineering skills.

Franciska de Jong confirmed that sustainability is indeed an issue. She indicated that this topic is a recurrent theme at several layers of SSHOC as a project. Some parties may have the possibility to sustain the tools and linguistic resources that they are developing. For example, CNR has committed to hosting the ARIADNE mapping tool for at least five years after the SSHOC phase has finished. There are plans to make all the mapped vocabs available as RDF downloads for reuse as well. BBT and BBTalk will continue to be maintained voluntarily by the Thesaurus Maintenance Working Group.

Franciska de Jong hopes that this vocabulary initiative will motivate the SSHOC WP leaders to work towards interoperability. SSHOC could develop a couple of use cases to demonstrate how WPs have achieved interoperability across platforms.

How do the editorial/curation processes influence vocabulary FAIRness?

Menzo Windhouwer pointed out that the end users do not know what types of relationships knowledge engineers apply during vocabulary mappings, so he proposed to set up a quality assessment of the mapping process that includes some provenance metadata with confidence metrics.

The panel concluded that more discussions are needed to come to a final recommendation for vocabulary platform(s) in SSHOC. Most likely, there will not be a single shared platform to manage the vocabularies. The moderator encouraged the SSHOC WPs to think about how they can achieve interoperability between the different platforms, for example, data exchange, vocabulary identification schemes, vocabulary maintenance and quality management.

Link to the recording:

Panel: SSHOC Considerations for the Vocabulary platforms ([recording](#))

Moderator: [Daan Broeder](#)

Panelists: [Menzo Windhouwer](#), [Suzanne Dumouchel](#), [Matej Ďurčo](#), [Melanie Bunel](#)

Outcomes & feedback

While the vocabulary virtual sessions we organized in September were meant to be informative, the

workshop provided a more in-depth discussion about vocabularies and vocabulary platforms to be used in SSHOC. One recurrent theme was that the SSH community should avoid reinventing the wheel and build on existing initiatives as much as possible. There should also be a clear distinction made between the requirements needed for vocabularies and requirements needed for vocabulary platforms.

Semantic and technical interoperability seem to be the most important requirements. This could be achieved by reusing the existing vocabularies. Reuse will prevent practitioners from creating double concepts. Users should also be able to link them to other semantic artefacts or integrate them in authoring environments.

Vocabularies should be published as Linked Open Data (i.e. RDF, SKOS, OWL formats) and provide comprehensive coverage of the domain through structured concept definitions and examples. Since SSH is a very diverse domain, concept definitions may vary as well leading to linguistic barriers, such as ambiguity. Therefore, mappings between different vocabularies are needed. However, since it is time-consuming to map entire vocabularies, it was proposed to adopt a flexible semantic mapping framework (e.g. [SEMAF EOOSC project](#)) targeted at specific interoperability goals.

For the SSH Open Marketplace, vocabularies are essential to describe the entries, improve search and retrieval, and foster discoverability. Furthermore, the external sources ingested into the marketplace will need to be aligned with the "local" vocabularies. Finally, yet importantly, since vocabularies are changing once they are integrated into a platform, a solid governance model is needed to ensure that updates and maintenance are done systematically and automatically. This will ensure quality.

The milestone report MS8.0 provides a set of recommendations for vocabulary publication platforms for SSHOC. Three out of the eight platforms seem to be the best candidates to host and publish the SSHOC vocabularies, namely ACDH-CH, CESSDA and CLARIN Vocabulary Services. However, none of these platforms seem to fulfil all the requirements as identified in the MS8 report, *Recommendations for vocabulary platforms in SSHOC*.

The panel of the virtual workshop concluded that more discussions are needed to come to a final recommendation for vocabulary platform(s) in SSHOC. Maybe instead of looking for a single platform to host and publish vocabularies, efforts should be made to achieve interoperability between the different platforms at the following levels: data exchange, vocabulary identification schemes, vocabulary maintenance and quality management.

We will use the output of the vocabulary event series to reassess the initial requirements for vocabulary platforms in SSHOC. Last but not least, we will further develop our SSH vocabulary registry, find a suitable platform to store it and investigate how this work could feed into the SSHOC products.

Overall, the participants found the event very useful and well-organised. They appreciated the diversity and variety of presentations, the discussion panel and the discussion in the chat.

Here are some testimonials about the impact of the event:

- "It will impact my work on the usage of vocabulary to common people. I have learned that I should make the technical words easier to understand to the common people. The more they understand, the more they are likely to cooperate and support."
- "It is inspiring to think about Controlled Vocabulary Management."

Post-event survey

Timestamp	1. Overall, how would you rate the events?	2. How did you learn about this event?	3. Did the event meet your expectations?	4. Do you see this event having a positive impact on your work and how?	5. What did you like most about the event?	6. What did you miss or could be done better at the event?	7. Is there anything else you'd like to share about your experience (your suggestions/needs for expansions on the topic, etc.)?	8. How well organized was the event (in terms of time management, length, venue/webinar system)?
2020/11/06 12:35:22 pm CET	Excellent	SSHOC lists	Matched expectations	Not directly, but in the back, since we are providing language specific vocabularies for one of the platform presented	a lot of discussions in the chat, and really interested presentations	perhaps more use cases / I know they where in September, but would be nice here as well	/	Well organized
2020/11/06	Excellent	The organizers	Greatly exceeded the	It will impact my work on the usage of vocabularies to	The Vocabulary	I hope that the event	I wish that they have a segment on	Excellently organized

4:40:12 pm CET		were excellent at preparing this event.	expectations	common people.. I learned that I must make the technical words easier to understand to the common people... The more they understand, the more they are likely to cooperate and support :)	Mapping Tool for Archaeology in ARIADNE + :)	will be longer next time..	vocabulary usage during the development of abstracts :)	
2020/11/09 11:47:03 am CET	Good	good	Matched expectations	yes	range of systems presented	a summing up and panel discussion	There was much discussion about the need to not keep reinventing the wheel. This is the crux of the matter.	Well organized
2020/11/09 1:10:04 pm CET	Excellent	well organized	Matched expectations	inspiring to think about Controlled Vocabularies management	diversity and variety of content contributions - really a good kaleidoscope	nothing	no	Excellently organized
2020/11/20 2:35:08 pm CET	Very Good	very good	Matched expectations	Positive impact, I hear about experience of other people.	Everything was OK.	Everything was OK.	No.	Well organized

ANNEX 3

SSHOC SPEECH-TO-TEXT WORKSHOP - LINKING SOCIAL SURVEY AND LINGUISTIC INFRASTRUCTURES THROUGH SPEECH INTERVIEWS WORKSHOP REPORT

By Ana Inkret (CESSDA/UL-ADP)

Background

The workshop – [SSHOC Speech-to-text Workshop - Linking Social Survey and Linguistic Infrastructures through speech interviews](#) – was held online on 16 April 2021 as an awareness raising event. It was organised by tasks 4.4 “Voice recorded interviews and audio analysis” (NIDI) and 6.2 (CESSDA/UL-ADP) to present the approach of the T4.4 to integration of research infrastructures and to engage researchers and technical specialists interested in the research potential of audio records collected in social surveys.

Workshop Overview & Format

The aim of the workshop was to introduce the project of collecting and analysing audio recording in quantitative social survey, present the tools that are used for collection, processing and analysis, and discuss certain aspects of the work with participants.

The content of the workshop was planned by Judith Koops (NIDI), the speakers, Joris Mulder (LISS panel), Henk van den Heuvel (CLARIN ERIC), and the moderator Giovanni Borghesan (EVS). CESSDA/UL-ADP provided technical support.

After a short welcome, the workshop started with an introduction to the work of the project and continued with two presentations. Participants then joined one of the four breakout rooms to continue discussion in smaller groups. Workshop concluded with a short summary of the discussions.

53 participants took part in the workshop (out of the 57 that registered to attend), indicating very good turn-out and high interest in the topic. Majority of the registered participants belong to two stakeholder groups, University and Research Institutions and Researchers (23 each), followed by Research Libraries and Archives (5 participants). The other categories were only individually represented.

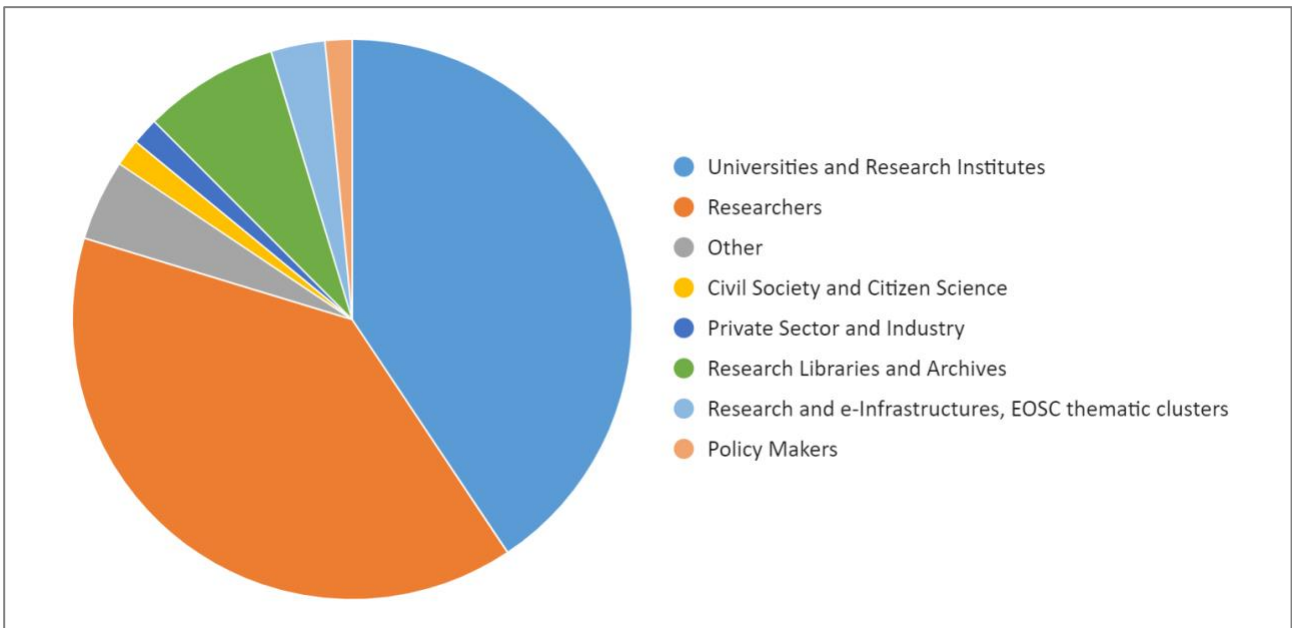


Figure 1 Participants by stakeholder categories

Participants came from 16 countries, the majority (88 %) from the EU countries.

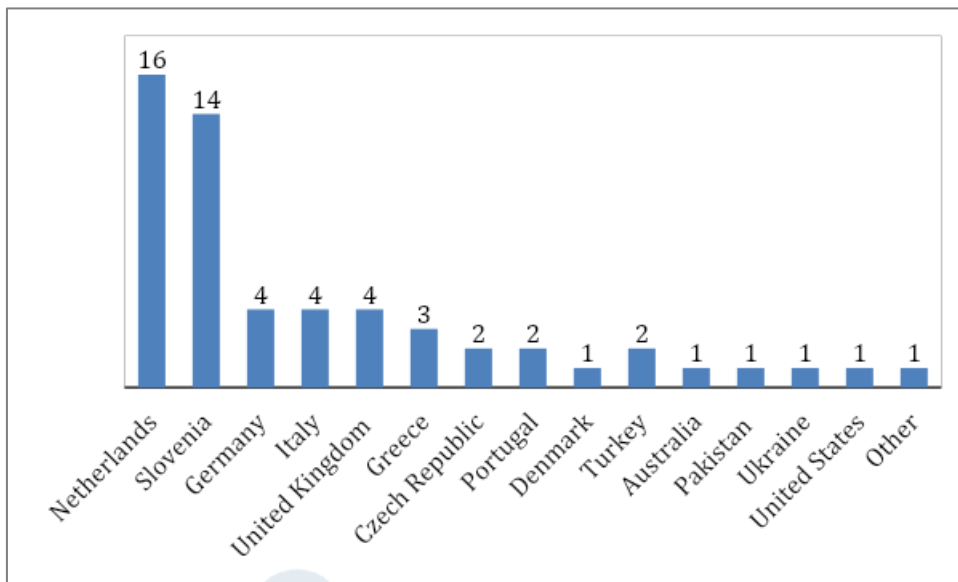


Figure 2 Participants by countries of origin

Presentations & Discussions: Key Points

Linking Social Survey and Linguistic Infrastructures

Speaker: Judith Koops, Generations and Gender Survey, Netherlands Interdisciplinary Demographic Institute

Main Points. In the innovative project that started earlier in April, audio records were collected by the social science infrastructures while the humanities infrastructure was used to automate the processing and analysis of audio data. This approach enriches the potential of data and provides a wealth of new information for both sides, such as the opportunity to connect linguistic information with socio-economic and demographic data or the ability to observe respondents' sentiments towards a topic.

Links to materials

Slides: <https://doi.org/10.5281/zenodo.4700104>

Video recording: <https://www.youtube.com/watch?v=8OmXr0Fr7DY>

Applying Speech to Text software in the Dutch LISS panel. Challenges and solutions of audio responses in a web survey

Speaker: Joris Mulder, CentERdata, Tilburg University

Main points. The audio recordings were collected in a social survey conducted in the LISS panel (Longitudinal Internet studies for the Social Sciences), a probability based online panel in the Netherlands. Using external software, speech-to-text technology allows the respondents to answer verbally rather than type the answers. Audio files are then manually checked and anonymised and are stored securely. Technical issues and pros and cons to the approach were presented, such as selective non-response. The collected data will be available in the LISS repository later this year.

Links to materials

Slides: <https://doi.org/10.5281/zenodo.4700104>

Video recording: <https://www.youtube.com/watch?v=8OmXr0Fr7DY>

Linguistic analysis of audio recorded social surveys by means of language and speech technology

Speaker: Henk van den Heuvel, CLST, Radboud University

Main points. The CLARIN infrastructure offers a range of tools for processing and analysis of the collected audio record, based on acoustical and text analysis. Audio records can be automatically transcribed using an open-source transcription chain. The project is testing and improving this feature for the Dutch language. Resulting text can be further analysed as qualitative data by e. g. applying tools for topical modelling, annotations, measuring statistical relations, producing word clouds, and even automatic summarization.

Links to materials

Slides: <https://doi.org/10.5281/zenodo.4700104>

Video recording: <https://www.youtube.com/watch?v=8OmXr0Fr7DY>

Plenary discussion

Main points. In the final part of the workshop, participants were able to lead a more focused discussion on the topics that were suggested by the three speakers and the moderator. After a short introduction of the starting points, participants formed smaller discussion groups in the four open breakout rooms, each moderated by one of the speakers. This format allowed participants to exchange knowledge and ideas and to examine certain concepts more closely. J. Mulder's group focused on the survey developer's perspective, J. Koops' group on the user perspective, H. van den Heuvel's group on the information extraction perspective and G. Borghesan's on the methodological and analytical insights. Discussion was not recorded, but the moderators provided a short summary in the main room before closing the workshop.

Outcomes & Feedback

Participants' response shows that the topic of the workshop was current and relevant. The workshop provided a valuable insight in the work-in-progress. With the restrictions that the social surveys face due to the health situation, audio recording seems a viable alternative to face-to-face interviews. Considering the additional information that can be gained by applying the linguistic analytical tools, audio recording will be an increasingly important source of information for social scientists and linguists alike. Participants joined the workshop to learn about the technical tools and of the more social implications and concerns. The post-event survey (though with a typically low response rate) confirms that the workshop provided a very useful overview of the project and of the practical insights on the tools for collection and analysis of data. There will undoubtedly be a lot of interest in the progress and results of the project.

The decision to turn to breakout rooms for a more intimate and efficient debate proved to be the right one. This format allowed the participants to tailor the workshop to fit their interests better, ask direct questions, network, and share their experiences with a smaller group. Notes that were taken during the breakout session will be useful in the future work of the project. A short summary of each discussion group was provided for the participants at the end. It might be useful to note for future events with similar agenda that a recording of the end summary would be practical and would not break the confidence established within the group.

After the workshop, video recording of the presentational part was published as well as the presentations themselves. As of the end of May 2021, the slides have been viewed 15 times and downloaded 12 times, while the video has had 25 views.

Post-event survey

Overall, how would you rate the workshop/ webinar?
5
4
4
5
Where did you learn about the workshop/webinar?
mailing list
Social Media
Colleagues
Social Media
How well organised was the workshop/webinar (in terms of time management, length, venue)?
5
5
5
5
What did you hope to gain from the workshop/webinar?
Inside on Tools for speech recognition
Speech-to-text techniques, tools, concerns
New views (technical, content) of conducting interviews
Knowledge
Did the workshop/webinar meet your expectations?
4
4
4
5
Do you see this workshop/webinar having a positive impact on your work and how?
Yes, very useful overview of a project and possible tools for speech recognition
Yes, I learned different approaches.
Yes, giving me more insight on potential (technically) how to conduct interviews.
Yes

What did you like most about the workshop/webinar?
Practical insides and examples
organisation, subjects, speakers
Breakout rooms
Overall execution
What did you miss or could be improved at the workshop/webinar?
A list of possible tools for different analyses. But a good starting point was given in form of a general link where to find some tools
-
More relaxed atmosphere, longer breakout rooms
Nothing

ANNEX 4

SSHOC ARCHAEOLOGICAL CASE STUDY WORKSHOP – THE ROMAN THEATRE IN CATANIA FROM SURVEY TO INTERACTIVE 4D VISUALISATION WORKSHOP REPORT

By Ana Inkret (CESSDA/UL-ADP) and Veronika Keck (CESSDA/GESIS)

Background

The workshop – [SSHOC archaeological case study Workshop - The Roman theatre in Catania from survey to interactive 4D visualisation](#) – was held online on 25 May 2021. It was organised by tasks 5.7 (DAI and CNR-ISPC) as an awareness-raising event, aiming to present the workflow developed for processing archaeological survey data and to exchange experiences in the broader context of archaeological data reuse and visualisation.

Workshop Overview & Format

The aim of the workshop was to present the progress on the case study of developing a workflow for the cloud-based processing of archaeological data and visualisation of archaeological sites. The workshop also intended to provide an overview of FAIR principles in archaeology, to present the SSHOC reference ontology and the Extended Matrix system, and to discuss the possible ways of integrating the new approaches in practice.

The content of the workshop was planned and delivered by the team of task 5.7: Wolfgang Schmidle (DAI), Alberto Bucciero, and Emanuel Demetrescu (both ISPC), as well as Athina Kritsotaki (FORTH), Holly Wright (ADS), and Bruno Fanini (ISPC). Technical support was provided by CESSDA/GESIS and CESSDA/UL-ADP, both from task 6.2.

The workshop had 33 participants. The majority (14) of participants were researchers, 9 belonged to universities and research institutions, and 4 participants could not associate themselves with any of the given categories. Research libraries and archives as well as research and e-infrastructures were each represented by two participants, while one participant belonged to the private sector (see the graphic below).

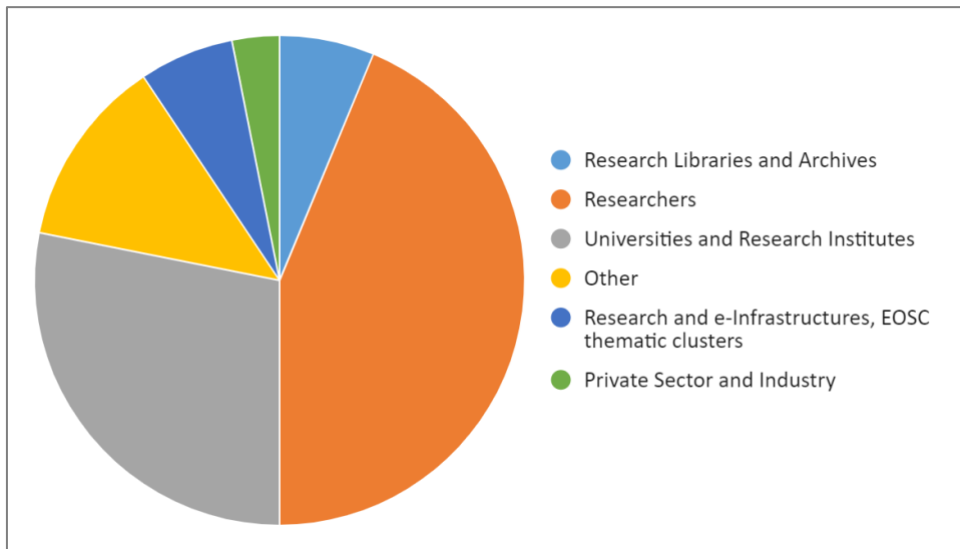


Figure 1 Participants by stakeholder categories

Participants originated from 11 countries, including Italy (10 participants), Slovenia (5), Greece (4), and even India (1).

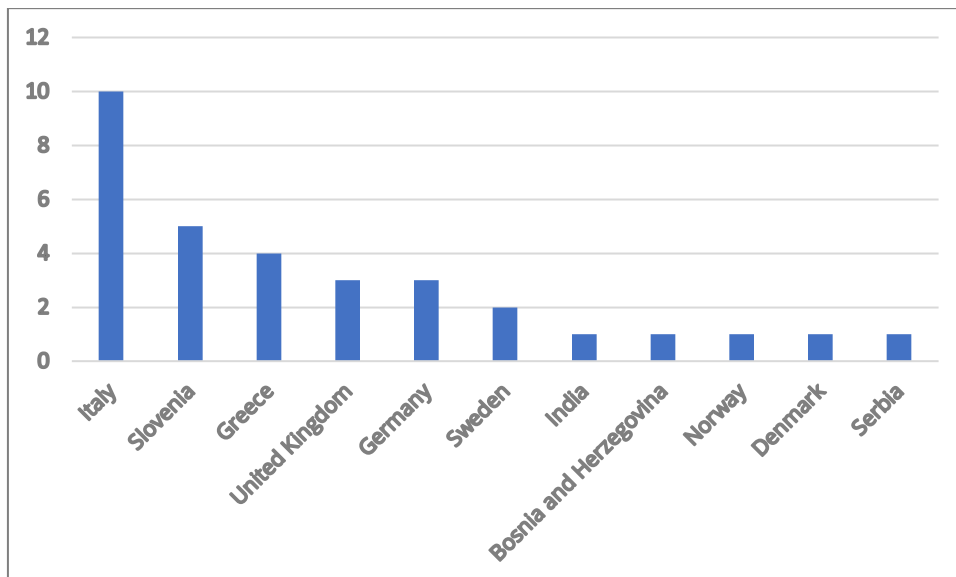


Figure 2 Participants by country of origin

A Mentimeter survey was conducted in the beginning of the workshop (see last chapter).

The workshop was divided into two main sessions. The first one started with a general introduction and continued with three presentations. After the break, the second session started with an introduction to breakout rooms. Participants then joined one of the three breakout rooms and reconvened after for a summary of breakout sessions, discussion of the next steps and a short Q&A session.

Presentations & Discussions: Key Points

Part 1: Work in progress

Introduction to the case study and the Roman theatre in Catania

Speakers: Wolfgang Schmidle (DAI), Alberto Bucciero (ISPC), Emanuel Demetrescu (ISPC)

Main points. Wolfgang Schmidle opened the workshop with a short introduction of the task that has been creating a workflow of processing archaeological survey or excavation data into a 3D or 4D visualisation. Alberto Bucciero then introduced the multidisciplinary team of the Lecce-based Institute of Heritage Science (ISPC) of the National Research Council that participated in the task, as well as the archaeological site of the Roman theatre in Catania, the best-preserved example of this type of construction with several building phases that had previously been well explored. Emanuel Demetrescu discussed previous archaeological projects in virtual reality that faced problems with software sustainability. This experience echoed in the Catania project, where metadata was available only as a publication and the existing reconstruction had to be annotated and linked to the sources of the reconstruction hypothesis.

Roman theatre in Catania: from survey to virtual reconstruction

Speaker: Francesco Giuri (ISPC)

Main points. Reconstruction of the Catania theatre required a complex workflow, consisting of four main steps: documentation (bibliography and archaeological excavation data), survey (laser scanner and digital photogrammetry), data processing (partial mesh reconstruction of the points cloud), and 3D reconstruction hypothesis. The metric characteristics were cross-checked with the published material. The hypothetical reconstruction was linked to the Antonine phase of the theatre, when the theatre was enlarged, and new *scena* and facade were built. Maxon CINEMA 4D software was used for the modelling, texturing and animation of the reconstruction.

Extended Matrix

Speaker: Emanuel Demetrescu (ISPC)

Main points. The importance of ingesting metadata for reconstruction during the project itself was noted in a previous project, conducted by ISPC (V-Must European project). Using metadata as a research tool in academic dissemination or as content validation for creative industries improves the way archaeologists share information. A tool that enables this is Extended Matrix, a formal language for the creation of virtual reconstructions for cultural heritage that keeps track of the reconstruction processes. Formally, it is an oriented graph that can be annotated with the metadata. It is an extension of an archaeological “Harris matrix”, a tool to describe the relative chronology of the elements discovered during the excavation. In EM, the process starts from the gap in material culture - a negative stratigraphic unit that needs to be reconstructed. All the elements documented in the regular matrix are filled with a specific horizon or epoch that is to be reconstructed. To annotate the reconstructive processes in the

case of Catania theatre, the EM process was reversed, starting from the reconstruction. Two tools were used, Blender (visualisation and editing software for 3D) and yED (a graph editor) plus two open source software that are part of the Extended Matrix Framework: EMtools¹³ and 3D survey Collection¹⁴ (3DSC). Since the EM is a very small file that can be easily shared, changes in the EM are immediately visible in the 3D environment even in a collaborative scenario of creation of a virtual reconstruction (a team distributed in different places and connected through cloud services). Creation of the representation model is an iterative, modular process that demands a team with understanding of stratigraphy (for the relative chronological relation of the elements), basic modelling skills (for the proxy models), and skills in communication (for the final representation model).

DAI systems and norm data

Speaker: Wolfgang Schmidle (DAI)

Main points. The iDAI.world is a system of open access web services for archaeological data. Data services include Arachne, a large database of archaeological objects, and idai.field, a system for documenting excavation data. They are complemented by norm data services for data interoperability and standardization, notably idai.gazetteer for place-related data and ChronOntology for temporal terms. The Roman theatre in Catania is represented in Arachne as a place and as an object/building with information about the building and linked objects. This creates a network of archaeological information. iDAI.world includes iDAI.field, a system for documentation of excavations and surveys that can be also used off-line for description, polygon drawing and creation of Harris matrix.

Integrating the systems and data

Speakers: Wolfgang Schmidle (DAI), Alberto Bucciero (ISPC), Emanuel Demetrescu (ISPC)

Main points. Process from the study of an archaeological site to determination of its reconstructive hypothesis and its 4D representation has many steps (from the remote study phase that is mostly biographical and produces references to publication, to field survey, then data processing that leads to the visualisation) and requires many tools. Each step produces data that is used in the next step. Focus on the result - publication - often means that the intermediate raw and processed data is lost and not shared. SSHOC project offered the opportunity to rethink the process. For the integration of the EM with the iDAI tools (Arachne), the possibility to connect to the different sources is important. The connections between buildings, places, objects, publications, norm data in Arachne is building towards Linked Open Data. As LOD works well in very standardised settings, more work is needed on the alignment of terms.

Part 2: The broader context

Breakout session 1: The SSHOC Reference Ontology (SSHOCro): Modeling the SSHOC data life cycle

¹³ <https://doi.org/10.5281/zenodo.4581456>

¹⁴ <https://doi.org/10.5281/zenodo.4556757>

Speaker: Athina Kritsotaki (FORTH)

Main points. SSHOCro is a workflow model that aims to describe the full data life cycle in SSH research. It is built on the ground of analytical methods used in various disciplines to inform a common research workflow. SSHOCro both uses and extends the CIDOC CRM (ISO21127), an event-based ontology. In terms of practical use, SSHOCro can be applied as a standard to devise and implement metadata capture schemes for tracking the data life cycle in individual projects, institutions, or disciplines; it can serve as a common language between SSH researchers and IT specialists; and, when encoded in a semantic data format (e.g. RDF), it can be used for mapping, transforming and integrating existing data across projects, institutions, and disciplines into interoperable pools (semantic repositories) of information for re-use and further exploitation.

Breakout session 2: Archaeology and the FAIR Principles

Speaker: Holly Wright (ADS)

Main points. Participants discussed long-term preservation and difficulties in preserving complex data with an example of a service with data from East Asia. Metadata and paradata were also discussed, as well as the problem of users' trusting the data and the findings of the SEADDA project, related to users' needs and preservation of complex data.

Breakout session 3: The EMviq tool

Speaker: Bruno Fanini (ISPC)

Main points. EMviq (Extended Matrix visual inspector and querier) is a cross-device, online interactive 4D visualisation tool for the Extended Matrices. It is based on ATON open-source framework because of its modern web-standard and open specifications, cross-device presentation (mobile devices, computers, displays), efficiency of queries, and scalable deployment. The EMviq tool has three main aspects in practice: exploration (users can navigate movement through both space and time in the multitemporal virtual environments), query (inspection of 3D shapes - proxies), and search (search or filter by proxy ID or by term). The tool works as a query interface to the matrix. In regard to data, the tool has a cloud-based workflow that allows for collaboration, with fine-grained access policies and live updating of the EM inside the tool.

Where do we go from here?

Speakers: Wolfgang Schmidle (DAI), Alberto Bucciero and Emanuel Demetrescu (ISPC)

Main points. In technical terms, the reconstruction process currently consists of heterogeneous steps with both online and desktop applications; these could be replaced by an online unique cloud service. Work will also be done on data integration and on the classification of time within the EM. As the EM is integrated into the iDAI.world system, more work on the alignment of vocabularies and granularity of data is needed.

Links to the materials:

Slides: <https://doi.org/10.5281/zenodo.5013565>

Video: <https://www.youtube.com/watch?v=Nz0Q3HORgf4>

Outcomes & Feedback

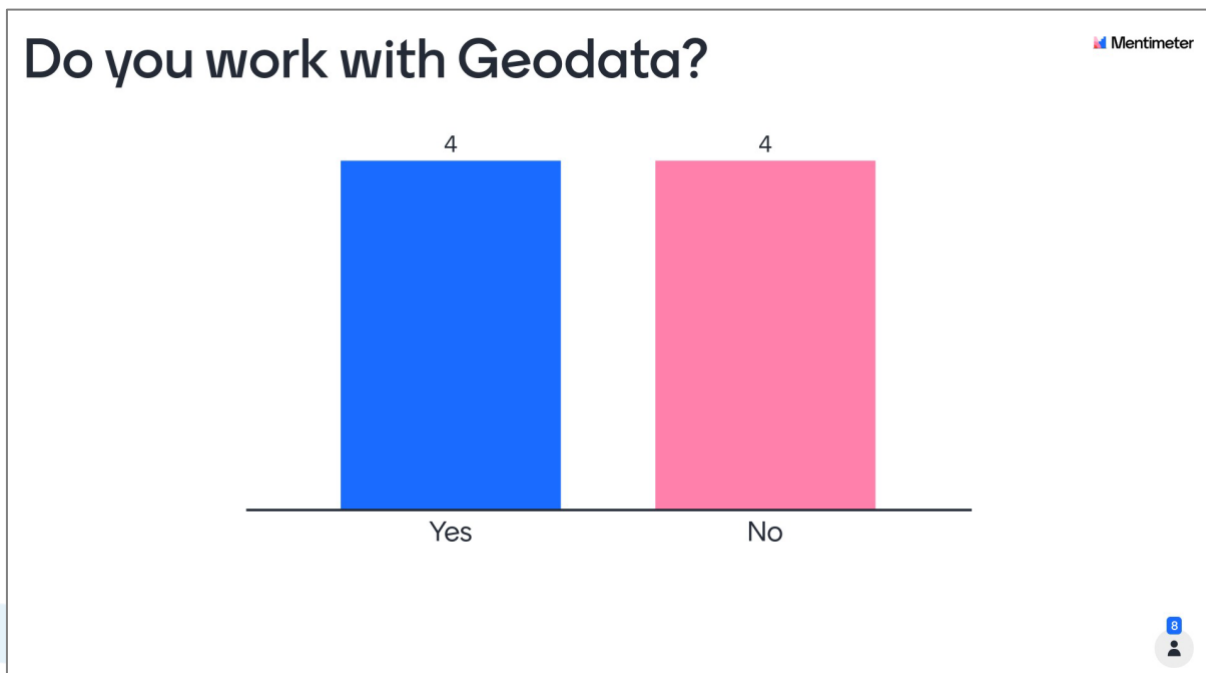
Only three participants responded to the SSHOC evaluation form (see Appendix 1). Overall, the event was positively evaluated. Two participants evaluated the event as “excellent”, one participant has chosen the option “very good”. The organization was evaluated as “excellently organized” by two participants, and “very well organized” by one.

Similarly, positive was the feedback and rating with regards to the question of whether the event met participants’ expectations. One participant reported that the event exceeded their expectations, and two agreed that the workshop greatly exceeded their expectations.

Participants expected to get “new experience and data”, “exchange interests” and “knowledge of the workflow”. Their expectations have been met. The positive feedback was given with regards to the content of the presentations. In particular, respondents reported that they found all of the presentations and discussions to be “interesting”. Responding to the question of what could have been better, participants indicated that nothing could have been improved.

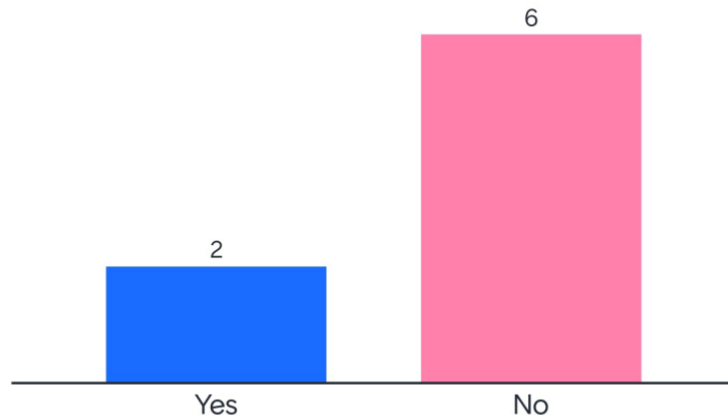
Announcements were shared on the SSHOC website, several mailing lists, partners’ websites, and SSHOC project social media. A blog post of the event has been prepared afterward which will be distributed through the SSHOC networks.

Mentimeter survey



Do you create 3D visualisations?

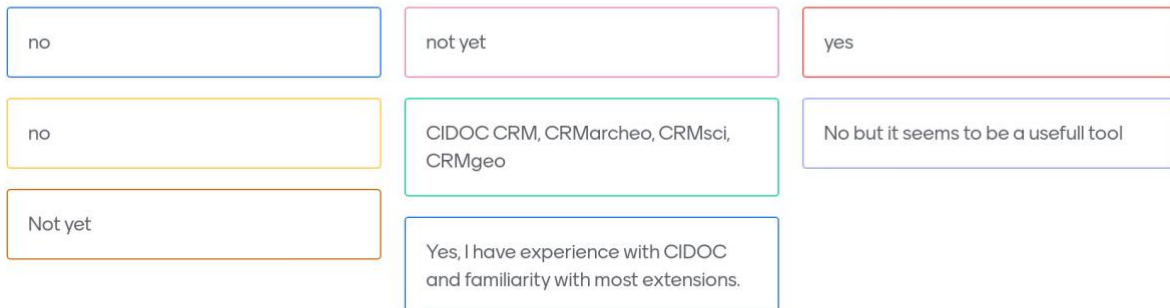
Mentimeter



8

Do you have experience with CIDOC CRM, or its extensions such as CRMaracheo or CRM geo?

Mentimeter



8

Do you have experience in implementing FAIR principles? If so, please describe.

Mentimeter

not yet

Yes, FAIR has become a core aspect of everything we do now.

no

Yes, in the field of social science

I'm starting just now

yes but more experiences are needed

conducting fair experiments

Our organisation has implemented them, and is continuing to improve

6

What are you especially interested in?

Mentimeter

both tools and theoretical aspects

3D virtual reconstruction

3D and metadata

More information about tools used for metadata

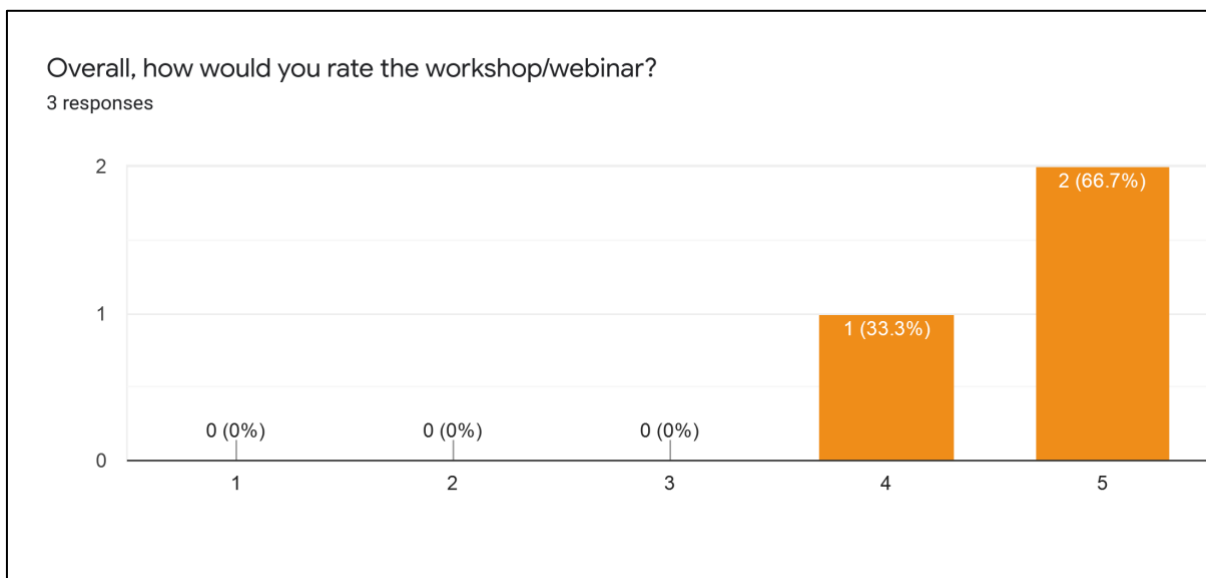
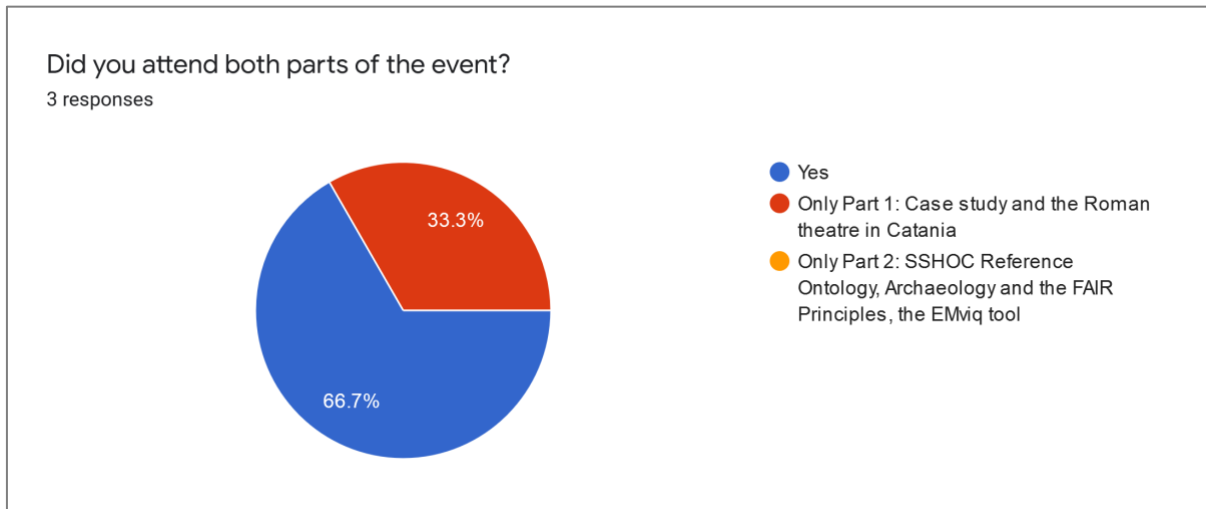
Really interested in the way metadata and paradata can be exported in this project and would like to learn more.

implementing EM into the current workflow and documentation

How metadata structure can be used for different fields

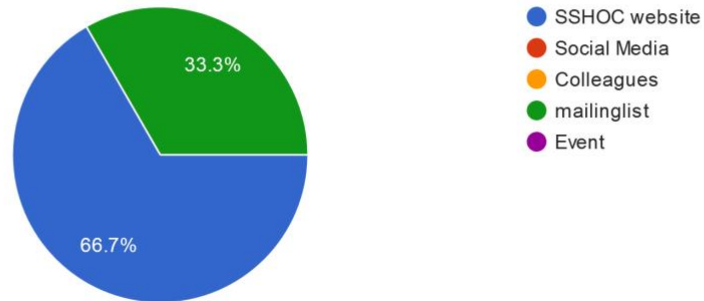
7

Results from the post-event evaluation survey



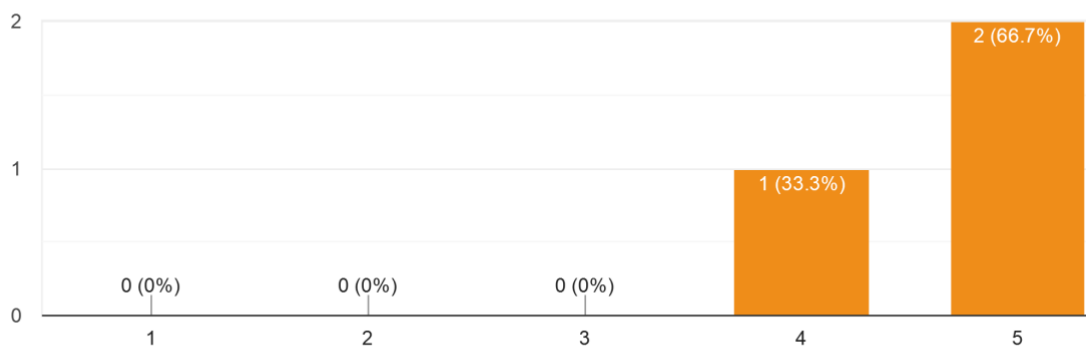
Where did you learn about the workshop/webinar ?

3 responses



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

3 responses



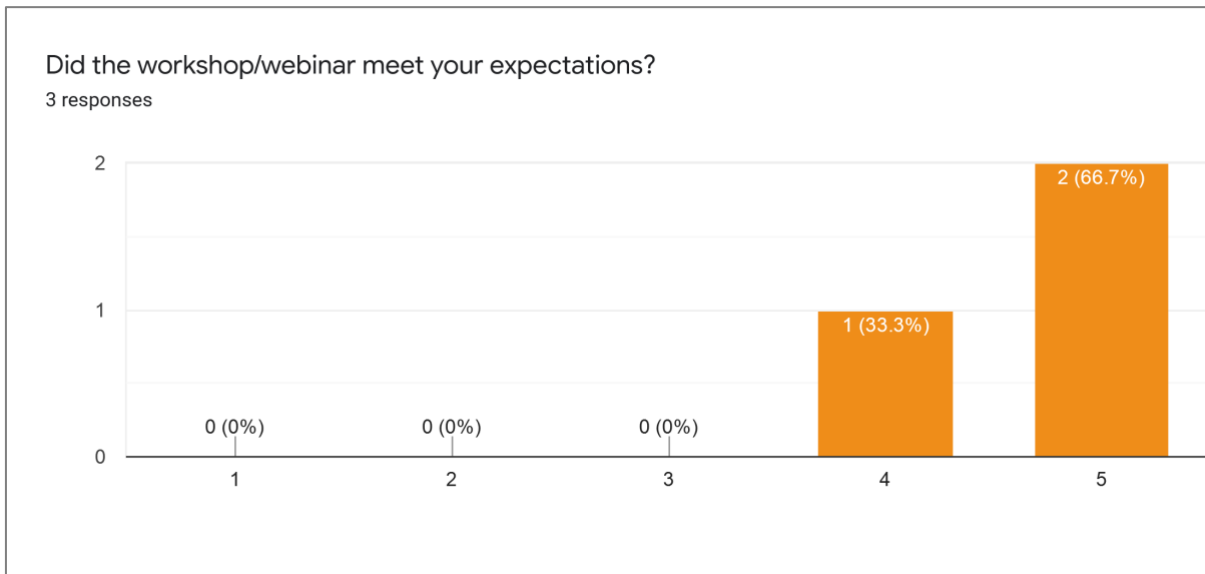
What did you hope to gain from the workshop/webinar?

3 responses

new experience and data

exchange interests

Knowledge of the workflow chosen in Archeological studies



Do you see this workshop/webinar having a positive impact on your work and how?

3 responses

yes

Yes

What did you like most about the workshop/webinar?

3 responses

all was interesting

subjects

xx

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

3 responses

no

nothing

xx

ANNEX 5

SSHOC DATAVERSE TRANSLATION WORKSHOP WORKSHOP REPORT

By Ana Inkret (CESSDA/UL-ADP)

Background

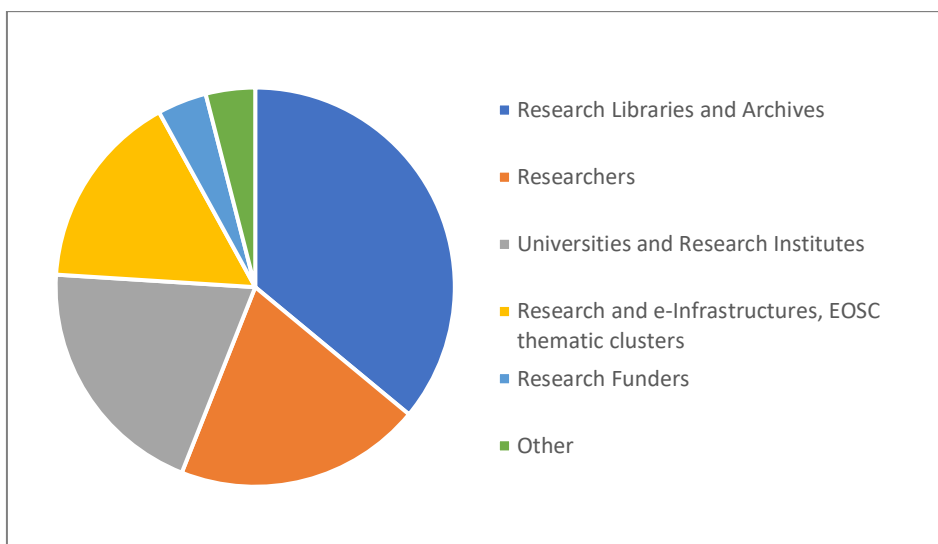
The [Dataverse translation workshop](#) took place online on 2 June 2021. It was organised as an engagement event to present the User Interface translation tool to service providers who had been following the SSHOC Dataverse adaptation and were interested in translating the UI to the national language.

Workshop Overview & Format

The aim of the workshop was to introduce the tool for translation of UI in Dataverse that SSHOC T5.2 had integrated into the installation pipeline. Workshop was also an opportunity to present the plans for the community of service providers.

The content of the workshop was planned and delivered by Marion Wittenberg, Laura Huis in 't Veld (both DANS), and Veronika Heider (AUSSDA). Technical support was provided by CESSDA/UL-ADP and GESIS.

The workshop had 29 registrants and 23 final attendants, including the organising team. Most of the participants belonged to the Research Libraries and Archives category (9), followed by Researchers and Universities and Research institutes (5 each). Research and e-infrastructures were represented by 4 participants. One participant identified themselves as a Research Funder and one as "other". Participants came from 10 European countries (Belgium, Bosnia and Herzegovina, Croatia, Cyprus, Finland, Germany, Italy, Lithuania, Netherlands, Norway, Poland, Portugal, Slovenia, Spain).



Workshop participants by stakeholder categories

The workshop started with an introduction of the speakers and the task, followed by an introduction of the Weblate tool and a presentation of experiences from the translation in German language. Workshop closed with a Q&A session and presentation of the future plans.

Presentations & Discussions: Key Points

Introduction

Speaker. Marion Wittenberg (DANS)

Main points. Session briefly introduced the SSHOC project, the main Dataverse functionalities developed within the task 5.2, and the two main speakers. Task 5.2 is adapting the Dataverse software to the needs of the social science and humanities researchers in Europe. 4 ERICs are involved in developing previewers, migration solutions, connections to local PID providers, interoperability with controlled vocabularies, as well as a translation service for the translation of the user interface in European languages.

Introducing the Weblate tool - Collaboration and procedures

Speaker. Laura Huis in 't Veld (DANS)

Main points. For the translation of the user interface, the Weblate tool was chosen because of its friendly interface, ability to keep track of one's progress (shows what parts need editing and review), and ability to automatically detect the untranslated strings after an upgrade of the software. Weblate also offers functionality to collaborate with other translators. User interface of Weblate and its collaboration functions were presented. These exist both on a level of single language and between languages, such as the possibility of saving and reviewing suggestions, possibility of comparing translations and of using

existing translations from previous versions of the software. Participants also learned of the key procedures.

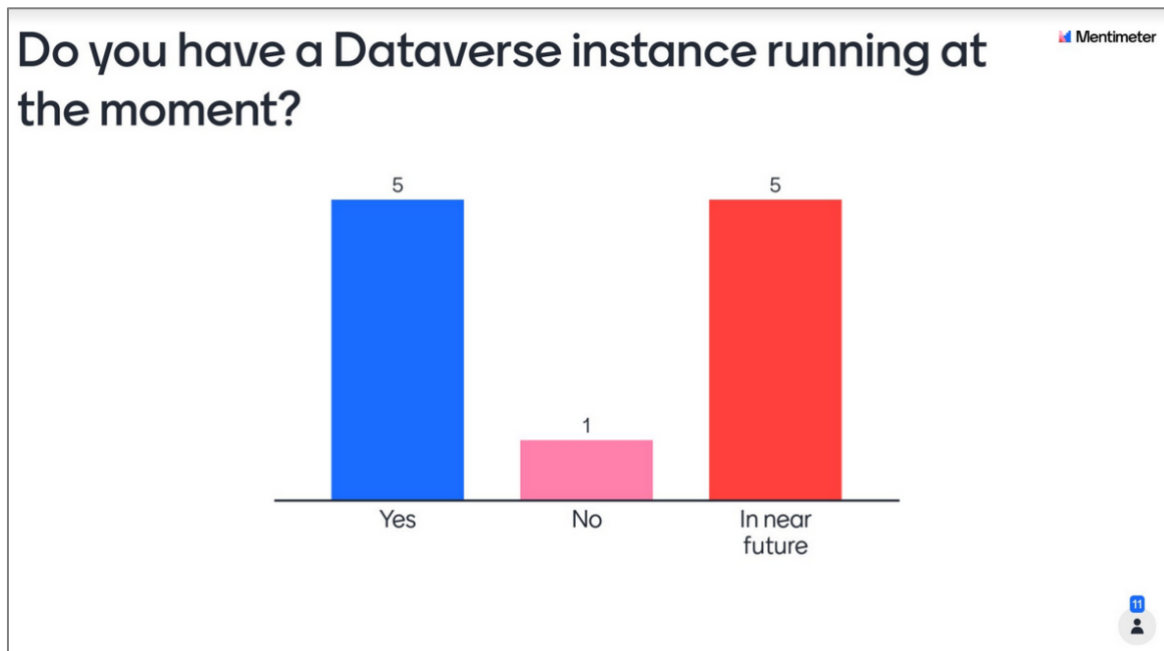
Experience from translating to German

Speaker. Veronika Heider (AUSSDA)

Main points. AUSSDA first translated elements of the UI in the DataverseEU project with the help of a professional translator to meet the deadline and ease the workload. Gender components in German, placeholders, and the missing context were challenging. In Weblate, on the other hand, it is easy to create new languages, add context for placeholders. Glossary function, adding contexts, cooperation with other users, and using other languages as reference points are clear advantages.

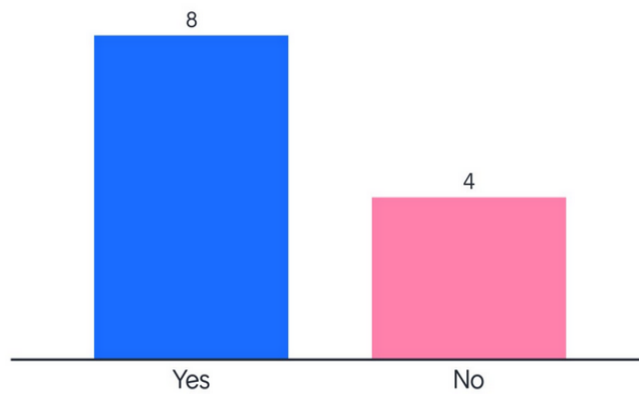
Q&A session

Main points. To find out more on the participants' knowledge of and experiences with the Weblate as well as their working languages for translations and plans to start a translation process, a Mentimeter survey was carried out.



Have you already had a look on the weblate tool?

Mentimeter



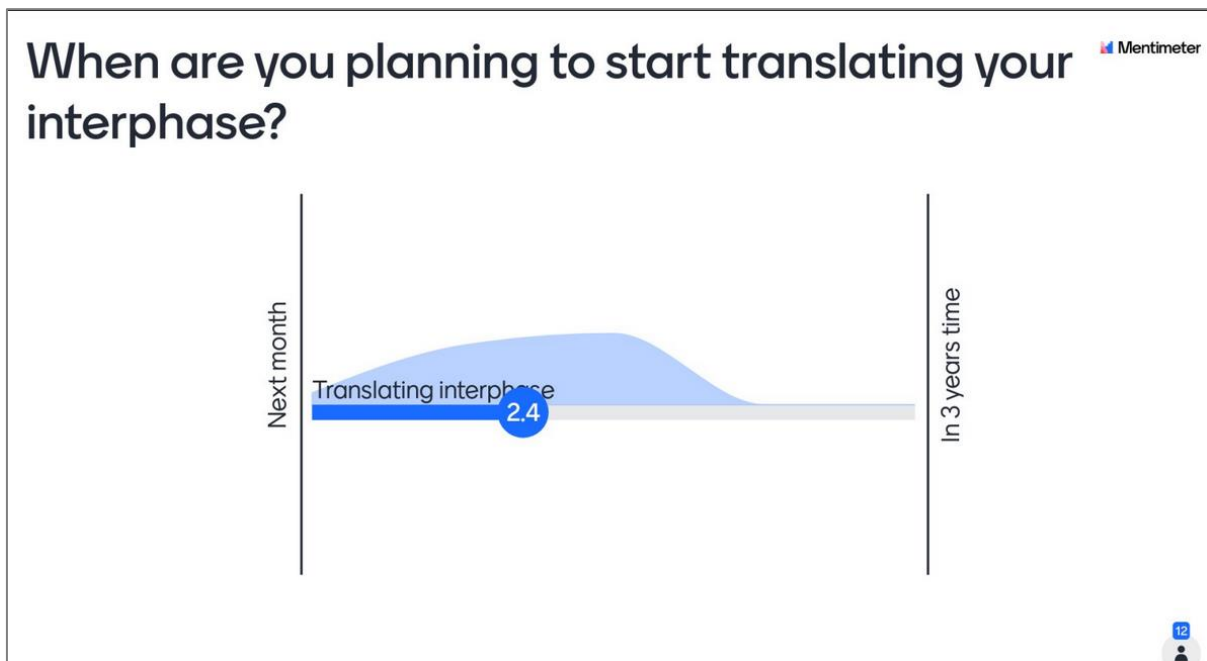
12

In which language would you like to translate your interphase?

Mentimeter



11



Speakers discussed questions that were collected at registration and posed during the presentations. Among other topics, participants asked about the translation process, adding new languages to Weblate and reusing existing translations. Questions and answers were edited after the workshop and are accessible on the [website of the workshop](#).

Links to materials

Slides: <http://doi.org/10.5281/zenodo.5017398>

Video: <https://www.youtube.com/watch?v=Wilxney3ORU>

Q&A: <https://sshopencloud.eu/events/sshoc-dataverse-translation-workshop>

Outcomes & Feedback

Only two participants responded to the SSHOC evaluation form (see Appendix 1). Overall, the event was positively evaluated. One of the participants evaluated the event as “excellent”, another one has chosen the option “very good”. The organization was evaluated as “excellently organized” by one participant, and “very well organized” by another.

Similarly, positive was the feedback and rating with regards to the question of whether the event met participants’ expectations. One participant reported that the event exceeded their expectations, and one agreed that the workshop greatly exceeded their expectations.

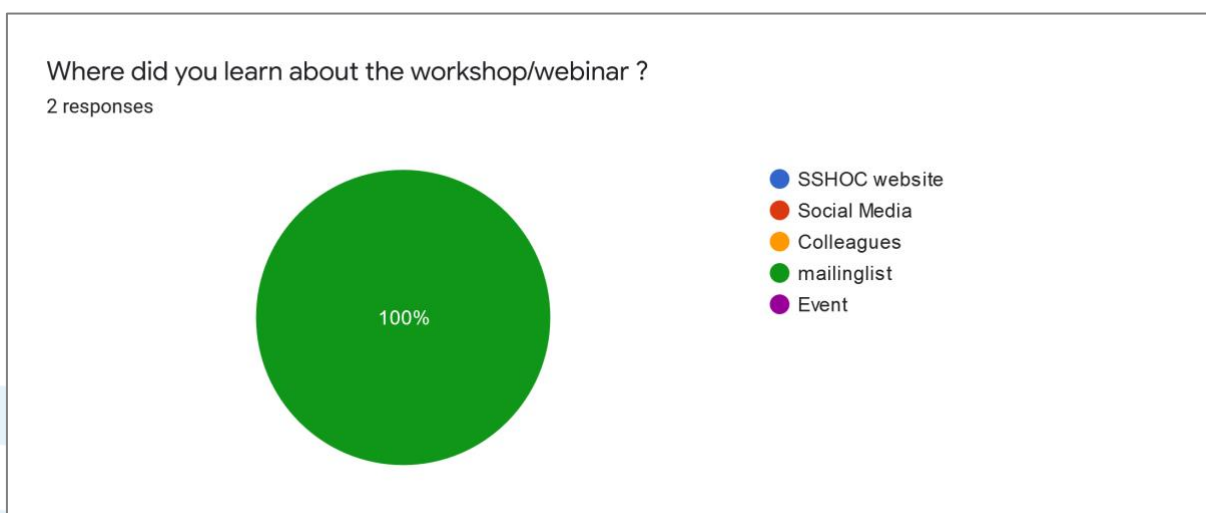
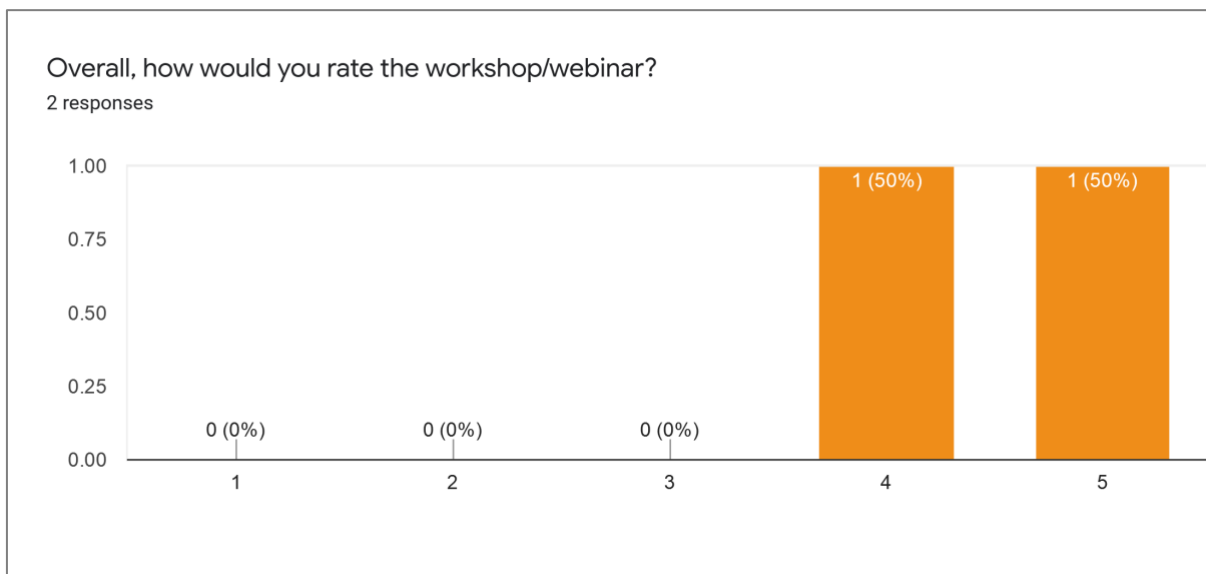
Participants expected to get “information on how to translate Dataverse in their language”, and “start with translation”. The positive feedback was given with regards to the usefulness of the workshop. In particular, respondents reported that they found all of the presentations and discussions to be

“practical/advisory” with great examples. Responding to the question of what could be improved, they stated that “in the next step it would be useful to present on exporting/integrating translation into github/dataverse” and to “show the translation process in an interactive way (no screenshot)”.

As was discussed during the workshop, the Weblate support from T5.2 will continue in autumn 2021 with a follow-up meeting to discuss participants’ user experience and gather feedback on any missing functionalities. Team is available for questions in the meantime.

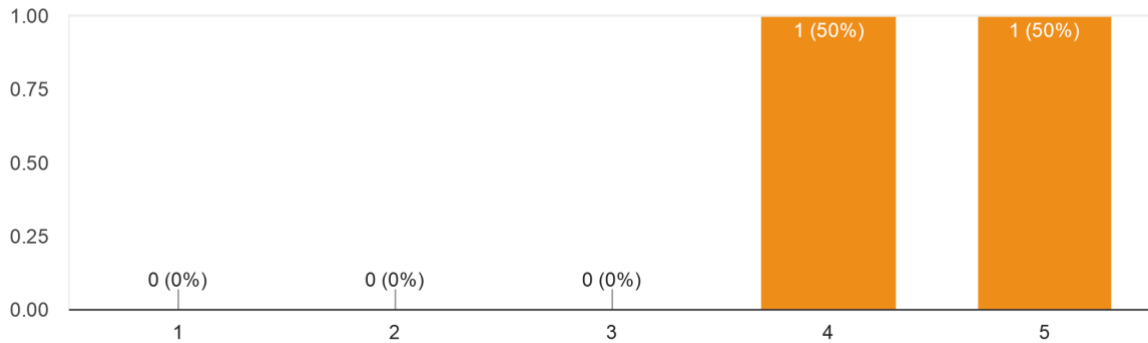
After the workshop, the video recording, slides, and answers to questions from the Q&A session were published on the [SSHOC website](#).

Results of the post-event survey



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

2 responses



What did you hope to gain from the workshop/webinar?

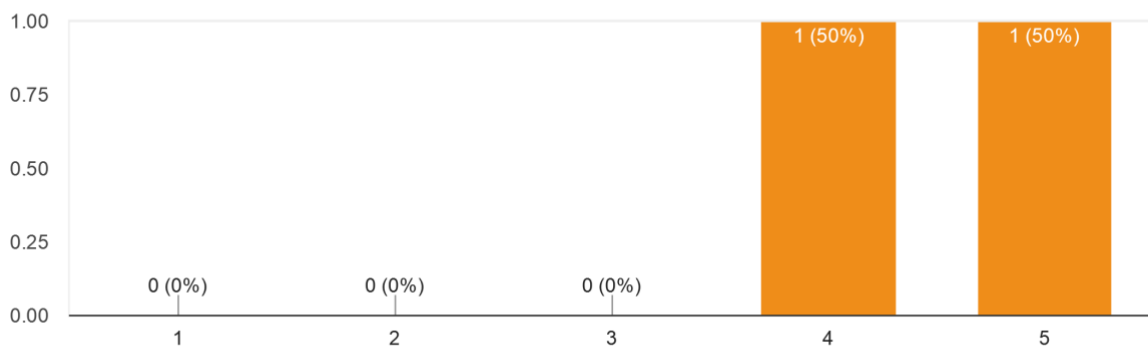
2 responses

Start with translation

Information on how to translate Dataverse in my language

Did the workshop/webinar meet your expectations?

2 responses



Do you see this workshop/webinar having a positive impact on your work and how?

2 responses

Yes, very useful, hands-on

Yep

What did you like most about the workshop/webinar?

2 responses

Practical/advisory

Examples

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

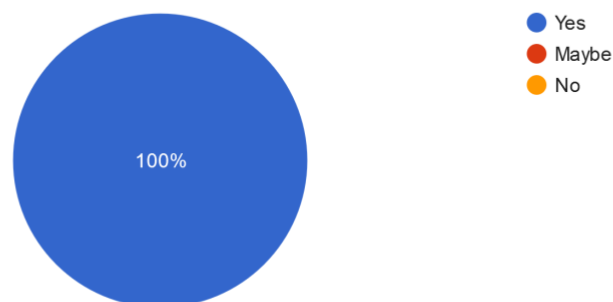
2 responses

Maybe in the next step it would be useful to present on exporting/integrating translation into github/dataverse

Show the translation process in an interactive way (no screenshot)

Are you planning to organize the translation of the Dataverse User Interface into your national language?

2 responses



ANNEX 6

SSHOC DATAVERSE TRANSLATION FOLLOW-UP EVENTS

REPORT OF THE WORKSHOPS

By Ana Inkret (CESSDA/UL-ADP) and Judith Wehmeyer (CESSDA/GESIS)

Background

The two engagement and awareness events covered by this report were organised to follow-up with the work that started with the [SSHOC Dataverse translation workshop](#) in June 2021. The initial workshop introduced the Weblate software, a tool for the translations of the Dataverse user interface and the support offered within the SSHOC project. The community of service providers and the organisers (SSHOC T5.2 “Hosting and sharing data repositories”) then met again on 8 September and 7 December to continue the discussion on their progress and needs.

1st Follow-up Meeting

Workshop Overview & Format

The aim of the [first follow-up event](#) was to offer the service providers who had begun using the translation tool an opportunity to share their first Dataverse translating experiences and discuss translating issues, the Weblate software, the content of the User Guide and future collaboration.

The event took place online on 8 September 2021. It was organised and delivered by SSHOC T5.2 (Marion Wittenberg and Laura Huis in 't Veld, DANS, and Veronika Heider, AUSSDA). It was technically supported by T6.2.

The workshop had 15 participants in total. Of these, 5 did not participate in the first workshop, which proves the reach of the materials that were openly available as well as relevance of the topic. Audience was quite diverse: the majority identified themselves as belonging to research libraries and archives (8),

3 as representatives of research and e-infrastructures and EOSC thematic clusters, and 2 as researchers. Universities and research institutes and research funders each had a single representative.

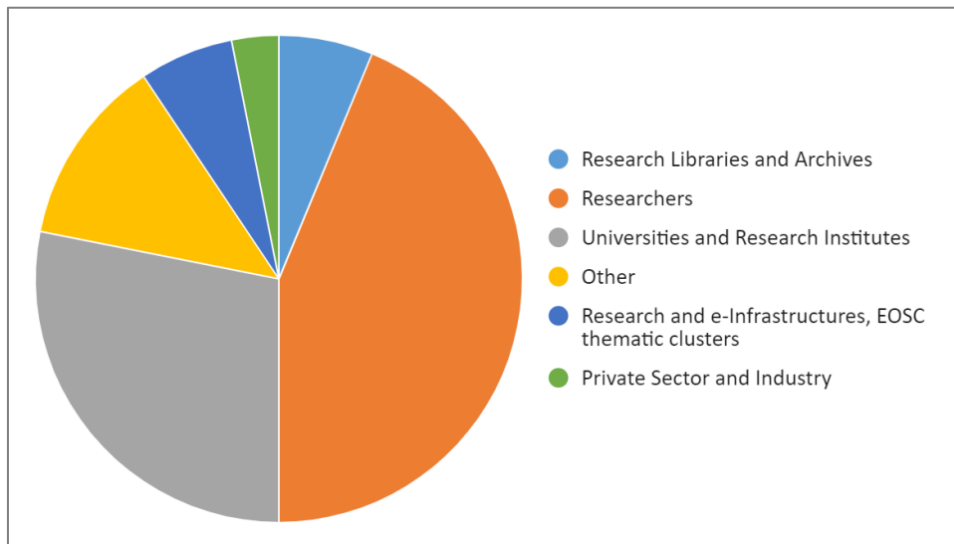


Figure 1 Participants by stakeholder categories

Discussion: Key Points

- Progress updates: CROSSDA starting translation with the help of a professional translator; ADP recently completed the translation and implemented it; Belgium State Archive is legally obligated to translate to all three national languages.
- Dataverse version used: ADP version 5.2, Belgium State Archive 5.6, AUSSDA will be updating in November to 5.5, DANS to 5.6
- Differences in Dutch and Flemish might mean producing two different translations.
- ADP's experience: a colleague needed about a month to do the translations without Weblate. It now took about two days to update everything in Weblate.
- Importance of context: need to double check strings which have to be found in the tool.
- For professional translation, engage translators with technical expertise.
- Weblate will have to be updated for the latest versions of Dataverse.
- Next meeting was set for December 7th at 15:00 CET.

The discussion was presented in a [blog post](#) that was posted after the meeting.

2nd Follow-up Meeting

Workshop Overview & Format

The aim of the [second follow-up event](#) was to give the participants further opportunity to exchange their translating experiences and discuss the translation of the new versions of Dataverse and future meetings.

The event took place online on 8 December 2021. It was organised and delivered by Marion Wittenberg, DANS, and Veronika Heider, AUSSDA, SSHOC T5.2. It was technically supported by T6.2.

A total of 13 participants joined the event. As expected, those identifying themselves as representatives of research libraries and archives were the majority (9 in total). Representatives of policy makers, research and e-infrastructures, EOSC thematic clusters, and universities and research institutes were also present. Participants joined from Austria, Belgium, Croatia, Denmark, Germany, Lithuania, Netherlands, Poland, Portugal, Slovenia.

The event started with a short introduction of the previous activities and resources. After an introduction of all the participants, a lively debate developed.

Discussion: Key Points

The discussion started with an introduction of all the participants and their experiences or plans for translation of the UI, which ranged from planning the work on institutional or even national installations to having completed the translation. The main points of the debate that followed were:

- Slava Tykhonov (DANS) explained how translators can access and work on the newer versions of Dataverse (5.6, 5.7, and 5.8). Translations of the previous versions are easily transmittable to Weblate and can be used as templates for any version that service providers are using.
- The translations will be available in the Global dataverse Consortium Community (GDCC) Github.
- A problem with special characters was noted due to an encoding issue that appears when importing the translation of the 5.2 version to work on the newer versions. The team will resolve this.
- Forming a community that would keep updating the versions in Weblate was discussed as a sustainability solution after the SSHOC project ends.
- The community will have to agree on the potential licences for the translations.
- The team has worked on support for controlled vocabularies (CVs) through SKOSMOS. This means that the translation of the CVs has to be done at the site of the responsible organisation for the CV.
- Bi- or multilingual metadata fields might be part of future work if the community requires them.

- A participant suggested to showcase the enhancements developed within SSHOC, their meaning and how to implement them to the broader Dataverse community even before the general Dataverse community meeting in June. This will be discussed with the Harvard office.
- Participants requested to continue the meetings. The next follow-up meeting, tentatively planned for April, will discuss the sustainability of the SSHOC Weblate community, licensing, and reuse of translations.

Outcomes & Feedback for the follow-up meetings

The second and third event of the series confirmed that the service providers who are implementing Dataverse at their institutions appreciate the opportunity to speak with the developers, share their experiences and needs with others, and receive advice for any issues they encounter. With the events, the SSHOC project has succeeded in both bringing the technical development closer to the users and in fostering an active community whose continued involvement with the development will be of benefit to the service providers as well as to the developers.

Blog posts were published after each of the events.¹⁵

¹⁵ Workshop Notes: SSHOC Dataverse Translation follow-up event, <https://sshopencloud.eu/news/workshop-notes-sshoc-dataverse-translation-follow-event>, and Meeting Notes: SSHOC Dataverse Translation Updates, <https://www.sshopencloud.eu/news/sshoc-workshop-notes-dataverse-translation-follow-event#overlay-context=news> (accessed 8 December 2021).

ANNEX 7

SSHOC WORKSHOP: PROVIDING CANONICAL TRAINING MATERIALS FOR SECURE DATA FACILITY PROFESSIONALS

WORKSHOP REPORT

By Ana Inkret (CESSDA/UL-ADP)

Background

The [Providing canonical training materials for secure data facility professionals workshop](#) took place online on 21 September 2021. This awareness-raising workshop was designed to introduce the training materials produced within T5.4 and provide feedback for their future development.

Workshop Overview & Format

The aim of the workshop was to discuss the format and development of the training materials, present the materials and how they could be adapted, and to get feedback that would enable T5.4 to further develop and improve the materials.

The content of the workshop was organised by Deborah Wiltshire (CESSDA/GESIS) from T5.4 who was also the main speaker. T6.2 (Trust-IT and CESSDA/UL-ADP) provided the technical support.

The workshop had 39 participants in total. They belonged to four stakeholder categories: libraries and archives (13 participants), researchers (12 participants), research and e-infrastructures (8 participants), and universities and research institutes (6 participants). Most participants came from Germany (13), United Kingdom and Netherlands (4 each) as well as Croatia and Slovenia (3 each). Representatives from Canada, Denmark, France, Italy, Norway, Portugal, Romania, Spain, Switzerland, and the United States were also present.

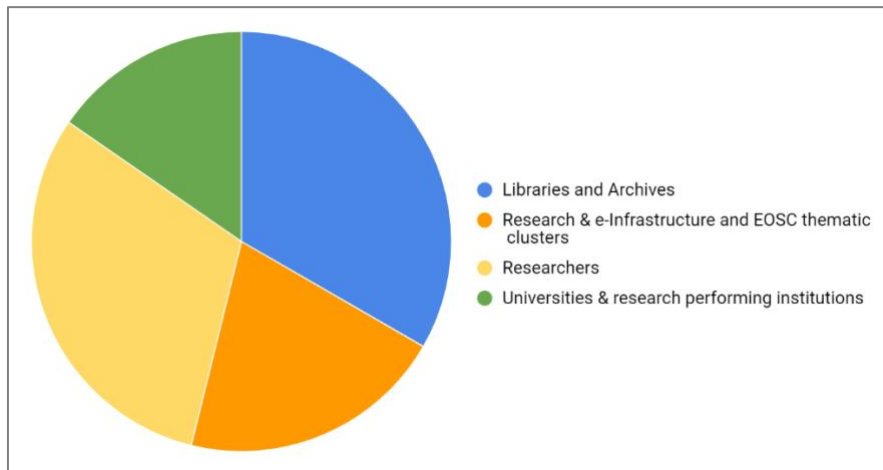


Figure 1 Distribution of participants

Workshop started with a presentation of the development of the training materials and an introduction of the modules. In the final part of the workshop, participants discussed the materials in smaller groups and reported their comments.

Presentations & Discussions: Key Points

The workshop was summarized in *Deliverable 5.20 Training materials of workshop for secure data facility professionals*.¹⁶

Links to materials	Slides: https://doi.org/10.5281/zenodo.5541587
	Video recording: https://youtu.be/svx-n4LEh6M
	Related deliverable: https://doi.org/10.5281/zenodo.5638596

Outcomes & Feedback

As was summarized in the resulting deliverable (see footnote), the workshop was valuable to the organizers as well participants and will facilitate further work in ensuring the safe use of secure data. The value of the training materials was recognized, feedback for the materials was positive and included helpful recommendations. As a result, a supplementary guide for trainers was created after the

¹⁶ Deborah Wiltshire. (2021, September 30). SSHOC Workshop: Providing canonical training materials for secure data facility professionals. Zenodo. <https://doi.org/10.5281/zenodo.5541587>
 Deborah Wiltshire. (2021). D5.20 Training materials of workshop for secure data facility professionals (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.5638596>

workshop. Participants of the workshop, a successful mix of those invited to participate directly due as representatives of secure data facilities and broader SSHOC audience, will continue to meet. A [blog post](#) was published after the workshop. Video recording, training materials, and supplementary guide¹⁷ are also freely available.

¹⁷ Both the training materials and the guide are available as part of the deliverable 5.20 (see above).