

## Social Sciences & Humanities Open Cloud

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

### Deliverable 6.14 Report on Training Workshops

Dissemination Level	PU
Due Date of Deliverable	28/02/2022 (M38)
Actual Submission Date	14/03/2022
Work Package	WP6 - Fostering Communities, Empowering Users, & Building Expertise
Task	Task 6.5 Coordinating Targeted Training in the Social Sciences and Humanities
Type	Report
Approval Status	Waiting EC approval
Version	V1.0
Number of Pages	p.1 – p.96

**Abstract:**

This report summarizes the organisation, structure and outcomes of training events organised by and delivered as workshops with the main aim to provide targeted training to the SSH community. The events offered both generic and highly topic-specific training content covering six themes: Data Science for the SSH, Data Science for Heritage Science, Data Protection and the GDPR, Data Stewardship and RDM, Data Citation, and Text Mining for the SSH. They especially targeted and successfully reached researchers, librarians and archivists, and representatives of research performing institutions. This report gives detailed information about the 11 workshops organised from July 2019 to February 2022.

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	04/02/2022	First draft	Kristina Pahor de Maiti (CLARIN/UL-FF) Darja Fišer (CLARIN/UL-FF)
0.2	11/02/2022	First review received	Tatsiana Yankelevich (LIBER)
0.3	14/02/2022	Implementation of changes required by first review	Kristina Pahor de Maiti (CLARIN/UL-FF) Darja Fišer (CLARIN/UL-FF)
0.4	22/02/2022	Second review received	Judith Wehmeyer (CESSDA/GESIS)
1.0	28/02/2022	Implementation of changes required by second review and submission of the final version	Kristina Pahor de Maiti (CLARIN/UL-FF) Darja Fišer (CLARIN/UL-FF)

## Author List

Organisation	Name	Contact Information
CLARIN/UL-FF	Kristina Pahor de Maiti	<a href="mailto:Kristina.Pahordemaiti@ff.uni-lj.si">Kristina.Pahordemaiti@ff.uni-lj.si</a>
CLARIN/UL-FF	Darja Fišer	<a href="mailto:Darja.Fiser@ff.uni-lj.si">Darja.Fiser@ff.uni-lj.si</a>

## Editor List

Organisation	Name	Contact Information
LIBER	Tatsiana Yankelevich	<a href="mailto:Tatsiana.Yankelevich@libereurope.org">Tatsiana.Yankelevich@libereurope.org</a>
CESSDA/GESIS	Judith Wehmeyer	<a href="mailto:Judith.wehmeyer@gesis.org">Judith.wehmeyer@gesis.org</a>

## Executive Summary

Targeted training events aimed to empower data users, data producers and data experts through skills and knowledge transfer to maximize the uptake of SSHOC resources and to promote data-driven and cross-disciplinary research directions. This goal was achieved through the organisation of twenty training workshops and webinars which were conceptualized as complementary training events. They primarily targeted and successfully reached researchers and representatives of research performing institutions and research libraries, but also other stakeholders identified as relevant for SSHOC (e.g., research infrastructures, private sector, and civil society). Because of the training design in the form of workshop-webinar pairs and due to pandemic-related restrictions that resulted in most of the events being delivered online, there was a considerable overlap of several organisational aspects for both types of events. For this reason, this report describes the common aspects of the organisation, structure and outcomes of both workshops and webinars, but also provides details about workshops when needed. For webinar-specific aspects, refer to *Deliverable 6.15 Report on Training Webinars* (Pahor de Maiti & Fišer, 2022).

In particular, this report gives details on eleven targeted training workshops which were organised from July 2019 to February 2022 and followed six thematic clusters:

- 1. Data Science for the Social Sciences and Humanities**
  - 1.1. The Case of Interview Data: A multidisciplinary approach to the use of technology in research using interview methods (07/2019)
- 2. Data Science for Heritage Science**
  - 2.1. Citizen Science & Cultural Heritage. Planning for success (03/2021)
  - 2.2. Digitising Museum Objects Using Basic Photogrammetry (03/2021)
- 3. Data Protection and the General Data Protection Regulation**
  - 3.1. Data Protection in Research Practice (10/2021)
- 4. Data Stewardship and Research Data Management**
  - 4.1. Caring for Sharing – Data Management and FAIRness of Migration Data (03/2020)
  - 4.2. Copyright Issues in Secondary Data Use (01/2022)
  - 4.3. Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing (02/2022)
- 5. Data Citation**
  - 5.1. Data Citation in Practice (06/2021)
- 6. Text Mining for the Social Sciences and Humanities**
  - 6.1. Using Corpora for Implementing Validation (09/2019)
  - 6.2. Exploration of Society Through the Lens of Labour Market Related Documentation (05/2021)
  - 6.3. ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates (05/2021)

The organisational process for each training event consisted of three stages: (1) specification of topics and content, target audience, speakers, venue, and scheduling, (2) promotion and event delivery, and (3) post-event activities. The specific topics and speakers were identified in close collaboration with subject-

matter experts from SSHOC and related communities in order to find most relevant topics and to provide state-of-the art content to the audience. The duration of the workshops varied (1.5h, half-day, full-day or multiple days) and depended on the content and the number of interactive activities planned. They consisted of presentations to transfer knowledge, hands-on activities to transfer skills, and dedicated time for a discussion to address open issues and future steps. Due to the COVID-19 pandemic, most of the workshops, which were initially planned as face-to-face events, had to be delivered online. Despite some downsides, the online and hybrid formats also brought certain advantages, especially in terms of an expanded outreach: firstly, people who were unable to travel could participate in online events, and secondly, since the online events can be easily recorded and shared, those who could not attend the workshop or webinar got the opportunity to watch the presentations afterwards.

In total, twenty training events (workshops and webinars combined) reached 2455 people (1102 participants of live (streaming) events and 1353 views of event recordings). The workshops reached 690 people: 464 people attended the live events, while playbacks of the recordings, published on the SSHOC YouTube channel, so far accounted for another 226 views (obtained on 28/02/2022).

## Abbreviations and Acronyms

FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
Q&A	Questions and Answers
RDM	Research Data Management
SSH	Social Sciences and Humanities
SSHOC	Social Sciences & Humanities Open Cloud
WP	Work Package

## Table of contents

<b>1. Introduction</b>	<b>7</b>
<b>2. Organisation of the Events</b>	<b>8</b>
2.1 Event planning	8
2.2 Event promotion and delivery	9
2.3 Post-event activities	10
<b>3. Summary of the Workshops</b>	<b>11</b>
3.1 Data Science for the SSH	14
3.1.1 The Case Of Interview Data: A Multidisciplinary Approach to the Use of Technology in Research Using Interview Methods	14
3.2 Data Science for Heritage Science	15
3.2.1 Citizen Science & Cultural Heritage. Planning for Success.	15
3.2.2 Digitising Museum Objects Using Basic Photogrammetry	16
3.3 Data Protection and the GDPR	17
3.3.1 Data Protection in Research Practice	17
3.4 Data Stewardship and RDM	18
3.4.1 Caring for Sharing – Data Management and FAIRness of Migration Data	18
3.4.2 Copyright Issues in Secondary Data Use	19
3.4.3 Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing	20
3.5 Data Citation	21
3.5.1 Data Citation in Practice	21
3.6 Text Mining for the SSH	22
3.6.1 Using Corpora for Implementing Validation	22
3.6.2 Exploration of Society Through the Lens of Labour Market Related Documentation	23
3.6.3 ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates	24
<b>4. Outcomes and Conclusions</b>	<b>25</b>
<b>5. Reference list</b>	<b>29</b>
<b>6. Annexes</b>	<b>30</b>

# 1. Introduction

This report concerns targeted training events organised by WP6, Task 6.5, in the form of training workshops and webinars which were designed as complementary events. Originally, the workshops were conceptualized as comprehensive and immersive in-person training sessions with an emphasis on hands-on activities in order to create a supportive learning environment. Webinars were conceived as thematically narrower and shorter online events that would precede or follow the workshops. Thus, the targeted workshops together with webinars formed a methodological training pair within Task 6.5. But due to pandemic-related restrictions, most of the events of both kinds had to be delivered online. For this reason, but also because of the training design in the form of workshop-webinar pairs, the organisational process as well as the outcomes were largely overlapping. In order to provide a concise report on the work done in SSHOC Task 6.5, this deliverable describes the common aspects of the organisation, structure and outcomes of both workshops and webinars, but also gives a detailed account of the workshops conducted. For webinar-specific aspects, please refer to *Deliverable 6.15 Report on Training Webinars* (Pahor de Maiti & Fišer, 2022).

The aim of the training events was to maximize the SSHOC impact among data users, data producers and data experts by providing them with relevant training content that helps develop skills and competences needed to foster cross-disciplinary cooperation in the SSH. They also focused on fully leveraging SSHOC services, tools and data in order to facilitate and promote data-driven research in the SSH. The events thus crucially contributed to one of the main SSHOC objectives: empowering individuals to maximise data re-use through Open Science and FAIR principles.

The focus of the training events was on showcasing best practices of working with data across six different thematic clusters (Data Science for the SSH, Data Science for Heritage Science, Data Protection and the GDPR, Data Stewardship and RDM, Data Citation, and Text Mining for the SSH). One workshop and one webinar per thematic cluster were planned in the SSHOC Grant Agreement. The organisation of the training events followed the outcomes of *Deliverable 6.2 Building Expertise Strategy*<sup>1</sup> (Torma et al. 2019) in order to harmonize efforts and consolidate training activities in line with WP6 objectives. The workshops were designed to reach diverse audiences in terms of their geographic location and SSHOC stakeholder profile as defined in *Deliverable 2.1 SSHOC Overall Communication and Outreach Plan*<sup>2</sup> (Schwabe et al., 2019).

The rest of the report is organised as follows: Section 2 of this report outlines the process followed in the organisation of the training events. Section 3 gives a summarized list of the workshops, and section 4 provides an overview of the outcomes. Full reports for all the workshops are annexed at the end (see Section

---

<sup>1</sup> Torma, et. al., *SSHOC D6.2 Building Expertise Strategy* (v1.0), <https://doi.org/10.5281/zenodo.4558294>

<sup>2</sup> Schwabe, A., et. al., *SSHOC D2.1 Overall Communication and Outreach Plan*, <https://doi.org/10.5281/zenodo.3595936>

6. Annexes).

## 2. Organisation of the Events

The organisation of targeted training events comprised three stages: event planning, promotion and event delivery, and post-event activities.

### 2.1 Event planning

The planning of the events included the following key aspects: specification of topics and content, target audience, speakers, venue and scheduling. The specific content of each event was determined in accordance with one of the six predefined topics<sup>3</sup> and the training needs of the SSH community in that topic area. The needs were identified based on *Deliverable 6.2* (Torma et al. 2019) and in close collaboration with Task 6.5 members and other project partners having specific topic expertise as well as with the activities of other SSHOC WPs in order to best showcase the SSHOC results. A similar approach was adopted when scouting for suitable speakers. These were first sought after within the SSHOC community, but the team of speakers was often joined by experts from other communities in order to encourage cooperation and broaden the reach of the SSHOC project outcomes. The key target audience for training events comprised three stakeholder categories: *Researchers, research networks and communities; Research libraries and archives; Universities and research performing institutions* (cf. Schwabe et al., 2019).

Workshop venues were selected for each workshop separately. Initially, all workshops were meant to be face-to-face events co-located with major SSH conferences or summer schools in order to maximize the outreach. However, the emergence of Covid-19 forced the team to alter the format of most workshops from on-site to online or hybrid events which meant that some workshops had to be delivered as stand-alone events. Migrating workshops to the virtual environment had substantial consequences for the workshop programme and for the delivery itself. In order to avoid potential negative effects arising from using the online format (e.g. technical difficulties, shorter and more fragile attention span compared to physical environment, lack of personal contact), special care was taken to use safeguards to avoid technical mishaps (e.g. organising test meetings weeks before the event, sharing best practices of online presenting with the speakers, connecting well before the meeting to check audio/video/materials/interactive options, using one central slide deck, having a dedicated person for technical support) as well as to consider the specifics of the virtual environment when creating the programme (e.g., breaking down the content into shorter sessions, preparing prompts to encourage interaction, engaging with participants in small groups in break-out rooms, rearranging content into a combination of activities and prompts provided to the participants before, during and after the event: the participants for example received reading materials, instructions for software download or additional activities). While the format was changed for most of the workshops, the team was able to keep most of

---

<sup>3</sup> Data Science for the SSH, Data Science for Heritage Science, Data Protection and the GDPR, Data Stewardship and RDM, Data Citation, and Text Mining for the SSH.



the events co-located with the identified conferences and summer schools since these were also migrated online. In case when the hosting conference was cancelled, the team identified an alternative hosting event or opted for a self-standing event if this was the only way to provide continuous training activities.

The scheduling of the training events followed two principles: a balanced distribution across the entire project duration and compatibility with the academic year calendar in order to ensure regular training and high participation rates. While showcasing SSHOC results was one of the priorities for this task, training events could not be strictly aligned with release dates of SSHOC services and tools, since these dates were often pushed to the final months of the project which could lead to unwanted cluttering of training events at the end of the project. To address this challenge and accommodate both objectives, i.e., promoting SSHOC results and ensuring continuous training offer, the training events embedded final or early versions of SSHOC results in their program whenever possible, but a regular pace of delivering events was followed in order to keep in touch with the community. Furthermore, additional events (i.e., workshops and webinars on top of the twelve events required in the Grant Agreement) were organised in order to address training needs of specific communities or present SSHOC results that were initially set aside because of incompatible scheduling of the service/tool release and the training event plan.

## 2.2 Event promotion and delivery

Once the preparatory stage was completed, cooperation between WP2 and WP6 proved fruitful in promoting the training event up to its delivery. The promotion included SSHOC internal communication means (Basecamp) and external communication channels (SSHOC social media channels, SSHOC website, and relevant channels of SSHOC partner institutions). In contrast to webinars which were one-hour events, the workshops had different durations. Depending on the type and extent of the content, level of interactive activities planned, and the online/on-site format, the workshops could be short (e.g., approx. 1.5 hours) or long events (half, full or multiple day(s)). A flexible timeframe was necessary, so that both the informative and interactive/demonstrative function of the workshops were properly covered. The workshops included presentations which were focused on transmitting knowledge, hands-on activities which were dedicated to skills transfer, and dedicated time for discussion where participants and speakers could freely engage in a debate. Each workshop included a short presentation of the SSHOC project, and the workshop content was situated in relation to the project goals and/or outcomes. The online training events were moderated by a host and technical support staff was provided in order to ensure smooth progression between programme sections. Feedback and event evaluations were collected during and after the event via live polling options and post-event surveys.

The T6.5 team needed to take on two challenges at the online training event delivery and evaluation stage. The first one concerned *virtual fatigue*. Due to the Covid-19 pandemic and related sanitary measures, most face-to-face events in academia were moved online which soon caused an unprecedented multiplication of online events. Given the large amount of training possibilities available online, combined with other professional and personal challenges caused by the pandemic, it was increasingly hard to attract targeted participants. This could result in lower participation rates and a less

engaged audience, and consequently, in a lower impact of SSHOC activities. As evident from reports on outreach (see Section 3), this challenge was successfully addressed by making sure the duration and format of the events maximally suited the content in order to optimize people's time and energy investment with regard to their knowledge gain, but also by putting more effort into targeted promotion. The second challenge concerned *survey fatigue*. For reasons explained above, post-event survey response rates dropped significantly which meant that the results were no longer representative and rarely helpful in terms of providing useful information for improving future training events. This challenge was addressed in two ways: first by reaching out to the participants with the survey several times, and second by collecting individual feedback from prompted and spontaneous comments during the workshop. These mitigating measures proved satisfactory, although additional feedback would have been preferred.

## 2.3 Post-event activities

The third and last organisational stage involved post-event activities which included reporting and dissemination of the materials. Each training event was followed by a blogpost, and publication of the presentation and recording (except for the on-site workshops) which are all freely available through the SSHOC website<sup>4</sup> and SSHOC Zenodo<sup>5</sup> and YouTube channels.<sup>6</sup> Links to the materials for each of the respective workshops are available in Section

---

<sup>4</sup> SSHOC website: <https://www.sshopencloud.eu>

<sup>5</sup> SSHOC Zenodo: <https://zenodo.org/communities/sshoc>

<sup>6</sup> SSHOC Youtube channel: <https://www.youtube.com/channel/UCw-mY8v84yeHW2z4KG3ZLtA/featured>

3. *Summary of the Workshops.* Each training event was also accompanied by a detailed report bringing information about its aim, organisers and speakers, concise descriptions of the presentations and participants' feedback. Full reports for the workshops are annexed at the end (see Section

6. Annexes).

### 3. Summary of the Workshops

Overall, eleven workshops were organised from July 2019 to February 2021, thus providing continuous training during the entire SSHOC project: six workshops were required as per the SSHOC Grant Agreement, and another five were organised on top of those at the initiative of the SSHOC training team, the lecturers, or the stakeholders who identified gaps in the existing training offer and saw the potential of SSHOC to fill this gap with knowledge transfer based on the resources, tools, best practices and knowledge generated within the project. Targeted training workshops covered six thematic clusters identified as relevant for the SSH community in the SSHOC Grant Agreement (see Table 1).

THEMATIC CLUSTER	WORKSHOPS
Data Science for the SSH	1
Data Science for Heritage Science	2
Data Protection and the GDPR	1
Data Stewardship and RDM	3
Data Citation	1
Text Mining for the SSH	3

Table 1: The number of workshops per thematic cluster

In total, the workshops were attended by **690 participants**: 464 participants attended the live events, and 226 participants (obtained on 28/02/2022) viewed the workshop recordings. The geographical coverage of the workshops was broad, since the participants represented on average 22 different countries per workshop covering Europe (EU and non-EU states), Africa, North and South America, Asia, Australia and the Middle East. It should be noted that, in contrast to training webinars which imposed no restrictions on the number of participants (cf. Pahor de Maiti & Fišer, 2022), some workshops accepted only a limited number of people in order to ensure adequate support to all the participants during the hands-on activities. For this reason, the attendance rates do not entirely reflect the interest expressed for a particular workshop.

Due to pandemic-related restriction, most of the workshops were adapted either to an online or hybrid format in order to ensure democratic access to the training as well as a broad outreach (see Table 2).

FORMAT	WORKSHOPS
Online	7
Onsite	3
Hybrid	1

Table 2: The number of workshops delivered in the indicated format

The workshops were of different duration (see Table 3). To accommodate the specifics of the virtual environment, the online workshops were either shorter than face-to-face events would be, i.e., around 1.5 hours, or divided into multiple shorter sections. The two longest workshops were delivered as part of

the summer school *Helsinki Digital Humanities Hackathon* which lasted for 10 days (for more, see Sections 3.6.2 EXPLORATION OF SOCIETY THROUGH THE LENS OF LABOUR MARKET RELATED DOCUMENTATION and 3.6.3 PARLAMINT – EXPLORING SOCIETAL ISSUES THROUGH COMPARABLE CORPORA OF PARLIAMENTARY DEBATES).

DURATION	WORKSHOPS
Approx. 1.5h	4
Half day	1
Full day	4
Multiple days	2

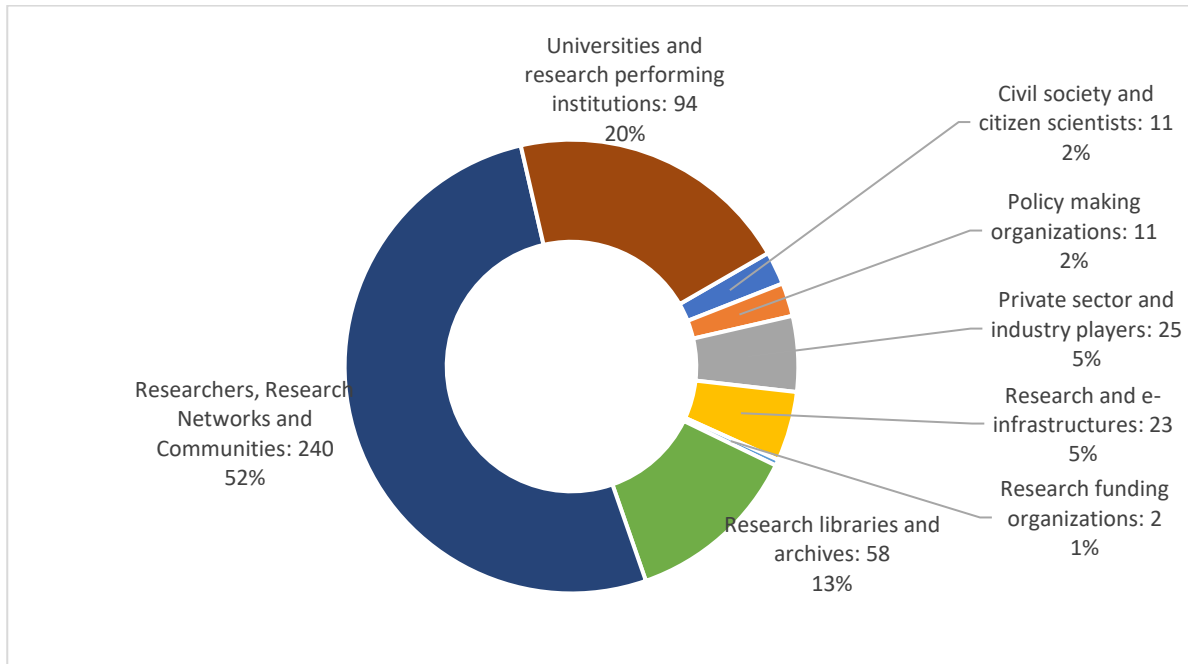
Table 3: The number of workshops according to their duration

As initially planned, most of the workshops were co-located with other prominent events since this brought additional exposure to the SSHOC project and its activities, as well as helped with reaching interested individuals (see Table 4). All SSHOC workshops were free to attend even when the conference itself had a registration fee. Some workshops were organised as stand-alone events in order to ensure a sufficient focus on the workshop content (i.e., visibility of the workshop can be reduced at events with many parallel sessions) as well as to keep the regular pace of providing training activities.

VENUE	WORKSHOPS
Stand-alone	3
Co-located	8

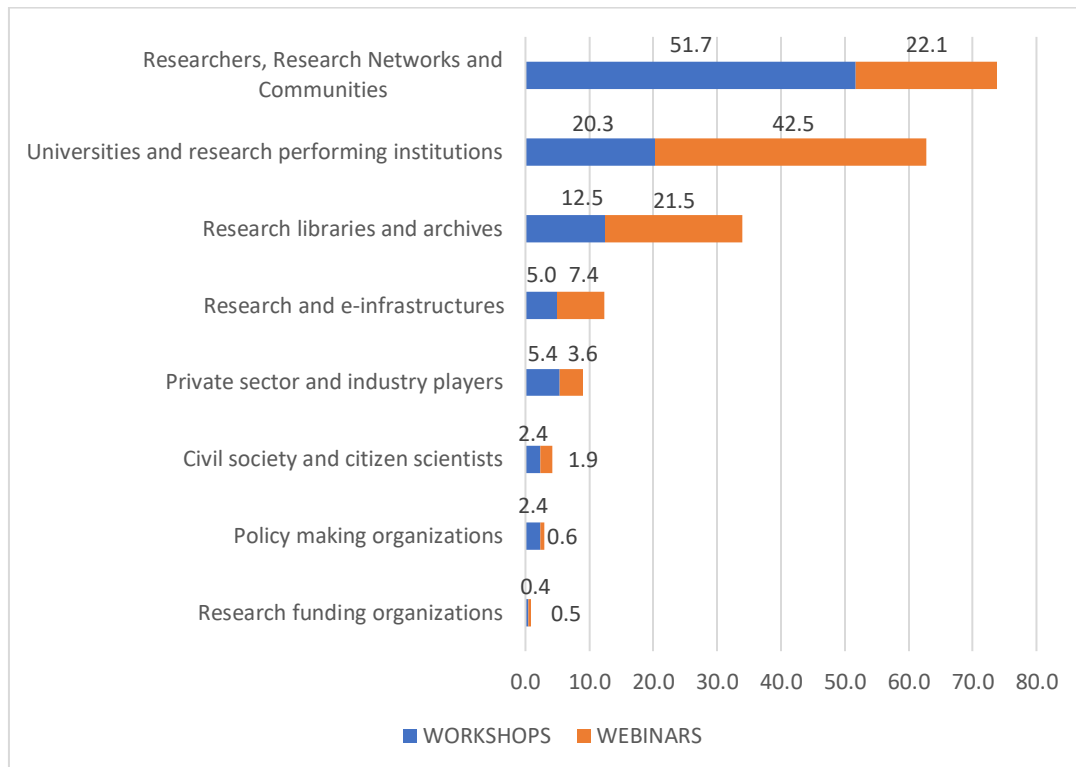
Table 4: The number of workshops per venue type

The distribution within the stakeholder categories is represented in Graph 1. A great majority (almost 85%) is taken up by the target groups which were identified as key target audiences for training workshops, i.e., *Researchers, research networks and communities; Research libraries and archives; Universities and research performing institutions*. Other stakeholder groups were also represented at the workshops, but with a smaller share.



Graph 1: The number and percentage of participants per stakeholder group

Graph 2 shows the comparison between the different stakeholder groups reached by workshops and webinars. While the overall share of participants from the three key target groups (researchers, research librarians and representatives of research performing institutions) was almost the same for workshops (85%) and webinars (86%), the individual shares for these three groups differ to a great extent: researchers more than twice as often attended the workshops in comparison to the webinars, while the participation of librarians and representatives of research performing institutions was around twice as frequent at webinars than at workshops. This is a good indicator of the preferred event format for different stakeholder groups—which can be usefully leveraged in the organisation of future events in order to offer trainings tailored to the needs of different communities.



Graph 2: The percentage of participants per stakeholder group

The following sections provide a list of the workshops according to the predefined topics in the SSHOC Grant Agreement together with a brief overview of the target audience, organisers and the links to published materials. Each workshop is described in further detail (e.g., the aim of the workshop, content of the presentations, participants' feedback) in the indicated workshop report annexed at the end (see Section



6. Annexes).

### 3.1 Data Science for the SSH

#### 3.1.1 The Case Of Interview Data: A Multidisciplinary Approach to the Use of Technology in Research Using Interview Methods

<b>Date and Venue</b>	07/2019 Utrecht, The Netherlands; co-located with the Digital Humanities 2019 conference
<b>Links to materials</b>	
Announcement	<a href="https://www.sshopencloud.eu/events/digital-humanities-2019-workshop-case-interview-data">https://www.sshopencloud.eu/events/digital-humanities-2019-workshop-case-interview-data</a>
Blogpost	<a href="https://sshopencloud.eu/news/working-interview-data-sshoc-workshop-multidisciplinary-approach-use-technology-research">https://sshopencloud.eu/news/working-interview-data-sshoc-workshop-multidisciplinary-approach-use-technology-research</a>
Presentation	<a href="https://zenodo.org/record/3521373#.YEieLV1Kj0p">https://zenodo.org/record/3521373#.YEieLV1Kj0p</a>
Recording	/
<b>Audience</b>	
Total participants	21
Recording playbacks	NA
Targeted audience	Researchers and others working with audio data
By stakeholder category	Researcher: 19 Private sector and industry player: 1 Research library and archive: 1
<b>Organisers</b>	T6.5 in cooperation with CLARIN ERIC and SSHOC T4.4
<b>Workshop report</b>	see Annex 1: Workshop report: The Case of Interview Data – A Multidisciplinary Approach to the Use of Technology in Research

### 3.2 Data Science for Heritage Science

#### 3.2.1 Citizen Science & Cultural Heritage. Planning for Success.

<b>Date and Venue</b>	03/2021
-----------------------	---------

	Online; co-located with the Sustainable Heritage Bidecennial Conference <sup>7</sup>
<b>Links to materials</b>	
Announcement	<a href="https://sshopencloud.eu/events/sshoc-workshop-citizen-science-cultural-heritage-planning-success">https://sshopencloud.eu/events/sshoc-workshop-citizen-science-cultural-heritage-planning-success</a>
Blogpost	<a href="https://sshopencloud.eu/news/sshoc-workshop-notes-citizen-science-cultural-heritage-planning-success">https://sshopencloud.eu/news/sshoc-workshop-notes-citizen-science-cultural-heritage-planning-success</a>
Presentation	<a href="https://doi.org/10.5281/zenodo.4776368">https://doi.org/10.5281/zenodo.4776368</a>
Recording	<a href="https://www.youtube.com/watch?v=Kv_ZvDQgydc">https://www.youtube.com/watch?v=Kv_ZvDQgydc</a>
<b>Audience</b>	
Total participants	80
Recording playbacks <sup>8</sup>	30
Targeted audience	Researchers and others interested in participative research
By stakeholder category	<p>Researcher: 33</p> <p>University and/or research performing organisation: 21</p> <p>Private sector and/or industry player: 8</p> <p>Civil society and/or citizen scientist: 5</p> <p>Policy making organisation: 5</p> <p>Research library and/or archive: 5</p> <p>Research funding organisations: 2</p> <p>Research and e-infrastructure: 1</p>
<b>Organisers</b>	T6.5 in cooperation with UCL
<b>Workshop report</b>	see Annex 2: Workshop report: Citizen Science & Cultural Heritage. Planning for Success

### 3.2.2 Digitising Museum Objects Using Basic Photogrammetry

<b>Date and Venue</b>	03/2021 Online; co-located with the Sustainable Heritage Bidecennial Conference <sup>9</sup>
-----------------------	---

<sup>7</sup> UCL, Sustainable Heritage Bidecennial Conference: Strategic Research Questions, <https://www.ucl.ac.uk/bartlett/heritage/events/2021/mar/sustainable-heritage-bidecennial-conference-strategic-research-questions>

<sup>8</sup> For all events, the numbers of recording playbacks have been collected on 28/02/2022.

<sup>9</sup> *ibid.*

Links to materials	
Announcement	<a href="https://sshopencloud.eu/events/sshoc-workshop-digitising-museum-objects-using-basic-photogrammetry">https://sshopencloud.eu/events/sshoc-workshop-digitising-museum-objects-using-basic-photogrammetry</a>
Blogpost	<a href="https://sshopencloud.eu/news/sshoc-workshop-notes-digitizing-museum-objects-using-basic-photogrammetry">https://sshopencloud.eu/news/sshoc-workshop-notes-digitizing-museum-objects-using-basic-photogrammetry</a>
Presentation	<a href="https://doi.org/10.5281/zenodo.4776384">https://doi.org/10.5281/zenodo.4776384</a> <a href="https://doi.org/10.5281/zenodo.4776387">https://doi.org/10.5281/zenodo.4776387</a>
Recording	<a href="https://www.youtube.com/watch?v=Tf1MMGc4bGg">https://www.youtube.com/watch?v=Tf1MMGc4bGg</a>
Audience	
Total participants	117
Recording playbacks	71
Targeted audience	Researchers and others interested in the field of imaging applications in the study and management of heritage
By stakeholder category	<p>Researcher: 39</p> <p>University and/or research performing organisation: 38</p> <p>Research library and/or archive: 19</p> <p>Private sector and/or industry player: 11</p> <p>Civil society and/or citizen scientist: 6</p> <p>Policy making organisations: 3</p> <p>Research and e-infrastructure: 1</p>
<b>Organisers</b>	T6.5 in cooperation with UCL
<b>Workshop report</b>	see Annex 3: Workshop report: Digitising Museum Objects Using Basic Photogrammetry

### 3.3 Data Protection and the GDPR

#### 3.3.1 Data Protection in Research Practice

<b>Date and Venue</b>	<p>10/2021</p> <p>University of Zadar, Croatia, and online; co-located with the 1st DARIAH-HR International Conference<sup>10</sup></p>
-----------------------	---

<sup>10</sup> Digital Humanities and Heritage, 1st DARIAH-HR International Conference, <https://dhh.dariah.hr/en/home/>

Links to materials	
Announcement	<a href="https://www.sshopencloud.eu/events/sshoc-workshop-data-protection-research-practice-gdpr-and-eldah-consent-form-wizard">https://www.sshopencloud.eu/events/sshoc-workshop-data-protection-research-practice-gdpr-and-eldah-consent-form-wizard</a>
Blogpost	<a href="https://www.sshopencloud.eu/news/sshoc-workshop-notes-data-protection-research-practice-gdpr-and-eldah-consent-form-wizard">https://www.sshopencloud.eu/news/sshoc-workshop-notes-data-protection-research-practice-gdpr-and-eldah-consent-form-wizard</a>
Presentation	<a href="https://doi.org/10.5281/zenodo.5770207">https://doi.org/10.5281/zenodo.5770207</a>
Recording	<a href="https://www.youtube.com/watch?v=lzP60vKnho0">https://www.youtube.com/watch?v=lzP60vKnho0</a>
Audience	
Total participants	34
Recording playbacks	23
Targeted audience	Researchers interested in ensuring GDPR compliance of their work
By stakeholder category	Researcher: 22 Research library and/or archive: 7 Research and e-infrastructure: 5
<b>Organisers</b>	T6.5 in cooperation with DARIAH and CLARIN ERIC
<b>Workshop report</b>	see Annex 4: Workshop report: Data Protection in Research Practice: The GDPR and the ELDAH Consent Form Wizard

## 3.4 Data Stewardship and RDM

### 3.4.1 Caring for Sharing – Data Management and FAIRness of Migration Data

<b>Date and Venue</b>	03/2020 Brussels, Belgium; co-located with the COST Action 16111 (ETHMIGSURVEYDATA) Work group plenary meetings and 2nd Annual Policy Dialogue Conference <sup>11,12</sup>
Links to materials	

<sup>11</sup> Ethmig Survey Data, <https://ethmigsurveydatahub.eu/>

<sup>12</sup> Agenda Brussels Meeting And Policy Dialogue Conference, <http://ethmigsurveydatahub.eu/wp-content/uploads/2020/03/AGENDA-BRUSSELS-MEETING-AND-POLICY-DIALOGUE-CONFERENCE.pdf>

Announcement	<a href="https://www.sshopencloud.eu/news/sshoc-workshop-caring-sharing-%E2%80%93-data-management-and-fairness-migration-data">https://www.sshopencloud.eu/news/sshoc-workshop-caring-sharing-%E2%80%93-data-management-and-fairness-migration-data</a>
Blogpost	<a href="https://www.sshopencloud.eu/news/caring-sharing-workshop-data-mgmt-and-fairness-migration-data">https://www.sshopencloud.eu/news/caring-sharing-workshop-data-mgmt-and-fairness-migration-data</a>
Presentation	<a href="https://doi.org/10.5281/zenodo.3736355">https://doi.org/10.5281/zenodo.3736355</a>
Recording	/
<b>Audience</b>	
Total participants	33
Recording playbacks	NA
Targeted audience	Researchers, policy-oriented professionals and others interested in quantitative survey research and/or EMM (ethnic and migrant minority) integration
By stakeholder category	Researcher: 21 University and research performing institution: 10 Policy making organisation: 1 Research and e-infrastructure: 1
<b>Organisers</b>	T6.5 in cooperation with CESSDA/UKDS
<b>Workshop report</b>	see Annex 5: Workshop report: Caring for Sharing – Data Management and FAIRness of Migration Data Annex 5: Workshop report: Caring for Sharing – Data Management and FAIRness of Migration Data

### 3.4.2 Copyright Issues in Secondary Data Use

<b>Date and Venue</b>	01/2022 Online
<b>Links to materials</b>	
Announcement	<a href="https://www.sshopencloud.eu/events/sshoc-workshop-data-management-planning-and-overcoming-challenges-social-sciences-data">https://www.sshopencloud.eu/events/sshoc-workshop-data-management-planning-and-overcoming-challenges-social-sciences-data</a>
Blogpost	<a href="https://www.sshopencloud.eu/news/sshoc-workshop-notes-data-management-planning-and-overcoming-challenges-social-sciences-data">https://www.sshopencloud.eu/news/sshoc-workshop-notes-data-management-planning-and-overcoming-challenges-social-sciences-data</a>
Presentation	Day 1: <a href="https://doi.org/10.5281/zenodo.6246973">https://doi.org/10.5281/zenodo.6246973</a> Day 2: <a href="https://doi.org/10.5281/zenodo.6246977">https://doi.org/10.5281/zenodo.6246977</a>
Recording	Day 1: <a href="https://youtu.be/yonxnA56jCg">https://youtu.be/yonxnA56jCg</a>

	Day 2: <a href="https://youtu.be/S-zik_yzo14">https://youtu.be/S-zik_yzo14</a>
<b>Audience</b>	
Total participants	38
Recording playbacks <sup>13</sup>	Day 1: 8 Day 2: 5
Targeted audience	Researchers and others interested in copyright principles
By stakeholder category	Researcher: 16 Research library and/or archive: 12 University and/or research performing organisation: 6 Research and e-infrastructure: 2 Private sector and/or industry player: 1 Policy making organisations: 1
<b>Organisers</b>	T6.5 in cooperation with CESSDA/UKDS and ADP
<b>Workshop report</b>	see Annex 6: Workshop report: Copyright Issues in Secondary Data Use

### 3.4.3 Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing

<b>Date and Venue</b>	02/2022 Online
<b>Links to materials</b>	
Announcement	<a href="https://www.sshopencloud.eu/events/sshoc-workshop-copyright-issues-secondary-data-use">https://www.sshopencloud.eu/events/sshoc-workshop-copyright-issues-secondary-data-use</a>
Blogpost	<a href="https://www.sshopencloud.eu/news/sshoc-workshop-notes-copyright-issues-secondary-data-use">https://www.sshopencloud.eu/news/sshoc-workshop-notes-copyright-issues-secondary-data-use</a>
Presentation	Day 1: <a href="https://doi.org/10.5281/zenodo.5938372">https://doi.org/10.5281/zenodo.5938372</a> Day 2: <a href="https://doi.org/10.5281/zenodo.5938388">https://doi.org/10.5281/zenodo.5938388</a>
Recording	Day 1: <a href="https://youtu.be/6THOdJhJ5ko">https://youtu.be/6THOdJhJ5ko</a> Day 2: <a href="https://youtu.be/Qy5m0F9BX08">https://youtu.be/Qy5m0F9BX08</a>
<b>Audience</b>	

<sup>13</sup> The number of views is lower than for other events, because the dissemination of the materials for this workshops started just before the deadline for this Deliverable.

Total participants	36
Recording playbacks <sup>14</sup>	Day 1: 15 Day 2: 16
Targeted audience	Early career researchers, especially from Social Sciences, and support staff
By stakeholder category	Researcher: 17 University and/or research performing organisation: 8 Research library and/or archive: 6 Research and e-infrastructure: 3 Policy making organisation: 1 Private sector and/or industry player: 1
<b>Organisers</b>	T6.5 in cooperation with CESSDA/UKDS and ADP
<b>Workshop report</b>	see Annex 7: Workshop report: Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing

## 3.5 Data Citation

### 3.5.1 Data Citation in Practice

<b>Date and Venue</b>	06/2021 Online
<b>Links to materials</b>	
Announcement	<a href="https://sshopencloud.eu/events/sshoc-workshop-data-citation-practice">https://sshopencloud.eu/events/sshoc-workshop-data-citation-practice</a>
Blogpost	<a href="https://www.sshopencloud.eu/news/sshoc-workshop-notes-data-citation-practice">https://www.sshopencloud.eu/news/sshoc-workshop-notes-data-citation-practice</a>
Presentation	<a href="https://doi.org/10.5281/zenodo.5140051">https://doi.org/10.5281/zenodo.5140051</a> <a href="https://doi.org/10.5281/zenodo.5140063">https://doi.org/10.5281/zenodo.5140063</a> <a href="https://doi.org/10.5281/zenodo.5138041">https://doi.org/10.5281/zenodo.5138041</a>
Recording	<a href="https://www.youtube.com/watch?v=16JD2ekW_Oo">https://www.youtube.com/watch?v=16JD2ekW_Oo</a>
<b>Audience</b>	
Total participants	36

<sup>14</sup> See footnote 14.

Recording playbacks	58
Targeted audience	Researchers and repository managers as well as anyone other interested in data citation in the SSH domain
By stakeholder category	University and/or research performing organisation: 11 Research and e-infrastructure: 8 Researcher: 7 Research library and/or archive: 7 Private sector and/or industry player: 3
<b>Organisers</b>	T6.5 in cooperation with CLARIN ERIC, SSHOC WP3 and SSHOC WP7
<b>Workshop report</b>	see Annex 8: Workshop report: Data Citation in Practice Annex 8: Workshop report: Data Citation in Practice

## 3.6 Text Mining for the SSH

### 3.6.1 Using Corpora for Implementing Validation

<b>Date and Venue</b>	09/2019 Leipzig, Germany; co-located with the CLARIN Annual Conference 2019 <sup>15</sup>
<b>Links to materials</b>	
Announcement	<a href="https://www.sshopencloud.eu/events/sshoc-workshop-using-corpora-implementing-validation-workflows-combine-quantity-and-quality">https://www.sshopencloud.eu/events/sshoc-workshop-using-corpora-implementing-validation-workflows-combine-quantity-and-quality</a>
Blogpost	<a href="https://www.sshopencloud.eu/news/using-corpora-implementing-validation-sshoc-masterclass">https://www.sshopencloud.eu/news/using-corpora-implementing-validation-sshoc-masterclass</a>
Presentation	<a href="https://polmine.github.io/ValidationWorkflows/">https://polmine.github.io/ValidationWorkflows/</a>
Recording	NA
<b>Audience</b>	
Total participants	4
Recording playbacks	NA

<sup>15</sup> CLARIN ERIC, CLARIN Annual Conference 2019 in Leipzig, Germany, <https://www.clarin.eu/event/2019/clarin-annual-conference-2019-leipzig-germany>



<b>Targeted audience</b>	Researchers and others, especially from political and social sciences, interested in analysing large text collections
<b>By stakeholder category</b>	Research library and/or archive: 2 Research and e-infrastructure: 1 Researcher: 1
<b>Organisers</b>	T6.5 in cooperation with CLARIN ERIC
<b>Workshop report</b>	see Annex 9: Workshop report: Using Corpora for Implementing Validation

### 3.6.2 Exploration of Society Through the Lens of Labour Market Related Documentation

<b>Date and Venue</b>	05/2021 Online; co-located with the Helsinki Digital Humanities Hackathon <sup>16</sup>
<b>Links to materials</b>	
Announcement	<a href="https://sshopencloud.eu/events/sshoc-workshop-exploration-society-through-lens-labour-market-related-documentation">https://sshopencloud.eu/events/sshoc-workshop-exploration-society-through-lens-labour-market-related-documentation</a>
Blogpost	<a href="https://sshopencloud.eu/news/sshoc-workshop-exploration-society-through-lens-labour-market-related-documentation-%E2%80%93-cbaquest">https://sshopencloud.eu/news/sshoc-workshop-exploration-society-through-lens-labour-market-related-documentation-%E2%80%93-cbaquest</a>
Presentation	/
Recording	/
<b>Audience</b>	
Total participants	65 <sup>17</sup>
Recording playbacks	NA

<sup>16</sup>Helsinki Centre for Digital Humanities, Helsinki Digital Humanities Hackathon 2021 #DH-H21, <https://www2.helsinki.fi/en/helsinki-centre-for-digital-humanities/helsinki-digital-humanities-hackathon-2021-dhh21>

<sup>17</sup> This is the total count of the audience reached. The group that worked specifically on the research problem of this workshop included 13 participants. To read more about the specifics of the event, please see Annex 10: Workshop report: Exploration of Society Through the Lens of Labour Market Related Documentation – CBAQuest and ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates.

Targeted audience	Researchers and students from computer science, data science, humanities and social sciences
By stakeholder category	Researcher: 65
<b>Organisers</b>	T6.5 in cooperation with CLARIN ERIC
<b>Workshop report</b>	see Annex 10: Workshop report: Exploration of Society Through the Lens of Labour Market Related Documentation – CBAQuest and ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates

### 3.6.3 ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates

<b>Date and Venue</b>	05/2021 Online; co-located with the Helsinki Digital Humanities Hackathon <sup>18</sup>
<b>Links to materials</b>	
Announcement	<a href="https://sshopencloud.eu/events/sshoc-workshop-par%C2%ADlamint-%E2%80%93-exploring-societal-issues-through-com%C2%ADpar%C2%ADable-cor%C2%ADpora-par%C2%ADlia">https://sshopencloud.eu/events/sshoc-workshop-par%C2%ADlamint-%E2%80%93-exploring-societal-issues-through-com%C2%ADpar%C2%ADable-cor%C2%ADpora-par%C2%ADlia</a>
Blogpost	<a href="https://sshopencloud.eu/news/sshoc-workshop-notes-par%C2%ADlamint-%E2%80%93-exploring-societal-issues-through-com%C2%ADpar%C2%ADable-cor%C2%ADpora-par">https://sshopencloud.eu/news/sshoc-workshop-notes-par%C2%ADlamint-%E2%80%93-exploring-societal-issues-through-com%C2%ADpar%C2%ADable-cor%C2%ADpora-par</a>
Presentation	/
Recording	/
<b>Audience</b>	
Total participants	65 <sup>19,20</sup>

<sup>18</sup> Helsinki Centre for Digital Humanities, Helsinki Digital Humanities Hackathon 2021 #DH-H21, <https://www2.helsinki.fi/en/helsinki-centre-for-digital-humanities/helsinki-digital-humanities-hackathon-2021-dhh21>

<sup>19</sup> This is the total count of the audience reached. The group that worked specifically on the research problem of this workshop included 6 participants. To read more about the specifics of the event, please see

<sup>20</sup> The number of participants for this event is not included in the final count, since the two events described in Section 3.6.2 Exploration of Society Through the Lens of Labour Market Related Documentation and Section 3.6.3 ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates were part of the same event and thus reached the same audience.

Recording playbacks	NA
Targeted audience	Researchers and students from computer science, data science, humanities and social sciences
By stakeholder category	Researcher: 65
<b>Organisers</b>	T6.5 in cooperation with CLARIN ERIC
<b>Workshop report</b>	see Annex 10: Workshop report: Exploration of Society Through the Lens of Labour Market Related Documentation – CBAQuest and ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates <sup>21</sup>

## 4. Outcomes and Conclusions

Targeted training events represent a crucial contribution to the efforts made by SSHOC members to promote the achievements of the project and transfer invaluable skills and knowledge to the communities. The general objective of targeted training activities was to ensure sustainable use of SSHOC outcomes in the future and to encourage data-driven interdisciplinary research approaches. Altogether, SSHOC Task 6.5 delivered twenty training events: the initial set of twelve training events (six workshops and six webinars) which covered a broad array of topics was expanded with five additional workshops and three additional webinars addressing specific training needs identified by the SSHOC community and/or showcasing tools and services developed within the SSHOC project.

ONE EXAMPLE IS THE WORKSHOP *DATA CITATION IN PRACTICE* (SEE SECTION 3.5.1 Data Citation in Practice FOR MORE) SHOWCASING THE WORK OF SSHOC WP3 WHICH FOCUSED ON MAKING DATA CITATIONS MACHINE ACTIONABLE. THE WORKSHOPS WERE SUCCESSFUL IN ESTABLISHING CONNECTIONS INSIDE THE SSHOC COMMUNITY AND BEYOND: FIRST, BY INVITING SPEAKERS FROM OTHER PROJECTS, INITIATIVES AND INSTITUTIONS, AND SECOND, BY BEING CO-LOCATED WITH MAJOR SSH EVENTS. A GREAT EXAMPLE OF IN-HOUSE COOPERATION WAS THE WORKSHOP *THE CASE OF INTERVIEW DATA* DELIVERED BY EXPERTS FROM WP4 (SEE SECTION 3.1.1 The Case Of Interview Data: A Multidisciplinary Approach to the Use of Technology in Research Using

<sup>21</sup> Since the two workshops *Exploration of Society Through the Lens of Labour Market Related Documentation – CBAQuest* (Section 3.6.2 Exploration of Society Through the Lens of Labour Market Related Documentation) and *ParlaMint – exploring societal issues through comparable corpora of parliamentary debates* (Section 3.6.3 ParlaMint – Exploring Societal Issues Through Comparable Corpora of Parliamentary Debates) were part of the same host event, they are described in the same detailed report outlining their respective characteristics where relevant.

Interview Methods FOR MORE). EXTERNAL COLLABORATION PROVED VALUABLE, FOR EXAMPLE, FOR THE WORKSHOP ON *DIGITISING MUSEUM OBJECTS USING BASIC PHOTOGRAMMETRY* (SEE SECTION 3.2.2 DIGITISING MUSEUM OBJECTS USING BASIC PHOTOGRAMMETRY for more) which featured one of the most visible experts in the field, Kira Zumkley (AHFAP,<sup>22</sup> Victoria and Albert Museum). This workshop is also just one of the examples of successful collaborations with major hosting events, in this case the Sustainable Heritage Bidecennial Conference<sup>23</sup>.

The fact that most workshops were fully booked confirms that the topics and speakers were highly relevant. Lively discussions as well as the positive participants' feedback prove that the presentations and demonstrations were valuable and thought provoking. This is further supported by the fact that on several occasions, participants or speakers expressed the need for follow-ups in the form of workshops or webinars. When possible, the event organisers catered to these needs, e.g., *Sharing Datasets of Pathological Speech* (see D6.15, Section 3.1.2) was an additional follow-up webinar to the workshop *The Case Of Interview Data* (see Section 3.1.1 THE CASE OF INTERVIEW DATA: A MULTIDISCIPLINARY APPROACH TO THE USE OF TECHNOLOGY IN RESEARCH USING INTERVIEW METHODS). It was organised as a response to the needs of a specific, but prominent research community studying communication disorders, but also to highlight the work done in SSHOC T5.4. Similarly, on the initiative of the speakers and participants, four Q&A sessions<sup>24</sup> were organised after the webinar *Hands-on Tutorial on Transcribing Interview Data* (see D6.15, Section 3.1.1) in order to offer support in using the platform showcased in the webinar.

THE ORGANISERS' EXPERIENCE SHOWS THAT WORKSHOPS OFFER A VALUABLE FORM OF TRAINING SINCE THEY ARE ESPECIALLY DESIGNED TO ENABLE THE TRANSFER OF BOTH INFORMATION AND SKILLS. THUS, THEY HELP PARTICIPANTS TO IMMEDIATELY PUT THEIR NEWLY GAINED KNOWLEDGE TO USE. THIS GUARANTEES TO A CERTAIN EXTENT THAT THE WORKSHOP TOGETHER WITH RELATED SSHOC RESULTS WILL HAVE A CONCRETE IMPACT ON THE PARTICIPANTS' WORK. WHILE PUTTING MUCH EMPHASIS ON HANDS-ON ACTIVITIES AND SOLVING CONCRETE CHALLENGES COULD LEAD TO A TUNNEL-VISION DESIGN OF THE PROGRAMME, THE SET OF WORKSHOPS SHOWS THAT THIS TRAINING FORMAT CAN BE EQUALLY VALUABLE FOR ADDRESSING GENERIC AS WELL AS TOPIC-SPECIFIC ISSUES. EXAMPLES OF SUCH TRAININGS ARE, ON ONE HAND, THE TOPIC GENERIC WORKSHOP *COPYRIGHT ISSUES IN SECONDARY DATA USE* (SEE SECTION 3.4.2 COPYRIGHT ISSUES IN SECONDARY DATA USE for more) applicable to anyone working with secondary data regardless of their scientific field. On the other hand, the highly specific and time-demanding workshop *ParlaMint – exploring societal issues through comparable corpora of parliamentary debates* (see Section 3.6.3 PARLAMINT – EXPLORING SOCIETAL ISSUES THROUGH COMPARABLE CORPORA OF PARLIAMENTARY DEBATES for more) which was a ten day intense workshop in the

<sup>22</sup> Association for Historical and Fine Art Photography's (AHFAP), <https://ahfap.org.uk/>

<sup>23</sup> UCL, Sustainable Heritage Bidecennial Conference: Strategic Research Questions, <https://www.ucl.ac.uk/bartlett/heritage/events/2021/mar/sustainable-heritage-bidecennial-conference-strategic-research-questions>

<sup>24</sup> SSHOC, SSHOC OH-Portal Questions & Answers Session 4, <https://www.sshopencloud.eu/sshoc-oh-portal-questions-answers-session-4>

framework of the Helsinki Digital Humanities Hackathon, connecting an interdisciplinary team of researchers from Computer Science, Sociology, Political Studies and alike.

The attendance results show that training events (webinars and workshops) reached a large and diverse audience. In total, twenty live (streaming) events (11 workshops, 9 webinars) have been attended by 1102 participants (464 at workshops, 638 at webinars). On average, each training event has been attended by almost 60 participants who represented 23 countries and 6 SSHOC stakeholder categories. The most prominent groups (accounting for around 85% of the audience) were those identified as key target audiences for training workshops, namely: *Researchers, research networks and communities; Research libraries and archives; Universities and research performing institutions*. It should be noted that the outreach is even bigger if we take into account the views of the event recordings published on the SSHOC YT channel. These account for another 1353 views of either webinar (1127 views) or workshop recordings (226 views), thus giving a final count of 2455 individuals that have been reached by T6.5 targeted training activities. Given that these recordings are freely accessible, the outreach is likely to extend even further in the future.

Specifically for workshops, the live (streaming) events were attended by 464 individuals. On average, there were 45 participants at each workshop who represented 22 countries and 5 stakeholder categories. Together with the playbacks of the workshop recordings when available (226), the total outreach includes 690 people. This confirms that the workshops were successfully organised and delivered since the Key Performance Indicators (Torma et al., 2019, p. 21) were mainly reached or exceeded: the attendance numbers were above the set minimum of 25 per workshop for all of the events except for two (see Section 3.1.1 THE CASE OF INTERVIEW DATA: A MULTIDISCIPLINARY APPROACH TO THE USE OF TECHNOLOGY IN RESEARCH USING INTERVIEW METHODS and 3.6.1 USING CORPORA FOR IMPLEMENTING VALIDATION for more); the distribution in terms of geographical location and stakeholder groups was very diverse; the three key target groups (researchers, research performing institutions, and research libraries) represented a great majority of the audiences; the webinars were equally distributed across the project's timeframe; and the final workshop count (11) was almost twice the required one (6 workshops required in the Grant Agreement).

Lastly, due to migration to a virtual environment, most of the targeted training workshops as well as all webinars resulted in a rich set of published online materials (blogposts, presentations, and recordings). That extends the positive impact of the training events beyond the actual live stream and continues to promote the vision of the SSHOC project after its formal conclusion. The co-location of the workshops with renowned conferences and summer schools also contributed to additional visibility of SSHOC. Apart from the training workshops themselves, when relevant, the evaluation of the organisational processes also resulted in a valuable exchange of experience among SSHOC members—especially with regard to best practices for organising impactful training workshops, especially when they need to be delivered in an online or hybrid format (e.g., hybrid does not mean simple streaming of the onsite happening; audio/video quality is of crucial importance; ice-breakers are not just for onsite events, and programmes needs a special adaptation for the onsite and online audience).

The results show that both training workshops and webinars received a very good reception among the SSH community. Undoubtedly, this could not be achieved neither without highly dedicated speakers nor without eager participants. The speakers showed a lot of flexibility in delivering their expertise in new, online environments, as well as a lot of selfless support by offering help even after the events. The participants showed a lot of perseverance and patience in order to attend mostly online events and fully engage in the activities—despite experiencing already a very high amount of screen time. The workshops, being a combination of presentations and hands-on exercises, proved as a valuable training method. They enabled immediate application of the newly gained knowledge and clarification of potential issues. Although the Covid-19 pandemic and subsequent migration of training activities to virtual space shrunk the networking potential of workshops, programme modifications ensured that the training remained engaging and participant-friendly. It should be added, however, that the organisers failed to specifically cater to the needs of persons with disabilities (e.g., by providing sign language interpretation, subtitles (with few exceptions where auto-generated captions in Zoom were used), or adapted materials). This is an important aspect to be considered for any future events. In general, it seems that the training support in the SSH for persons with disabilities is still largely neglected. Best practices for developing inclusive training materials are lacking, so the community needs to strive in its future efforts to develop sustainable guidelines that would provide accessible content to a variety of users. The work performed by SSHOC on the language and speech technologies undoubtedly provide an excellent basis for future steps in the direction of greater inclusiveness. The pandemic-related situation also brought certain advantages. In particular, the online training events could also be attended by participants who would otherwise miss the opportunity due to travel restrictions and other limitations (e.g., incompatible time zones, work obligations, disabilities, or time constraints related to childcare or care for ill relatives). Moreover, the published online materials continue to provide useful training content to interested individuals, including persons with disabilities who can process them with adapted software.

Therefore, it can be concluded that the training events importantly contributed to the SSHOC efforts to promote maximal reuse of tools, services and data in line with the Open Science and FAIR principles. The attendance numbers and the participants' feedback confirm that the events addressed a variety of pertinent training topics which were intertwined with the work done in SSHOC, successfully engaged with diverse user communities, attracted a high number of participants and established collaborations between speakers that extend beyond SSHOC. By publishing most of the materials online, the concerted efforts of Task 6.5 and their partners have also resulted in an extensive set of valuable and openly accessible training content for future use.

## 5. Reference list

Pahor de Maiti, K. & Fišer, D. (2022). SSHOC D6.15 Report on Training Workshops. Zenodo.

Schwabe, A., Ausserhofer, J., Marino, L., Willems, M., Kalaitzi, V., Vipavc Brvar, I., Smith, E., & Muscella, S. (2019). *SSHOC D2.1 Overall Communication and Outreach Plan (approved 18 Nov 2019)*. Zenodo.  
<https://doi.org/10.5281/zenodo.3595936>

Torma, M., Kalaitzi, V., Dijk, E., Wittenberg, M., Willems, M., Fišer, D., Pahor de Maiti, K., Durco, M. & Vipavc Brvar, I. (2019). SSHOC D6.2 Building Expertise Strategy (v1.0). Zenodo.  
<https://doi.org/10.5281/zenodo.4558294>

## 6. Annexes

---

# ANNEX 1: WORKSHOP REPORT: THE CASE OF INTERVIEW DATA – A MULTIDISCIPLINARY APPROACH TO THE USE OF TECHNOLOGY IN RESEARCH

### Background

This report concerns the first workshop organised by T4.4 in cooperation with T6.5 of the SSHOC project on the pre-defined topic of Data Science for the Social Science and Humanities. The event — [The Case of Interview Data – A Multidisciplinary Approach to the Use of Technology in Research](#) — was held on 6 July 2019 at the [Digital Humanities 2019 conference](#) in Utrecht, the Netherlands.

This workshop addressed the data-creation phase of the data lifecycle within the Social Sciences and Humanities, which is one of the major goals of the SSHOC project and specifically of Work Package 4 (WP4). SSHOC partners, who also spoke at the workshop, are collecting, processing and analysing audio data in the form of voice-recorded interviews with automatic speech recognition. These partners are developing a survey module capable of integrating audio recordings and their processing into the traditional data collection process.

### Workshop Overview & Format

**Aim.** Interviews as a data type are dealt with various scholarly approaches. These are usually discipline specific. Researchers are sometimes reluctant to try new methods and tools, despite the potential benefits for their research, because they fear a considerable time investment to acquire the required technical skills. The workshop organisers believe that this attitude can be changed for the benefit of the broader scientific community by offering tailored information about the available technological resources and by sharing hands-on experience.

The workshop on interview data was therefore dedicated to presenting the latest developments of the [Transcription chain](#) (T-Chain – combining various open source tools to support transcription and alignment of audio (oral) and text (written) in various languages) and new NLP tools to process the resulting texts.



**Speakers.** The workshop was delivered by five speakers. Three of them are part of the [Oral History & Technology research group](#):

- Christoph Draxler (Ludwig-Maximilians-Universität München),
- Stef Scagliola (C2DH, University of Luxemburg),
- Louise Corti (UK Data Archive).

The other two are external colleagues of the research group:

- Jeannine Beeken (UK Data Archive),
- Khiet Truong (University of Twente).

**Organisers.** The workshop was organised in cooperation with the partners in T6.5 and members of the Oral History & Technology research group, participating in T4.4 of the SSHOC project: Arjan van Hessen (Utrecht University), Christoph Draxler (Ludwig-Maximilians-Universität München), Stef Scagliola (C2DH, University of Luxemburg), Louise Corti (UK Data Archive), Norah Karrouche (Erasmus University Rotterdam), Silvia Calamai (Università di Siena, sede Arezzo), Henk van den Heuvel (Radboud University Nijmegen).

**Participants.** The workshop was attended by 21 participants: 12 from Europe, six from Africa, two from the U.S. and one from Israel. They represented a great variety of disciplines from (oral) history, linguistics and the field of language and speech technology to sociology and psychology, including psycholinguistics, mental health studies and the field of social signal processing. They are part of three stakeholder categories as defined in D6.1: researchers, research networks and communities; research libraries and archives; private sector and industry players.

**Brief summary of the event structure.** The half-day workshop was divided into four sessions. It featured both lectures and hands-on activities, and concluded with an open discussion between participants and speakers.

The workshop covered the following topics:

1. conversion of the AV-material into a suitable format, use of automatic speech recognition via the [OH-portal](#) (the prototype of the T-Chain), correction of the automatic speech recognition (ASR) results and the annotation of the resulting text;
2. linguistic text-analysis including the creation of results visualizations;
3. emotion extraction with the [Open Smile](#) tool.

## Presentations & Discussions: Key Points

**First Session.** ‘Digital Humanities approaches to interview data - can historians, linguists and social scientists share tools?’

**Speakers.** [Louise Corti](#) and [Stef Scagliola](#)

**Main points.** Louise Corti and Stef Scagliola started out by explaining why interviews are so important for oral historians. Next they proceeded to the role of interview data in the work and methodologies of social scientists, computational linguists, social linguists, and affective social signs processing scholars. Thus they made the point that interview data as source of research spans a wide field of scientific disciplines.

**Links to materials.** [Presentations](#)

---

**Second Session.** Preparing your audio-data, uploading audio to the portal and automatically recognizing the speech. Correcting the ASR-results. Downloading the (corrected) results and improving the readability.

**Speaker.** [Christoph Draxler](#)

**Main points.** The speaker elaborated on the preparation of audio data, automatic speech recognition (ASR), ASR results correction and possible steps for improved readability. He then demonstrated the OH-Portal (v1.0.2) which is the prototype of the T-Chain: it is easy to use, the different steps are clear and the final results/outputs are easy to download. He also shared recommendations and comments for a successful use of the portal:

- There is no free lunch. You will have to proper work to obtain good results
- Split long interviews in meaningful chunks of 5-15 minutes.
- With good audio quality and a good recogniser, you can transcribe more quickly with ASR and manual correction than through only manual transcription.
- The full Transcription Chain from upload to word alignment is in operation.
- Users have a 24-hour window to download files before the OH portal removes the files.
- There is also a conversion tool for various audio formats at the Bavarian Archive for Speech Signals called AudioEnhance.

**Links to materials.**

- [Short paper about the Transcription Chain](#)
  - [Presentations](#)
- 

**Third Session.** Introducing linguistic analysis of text: free tools

**Speaker.** [Jeannine Beeken](#)

**Main points.** The speaker pointed to three key challenges for linguistic tools analysing human communication:

1. Separate/Disambiguate: by using spelling correctors, punctuation, and tools for grammatical analysis,
2. Cluster/Group words: by bringing them back to their dictionary form, grouping synonyms and idiomatic expressions,

3. Control/ extract information: by locating stopwords, keywords and using text summarizers.

Beeken made interesting and relevant observations on punctuation. She noted that a standard speech recogniser does not output punctuation, despite punctuation's essential role in dividing a stream of words into meaningful entities. As an example of how punctuation (in this case the comma) makes an essential difference, she gave the following example:

*The panda eats shoots and leaves* (a reference to the food eaten by the panda) vs. *The panda eats, shoots and leaves* (a killer panda!)

Next, she briefly pointed to a [keyword extractor](#), a [summarisation tool](#) and offered a short hands-on activity on her demo corpus using the [SketchEngine](#). The participants were able to carry out a number of tasks such as keyword extraction based on frequency, token/types, lemmas. In addition, they worked on the concordance option to see words in context (e.g., to disambiguate them) and the thesaurus (to find synonyms and similar words belonging to the same semantic cluster).

**Links to materials.** [Presentations](#)

---

#### **Fourth Session.** Emotion extraction with Open Smile

**Speaker.** [Khiet Truong](#)

**Main points.** The speaker, who presented remotely, presented one of the most surprising dimensions of analysing a dialogue between interviewer and interviewee. She demonstrated a simple picture and asked the participants to share their thoughts about the activity in the picture.



The immediate observations of participants varied. This showed that it is easy to make assumptions, yet these can seriously colour the interpretation of research results. In a similar way, when a researcher reads an oral history but does not listen to it, they are missing emotions that may underpin the conversation. A social signal (or emotion) can be a complex installation of behavioural cues. Studying social-sign processing opens up the option of re-interpreting an interview (e. g., by reflecting on the function of the silence, or tone and whether they occur as a generic or a specific feature of

communication within a corpus /collection of interviews). The speaker also showed how she works on tools for automatic emotion recognition directly from the speech signal.

**Links to materials.** [Presentations](#)

---

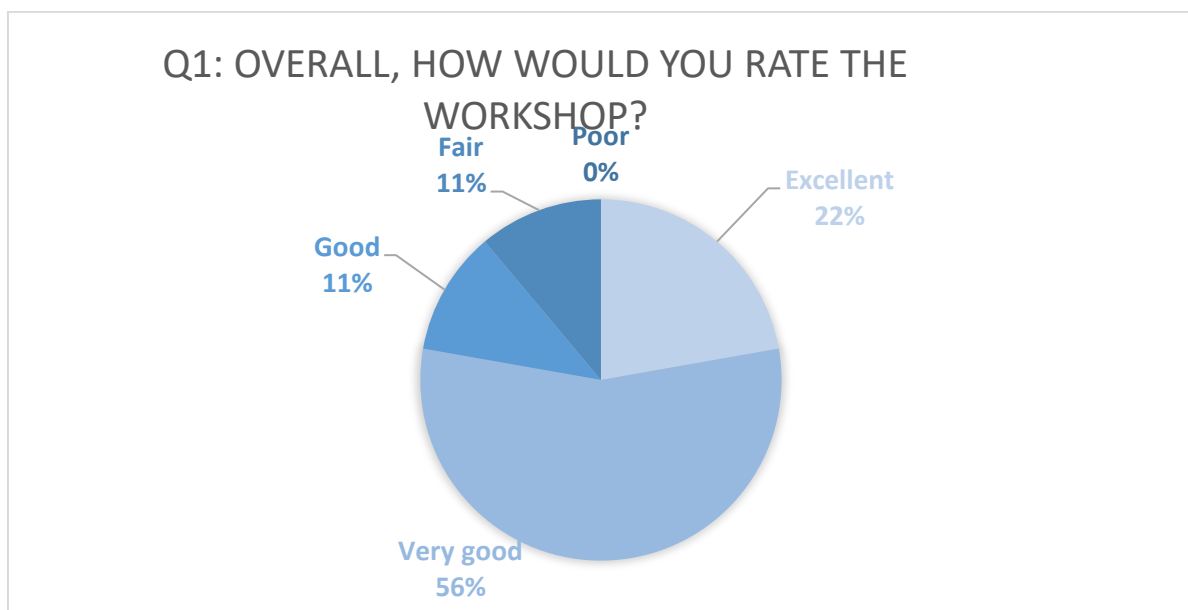
#### **Fourth session.** Open Discussion

**Main points.** The workshop concluded with a discussion. This proved to be a very fruitful and valuable element of the workshop both for the organisers and participants. Many questions were asked, for example, about the quality of ASR for regional or dialectal forms of speech and the data protection after the uploading to the server. Participants also expressed requests for clear guidelines on how to use the tools, for recognisers that can cope with multilingual conversations/data, for easy to read text output formats that show the contributions of the various speakers, etc.

## Outcomes & Feedback

**Participant satisfaction.** Three-quarters of participants rated the workshop Excellent or Very Good, and many asked for an extended version of the workshop in order to have more time for discussion with lecturers and hands-on activities (cf. Q1, 4, 6 & 7). Several participants said the workshop would have a positive impact on their future work, as can be shown from the following answer from the Respondent No. 2 (cf. Q5):

*»/.../ I have been able to learn new tools and theories in digital scholarship and have started applying them to my work«.*



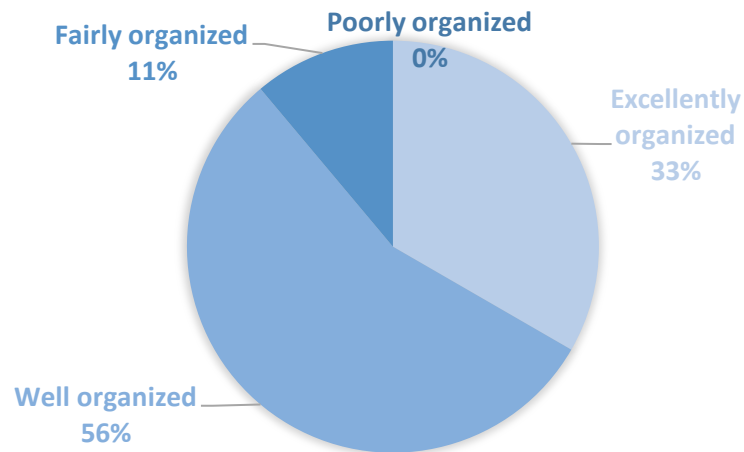


**Feedback regarding the organisation.** Participants generally felt the workshop was well organised (cf. Q9). However, they would prefer a longer workshop and wished for a better venue as expressed by Respondent No. 6 (cf. Q10):

*»The room was not really adapted to such a workshop. Unless I didn't see them, there were no power outlets, no room on the desks to rest one's laptop while taking manual notes if needed etc... But thanks to the warm attitude of the team and the short length of the workshop, it was ok. Thanks to you all.«*

Furthermore, the participants' answers show that more advertisement could be done through SSHOC-related channels as none of the participants found out about the workshop via those channels (cf. Q2). The majority of the participants learned about the workshop on the website of the DH conference where this SSHOC workshop was co-located.

### Q9: HOW WELL ORGANIZED WAS THE WORKSHOP (IN TERMS OF TIME MANAGEMENT, LENGTH, VENUE)?



**Future work.** Interview data is a category that is used in many research disciplines, and allows great knowledge transfers. Diverse research interests of the workshop participants confirm this hypothesis and their feedback shows that the T-Chain meets the needs of many researchers in the various disciplines. For this reason, the Oral History & Technology research group plans new workshops in the near future. In addition, the SSHOC community will be also able to benefit from a webinar on the use of the T-Chain which is planned in the first half of 2020.

---

## ANNEX 2: WORKSHOP REPORT: CITIZEN SCIENCE & CULTURAL HERITAGE. PLANNING FOR SUCCESS

### Background

The workshop —[Citizen Science & Cultural Heritage. Planning for success](#)— held on 24<sup>th</sup> March 2021, was organised by SSHOC T6.5 in cooperation with [UCL ISH](#) as a follow up to a successful webinar delivered in spring 2020 titled [Use and Re-use of Scientific Data in Archaeology and Heritage](#). The workshop enabled knowledge transfer on the topic of participative research which is highly relevant and increasingly popular yet still largely unknown in the Heritage Science domain as well as more broadly.

The workshop contributed to SSHOC objectives by focusing on the promotion of best practices in participative research and highlighting the importance of considering each and all of the steps in the data lifecycle. By presenting use cases tailored to heritage science researchers, it directly targeted this specific yet vast community of scholars, while at the same time reaching out to other domains by pinpointing the key aspects to be considered in citizen science projects.

### Workshop Overview & Format

**Aim.** Crowdsourcing of data and participative research is becoming much more accessible thanks to advances in digital technology and data science. In reality, however, citizen science projects are highly complex and require significant resources to be successful. This methodological approach is becoming increasingly popular, especially in cultural heritage where opportunities for participative data are vast. However, this approach is not without dangerous pitfalls for the success of the project. Therefore, the purpose of this workshop was to offer the participants an interactive session to discuss best practices in conducting participative research while simultaneously providing clear guidelines how to plan a citizen science project in order to ensure quality of data and safety of the participants. This was supported with the use cases which were focused on cultural heritage: one involving data collection and one involving data tagging.

**Speakers.** The workshop was delivered by 3 speakers:

- [Josep Grau-Bové](#) (UCL): a lecturer in Science and Engineering in Arts, Heritage and Archaeology. His work explores the interface where technology meets preventive conservation. Josep is interested in how new computational developments can support preventive conservation management and is currently working on applying information modelling, citizen science, and system dynamics in heritage science.

- [Rosie Bringham](#) (UCL ISH): a software engineer and researcher at the Institute for Sustainable Heritage, (ISH) University College London. She is the founder of Monument Monitor, a ground-breaking citizen heritage science project that collects conservation data from visitor photographs. Her PhD explores how artificial intelligence can improve conservation management processes and her previous publications have investigated the reliability data provided by citizen scientists.
- [Alejandra Albuerne](#) (UCL, SSHOC T6.5): SSHOC representative and moderator

**Organisers.** The workshop was organised in cooperation with partners in T6.5 of the SSHOC project and [UCL ISH](#), which was the principal organiser of the [Sustainable Heritage Bidecennial Conference](#) that co-hosted this SSHOC workshop.

**Participation.** There were 80 participants of the workshop.<sup>25</sup> The majority of them (75%) came from European countries, both EU and non-EU, while the remaining 25% represented countries on other continents (China, Colombia, Egypt, India, the USA, etc.). The audience of the livestream covered all eight stakeholder categories identified in D6.1. The biggest portion of the audience (75%) represented the broad scientific community (categories “Researchers, Research Networks and Communities”, “Universities and Research Performing Institutions”, “Research Libraries and Archives” and “Research and e-infrastructures” in that order). “Private Sector and Industry Players” category represented 10% of the viewers, “Civil Society and Citizen Scientists” and “Policy Making Organisations” around 7% each, and “Research Funding Organisations” 2% of the audience.

**Brief summary of the event structure.** Due to the Covid-19 pandemic, the conference [Sustainable Heritage Bidecennial Conference](#) which co-hosted this workshop and consequently the workshop itself were held online. The workshop lasted for an hour and was divided into three parts: two shorter ones dedicated to the presentation of SSHOC and to UCL ISH work and participative research in general; and a longer presentation delivered by the lead presenter Rosie Bringham on the main topic of the workshop with specific use cases. The event concluded with a thought-provoking discussion.

## Presentations & Discussions: Key Points

**First Session.** Introduction to SSHOC

**Speakers.** [Alejandra Albuerne](#)

**Main points.** The workshop was given in the framework of the SSHOC project. The speaker presented the goals and the expected impacts of the SSHOC project and situated the workshop in relation to them.

---

<sup>25</sup> The number consists only of viewers of the livestream. It should be noted that the overall reach is bigger since this number does not include the views of the workshop recording which is available on the SSHOC YouTube channel and promoted through SSHOC and its partners.



**Links to materials.**

- [Workshop recording](#) (min: 0:00–9:30)
- 

**Second Session.** A Quick Introduction to Citizen Science in the Institute for Sustainable Heritage.

**Speaker.** [Josep Grau-Bove](#)

**Main points.** UCL ISH has considerable experience in performing citizen science, and despite its high appeal for cultural heritage domain, the presenter was clear that participative research approach can be exploited by many other disciplines. In fact, there are several types of citizen science models that vary with regard to project's aim, resource availability, the level of participants' skills required, etc. More generally, citizen science projects can serve two distinct purposes that can overlap to various degrees: educational and scientific purpose, and it is precisely at this intersection that citizen science projects establish their specific added-value and importance for the scientific community and our contemporary society.

**Links to materials.**

- [Workshop recording](#) (min: 9:31–16:45)
- 

**Third Session.** Citizen Science & Cultural Heritage. Planning for success

**Speaker.** [Rosie Bringham](#)

**Main points.** Rosie Bringham presented two distinct projects in order to showcase two different applications of crowdsourcing and highlight the general characteristics of any citizen science project. The two projects presented have very different objectives and requirements, and thus employ different approaches in order to promote community engagement and participation. The first project, *Plant Monitoring at Angkor Wat*, aims to monitor the physical change of a historic site and is focused on data collection, while the second, *Digitizing the Naga 'Queen' diaries*, aims to digitize a collection of hand-written texts and make it available online, and is thus focused on data tagging.

The key difference between a citizen science project and other research projects is that a successful citizen science research project needs to be engaging for the community, both at an intellectual and practical levels. The essential step in the planning stages of a crowdsourcing project is community scoping. So, the researcher should start with answering the first question: considering what data you need and how complex this data might be to obtain, whom should you engage?

**Plant Monitoring at Angkor Wat – citizen science and data collection.** Monitoring physical change through time can be effectively done by processing simple imaging data: photographs. In popular cultural heritage sites, visitor photographs are a rich data source. Once it is established that there is potential for a citizen science project given (1) the appropriateness of the task and (2) existence/availability of a relevant and interested community, the second question that needs to be answered is what are the key considerations in planning crowdsourcing data collection. Workshop participants contributed to compiling a list of wide-ranging factors, including how to mobilise participation; collect photograph

settings and metadata; ensure the entire site is covered and not only the popular areas; develop privacy policies and image processing approaches of people featured in photographs. The discussion highlighted the complexity of the task. From a technical perspective, an essential step that needs to be fully planned and scoped before the project begins, is planning end user access to the sourced data. Data storage needs to be planned and budgeted from the start, considering possible duration of the project and relevance of the data. Data transfer from collectors to users also deserves careful consideration: while social media might be considered a readily accessible means where engagement can be channelled, most social networks compress images and strip their metadata, possibly rendering them unusable for the desired analysis. Last but not least, contributor perspective should also be considered. It is important to clearly define the requirements of the images that are collected for a project addressing the questions how specific does the data need to be, and how will this be communicated to participants.

**Digitizing the Naga ‘Queen’ diaries – citizen science and data tagging.** Crowdsourcing can also be effective in research that requires the processing of large amounts of data, where that processing needs a degree of human validation, such as in the digitizing of hand-written texts. Workshop participants contributed to identifying the key issues to consider in a data tagging project linked to the processing of personal diaries and correspondence. Again, a broad range of challenges was discussed, from legibility of the texts to ethical issues arising from the contents or authenticity concerns. It was stressed that the end use of the digitized documents needs to be clearly understood when planning the crowdsourced data tagging exercise. Data augmentation and searchability are two essential factors. To this end, keywords can be considered as part of the participative process. Researchers can help themselves by using platforms designed for citizen science, like [Zooniverse](#), that offer capabilities for enhancing searchability. A level of technical literacy is, however, required to process the data output from these platforms.

**The challenge of community engagement.** A key difference between the two case studies discussed in this workshop is the need for participant skills and level of engagement. While a data collection project can be streamlined to arrive at a simple yet effective mode of participation, a more complex data tagging task might require more significant contributions from individual participants, often including some training and input validation. These differences will impose different strategies for community engagement, and different challenges that go beyond the technical. Community engagement in heritage science projects is an exciting and effective way of making participants aware of the conservation and management challenges of heritage sites. The core motivations for participants is often the opportunity of participating in scientific research and contributing actively to the conservation of a heritage site or asset. Participation of communities comes with responsibilities for researchers: it should be conceptualised as a two-way relationship between contributor and researcher, with regular feedback and accountability on how participation has served the research. This meets not only ethical principles, but can help maintain engagement throughout the project, which is an essential element for success.

#### Links to materials.

- [Presentation slides](#)
- [Workshop recording](#) (min: 16:46–1:02:05)

**Fourth Session. Q&A Session.****Speaker.** [Rosie Brigham](#) and [Josep Grau-Bove](#)

**Main points.** Some questions addressed specific, technical aspects of collecting data, such as how to work with younger data providers or how to direct data providers to one exact spot that one would like to observe, and the solution for this is to be able to communicate appropriately the content and the needs of the project taking into account the educational level of the community engaged. Providing community with detailed guidelines that show what is interesting for scientists and why, and what is not, is also essential.

Other questions touched upon the conceptual level of citizen science: what are the actual outcomes and how useful are they, how well does citizen science work with quantitative and qualitative data collection. The speakers enumerated several positive impacts that participative research has (e.g. the [MonumentMonitor](#) project), among other things citizen scientists' intervention alert site managers to site vandalism, heritage crime or simply (natural) accidents on the site, but also lower the amount of anti-social behaviour on the heritage site. Considering the quantitative data collected, e.g. tracking flood or erosion levels, it is true that such data needs heavy processing after collection and that other approaches are surely more time-effective (e.g. planting a camera on the site). However, the main added-value of such a project is to promote knowledge sharing and communicate science in an engaging way. As for the qualitative data collection in a participative way, the speakers see it as an even greater opportunity for community engagement and our understanding of the importance of cultural heritage since one can get direct access to participant's reactions to the heritage site/object in focus.

**Links to materials.**

- [Workshop recording](#) (min: 1:02:06–1:13:01)

**Outcomes & Feedback**

This workshop provided a clear and comprehensive introduction to the planning of citizen science and data crowdsourcing projects in the context of heritage science. After providing an overview of the range of opportunities for the application of citizen science in the area of heritage, the planning stages for a crowdsourcing project were discussed. Focus was placed on two elements: first, the importance of understanding from the start of the project the methods and requirements for collecting and processing data; and second, the need for responsible and effective community engagement for successful citizen science projects. Through the use of two speculative case studies, the key differences between data sourcing and data tagging projects was explored, identifying different participants' skills and engagement level. Furthermore, this structure offered the possibility of discussing existing crowdsourcing platforms, such as Zooniverse. Ethical considerations of involving communities in research activities were discussed, highlighting it should be a two-way collaboration, with the need to provide feedback and share outcomes. Thus, the workshop led the participants step by step through the process so they were able to acquire

the necessary understanding of all the phases of a citizen science project in order to ensure a successful outcome and a pleasing experience for all parties involved.

**Viewers' satisfaction.** Most of the viewers rated the workshop as Excellent or Very good (cf. Q1). For half of the respondents the workshop exceeded expectations, while for the other it matched their expectations (cf. Q4).

The viewers were enthusiastic about the possibilities offered by participative research approach not just for the scientific community, but also for the public with its educational potential. The workshop left a highly positive impression on the viewers: they were able to acquire the necessary understanding of the implications of involving citizen scientists into a research project from the start till the end (cf. Q5) and build the “confidence on relevance of approach” (Respondent No. 5, Q5). The answers show that the workshop also sparked concrete interest in the participants to organise their own citizen science projects (Respondent No. 3, Q5).

**Feedback regarding the Organisation.** Most of the viewers felt that the event was Excellently organised (cf. Q9). Regarding the format, they appreciated its “accessibility via Zoom” (Respondent No. 1, Q6), the fact that the recording will be available later to refer back for some details (Respondent No. 6, Q5 & Respondent No. 7, Q5), as well as the event structure itself with use cases that neatly demonstrated the various factors to be considered, interactive tasks and speaker’s engaging presentation (cf. Q6).

The information channels through which the viewers learned about the event were very diverse: from social media and mailing lists to personal invitations among colleagues. Since the promotion activities through SSHOC channels were similar to our other events where such promotion proved successful, the absence of SSHOC-related news sources from the pool of information channels in this particular case might indicate that heritage science and archaeology research communities are strongly connected through community-specific channels and less through general purpose news sources, similar to those provided by SSHOC. This can be a helpful insight into how to adapt promotion activities for this specific research community.

**Future work.** The participants of the survey joined the workshop because they were curious “to learn more about the potential and limitations of running citizen science activities” (Respondent No. 3, Q3), since the concept is not yet well known (Respondent No. 2, Q3). They also noted that it would be great to have more micro-presentations of ongoing citizen science projects (Respondent No.8, Q7), and to receive short handouts or similar breakdown of the crucial factors, since this would serve as a great reference point for further work (Respondent No. 7, Q8). Overall, the high number of viewers shows that this is a popular topic that would deserve additional attention in the future not just in the field of cultural heritage, but also in other SSH domains that could be well supported by the participative research approach.

---

# ANNEX 3: WORKSHOP REPORT: DIGITISING MUSEUM OBJECTS USING BASIC PHOTOGRAMMETRY

## Background

The workshop — [Digitising Museum Objects Using Basic Photogrammetry](#) — was held on 25<sup>th</sup> March 2021 and was organised by SSHOC T6.5 in cooperation with [UCL ISH](#) as a second follow-up workshop to a successful webinar delivered in spring 2020 titled [Use and Re-use of Scientific Data in Archaeology and Heritage](#). The workshop enabled knowledge transfer on the topic of photogrammetry – an imaging technique that is increasingly gaining popularity especially in the heritage science.

The workshop contributed to SSHOC objectives and its commitment to further Open Science principles by introducing the key aspects of an affordable geometry capture technique that can return quality research data and promote the involvement of a wider team of (citizen) scientists. By systematically presenting all of the steps in the process of 3D model creation, the participants were able to acquire crucial understanding of the technique and develop skills that form the basis of each successful photogrammetry project.

## Workshop Overview & Format

**Aim.** There is a wide range of imaging techniques that have high potential for their application to the study, conservation and management of cultural heritage. One of these techniques is photogrammetry, the science of extracting 3D geometric data from 2D photography. The accessibility of quality digital photo cameras, including those embedded in mobile phones, makes this an affordable geometry capture technique in growing demand. The general aim of this training workshop was to introduce participants to the use of photogrammetry in heritage. Through participation in the workshop, participants were able to gain insights into the photogrammetric process, understand what is achievable with basic equipment as well as become more aware of the requirements of 3D digitisation of museum objects and how these might differ from everyday objects.

**Speakers.** The workshop was delivered by 3 speakers:

- [Kira Zumkley](#) has worked in the heritage sector for over 10 years as an archaeologist, photographer, and researcher. She is the Chair of the Association for Historical and Fine Art Photography and Honorary Lecturer at the Centre for Digital Humanities at University College London. In addition, she is currently working as a researcher on a joint project between the

Victoria and Albert Museum, University of Brighton and University College London. Prior to moving into academia, Kira worked as the photography manager at the Science Museum Group and carried out her own creative photography practices focusing on the contrasting experiences of urban and natural spaces.

- [Adam Gibson](#) (UCL ISH) is a Professor of Heritage Science at the UCL Institute for Sustainable Heritage. His group applies imaging techniques to aspects of heritage, predominantly multispectral and hyperspectral imaging. His work has looked, among other things, at painting analysis and feature recovery in historical documents.
- [Alejandra Albuerne](#) (UCL, SSHOC T6.5) is SSHOC representative and moderator.

**Organisers.** The workshop was organised in cooperation with partners in T6.5 of the SSHOC project and [UCL ISH](#) which was the principal organiser of the [Sustainable Heritage Bidecennial Conference](#) that co-hosted this SSHOC workshop.

**Participation.** There were 117 participants at the workshop.<sup>26</sup> The majority of participants (78%) came from European countries, both EU and non-EU, while the remaining 22% represented countries on other continents (Algeria, China, Egypt, Indonesia, Saudi Arabia, South Africa, the USA, etc.). The audience of the livestream covered seven stakeholder categories identified in D6.1. The biggest portion of the audience (83%) represented the broad scientific community (categories “Researchers, Research Networks and Communities”, “Universities and Research Performing Institutions”, “Research Libraries and Archives” and “Research and e-infrastructures” in that order). “Private Sector and Industry Players” category represented 10% of the participants, “Civil Society and Citizen Scientists” around 5% and “Policy Making Organisations” approximately 2% of the participants. The category “Research Funding Organisations” was not represented in the audience.

**Brief summary of the event structure.** Due to the Covid-19 pandemic, the conference [Sustainable Heritage Bidecennial Conference](#) which co-hosted this workshop and consequently the workshop itself were held online. The workshop lasted for one and a quarter of an hour and was divided into three parts: two shorter ones were dedicated to the presentation of SSHOC and to UCL ISH work on imaging; and a longer presentation delivered by the lead presenter Kira Zumkley on the main topic of the workshop. The main speaker answered questions throughout the event.

## Presentations & Discussions: Key Points

**First Session.** Introduction to SSHOC

---

<sup>26</sup> The number consists only of participants of the livestream. It should be noted that the overall reach is higher since this number does not include the views of the workshop recording which is available on the SSHOC YouTube channel and promoted by SSHOC and its partners.

**Speakers.** [Alejandra Albuerne](#)

**Main points.** The workshop was given in the framework of the SSHOC project. The speaker presented the goals and the expected impacts of the SSHOC project and situated the workshop in relation to them.

**Links to materials.**

- [Workshop recording](#) (min: 0:00–9:30)
- 

**Second Session.** Imaging at UCL Institute for Sustainable Heritage.

**Speaker.** [Adam Gibson](#)

**Main points.** UCL ISH works on various aspects of imaging, photogrammetry being just one. Its research focus lies in multimodal imaging, image processing and chemical imaging. In his presentation, the speaker talked briefly about different types of imaging other than photogrammetry by listing them as methods that offer from less to more details about the object. He first presented *reflectance transformation imaging*, which is an extension of photogrammetry and is highly useful to restoration specialists. This method enables high techniques and allows moving artificial lights around the image which helps, unlike the standard imaging, to detect the most deteriorated areas to be restored. The second method is *multispectral imaging* which utilizes high resolution cameras and computer controlled lightning in different colours. This technique helps uncover elements of the object that are not seen with the naked eye. Next method – hyperspectral imaging – is an extension of the multispectral imaging and offers even more detailed image of an object since by splitting the white light. 600 colours can be used in producing the image in comparison to 16 colours used in multispectral imaging. Furthermore, this technique can also be used to uncover chemical properties, such as the paper quality, which is an area of research with growing importance.

**Links to materials.**

- [Presentation slides](#)
  - [Workshop recording](#) (min: 9:31–18:00)
- 

**Third Session.** Digitising Museum Objects Using Basic Photogrammetry.

**Speaker.** [Kira Zumkley](#)

**Main points.** The presentation started with a grounded introduction to the photogrammetry covering different aspects such as the actual technique and its usage, theoretical background, and the science behind it.

The applications of photogrammetry in heritage are many. It is increasingly used for both documentation and interpretation applications that are aided by the fast creation of 3D models. It is applied to all scales of tangible heritage, from small objects or artefacts, to full archaeological or architectural sites. 3D models can be the basis for interactive interpretation models, 3D reproductions, architectural drawings or topographical maps, among many other uses.



The scientific fundamentals rely on triangulation: photographs from at least two different locations are taken and lines of sight are established from each camera position to focus on points of the object. Where these lines of sight intersect, three dimensional coordinates of the point of intersection can be measured. When it comes to small and medium-size objects such as those generally found in museums, there are two ways of creating a photogrammetric model: walking around the object or using a rotational table. Both present limitations: a rotational table can only be used with movable objects of a certain size. Moving around the object doesn't have such restrictions and can be used also for sites, but it requires a very rigorous methodology to achieve precise 3D models and reproducible results.

Some key messages regarding methodology are:

- Lighting plays a key role: diffused lighting that does not create harsh shadows is ideal. If light cannot be controlled, consider using a flash.
- Control your camera settings: shoot in .raw format, select the Adobe RGB colour space.
- Aim for the entire object to be in focus: you can use a depth of field calculator to find the right aperture for your camera, lenses and object dimensions.
- Control the light by combining shutter speed and ISO – a lower ISO is preferable, so best use a tripod to minimise image noise.
- Assess your object to identify the most challenging points – consider how many images you need to take from underneath.
- Aim for the object to take up the majority of your view, without cropping the object.
- Aim for a minimum 2/3 overlap between images and take pictures in order whenever possible.

The quality of commercial photogrammetry software is constantly improving to ensure that the most accurate model is created from the photographs at hand. There are free versions of some of these software, but typically capped by the number of photographs they can process to create the model. In case of using these free versions, photographs need to be carefully planned to cover the entire object with sufficient overlap without exceeding the maximum number of images.

Some image post-processing is typically applied to all images before inputting them into the photogrammetry software. Images are often colour-treated to soften shadow contrast, for example, in order to maximise the capabilities of the photogrammetry technique. It is recommended to carefully record any post-processing, especially when models are disseminated through [Sketchfab](#) (the *youtube* of 3D models).

It is essential to remember that the quality of the 3D model will depend on the quality of the photographs. Sometimes the processing struggles to align specific images where there is not enough overlap or where there are misleading features. Identifying misaligned images and disabling or realigning them can help improve the accuracy of the model.

#### **Links to materials.**

- [Presentation slides](#)
- [Workshop recording](#) (min: 18:00–1:16:23)



**Speaker.** [Kira Zumkley](#)

**Main points.** Most of the questions were very specific and related to functionalities of software presented. Kira Zumkley suggested that the best object for the first try at photogrammetry is a sports shoe since not one side is the same as the others, it has numerous details and everyone has it at home. One viewer raised a question regarding the bottom of the object and how to process it via software. Unfortunately, the software often does not automatically recognize that image as the bottom part, so this link should be established manually by indicating the points on the upper and bottom image that go together. Another important issue was related to background artefacts and what to do with them. The speaker underlined that when dealing with heritage objects, care should be taken to minimize the editing interventions since this can negatively impact the final output and its interpretation. Therefore, metadata and paradata should be meticulously noted and attached to the final model. There was also a question with regard to using photogrammetry on painting to which the answer was that this technique works extremely well on paintings as long as two thirds overlap is ensured when taking the images.

**Links to materials.**

- [Workshop recording](#) (min: throughout the event)

## Outcomes & Feedback

Photogrammetry is still an evolving technique that is constantly improving in order to overcome its limitations (e.g. capturing transparent or highly reflective surfaces). It is unquestionably gaining prominence in the heritage sector as a non-invasive, easy to use and less-costly geometry capture technique and one of the most broadly-spread imaging techniques in the sector. This workshop enabled participants to familiarise themselves with the field of imaging applications in the study and management of heritage. They were able to get an overview of the photogrammetric process and familiarise themselves with the scope, advantages as well as limitations of this method. Last but not least, participants were given concrete examples and suggestions on how to apply this method to their work by using basic photogrammetry equipment.

**Participant's satisfaction.** Almost all of the participants rated the workshop as 'excellent' or 'very good', however one participant was only fairly satisfied with the event due to apparently too long of an introduction (cf. Q1). For the great majority, the workshop (greatly) exceeded expectations, while for the others, it matched their expectations, except for the one participant already mentioned (cf. Q4).

"Really enjoyable session and very informative." (Respondent No.16, Q8)

The participants were particularly grateful for and excited about the main speaker, complimenting her expertise, clarity, being to the point and the presentation format including many examples (cf. Q6). One respondent also noted that it was good to have "different views on one topic" (Respondent No. 18, Q6).

They felt that the event will have direct positive impact on their work, encouraging them to give it a try as the following extracts from the survey demonstrate:

- “I will be using photogrammetry soon and this was an excellent introduction. I look forward to seeing the recording /.../ (Respondent No. 3, Q5).
- The workshop “gave enough detail and encouragement to attempt photogrammetry” myself (Respondent No. 6, Q5).
- The workshop “has enabled me to think of ways to provide museum experiences online and engage with a wider audience” (Respondent No.16, Q5).
- “I am more interested now about the technique and will look more into it” (Respondent No.17, Q5).

**Feedback regarding the organisation.** Most of the participants felt that the event was well or excellently organised, while a couple thought that improvements could be made (cf. Q9). The main reproach was related to the format of the workshop where apparently too much time was dedicated to the introduction of the SSHOC project and the UCL ISH work (Respondent No.11, Q7). Even though some praised the SSHOC introduction (Respondent No. 14, Q6), the general frustration was related to the feeling that much more information on the photogrammetry could be shared if the entire intro was shorter (although taking up only a quarter of an hour, while the main speaker at the same time prolonged her talk for the same amount of time): “A lot of time was taken up with the intro - interesting but we needed the hour for photogrammetry” (Respondent No. 14, Q7). This frustration with a lack of available and free events on the topic of photogrammetry probably partially explains the contradictory opinions regarding time management which for some was good enough to be highlighted as a positive point, and for others bad enough to be noted as a negative point (cf. 6 & 10). In terms of introduction, one respondent suggested that next time, the workshop description should be made clearer, since “it wasn’t super clear what the workshops wanted from us as participants” (Respondent No. 18, Q6). As a point aside, Respondents No. 11, Q8 & 14, Q6 also remarked that events on this topic are mostly commercial and therefore often hardly accessible, so it was very welcome that this event was free of charge.

The information channels through which the participants learned about the event were very diverse: from social media and mailing lists to personal invitations among colleagues. However, again similar to the [SSHOC workshop](#) on citizen science the day before, it is interesting that SSHOC channels were not enumerated among the information sources. Since the promotion activities through SSHOC channels were similar to our other events where such promotion proved successful, the absence of SSHOC-related news sources might indicate that heritage science and archaeology research communities are predominantly connected through community-specific channels, which should be taken into account when developing outreach activities. Furthermore, one respondent noted that the event could be more advertised (Respondent No. 6, Q7). However, even with the given amount of advertisement, the workshop received one of the highest attendance numbers.

**Future work.** The participants of the survey joined the workshop because they want to use this technique in their future work and include the newest trends in their teaching (cf. Q3) or “to know to what

extent photogrammetry /can/ help in /.../" their current projects (Respondent No. 9, Q3). Many of the respondents noted that given the topic, the workshop should definitely be longer in order to cover various steps of the process in more details, have the opportunity for more hands-on work and more time for a discussion (cf. 7 & 10). One respondent noted that they “would like to attend more advanced levels” of the workshop (Respondent No. 14, Q8) and that future events could also cover “the topic of hyperspectral and multispectral imagining for Cultural Heritage” (Respondent No.9, Q8). All the positive feedback, as well as the negative remarks, show that this topic is in high demand and that, according to respondents, not enough opportunities are currently offered to cover the needs for knowledge transfer in this area of research. Therefore, this SSHOC workshop helped uncover a trending area for the research community on the one hand, and on the other, contributed a small part to help fill the gap in scientific training, thus addressing timely and concrete needs of researchers and professionals in the field.

---

# ANNEX 4: WORKSHOP REPORT: DATA PROTECTION IN RESEARCH PRACTICE: THE GDPR AND THE ELDAH CONSENT FORM WIZARD

## Background

The workshop — [Data Protection in research practice: The GDPR and the ELDAH Consent Form Wizard](#) — was held on 13 October 2021 and was organised as a hybrid event at the University of Zadar and online in cooperation between the [DARIAH ELDAH working group](#) and T6.5. The workshop was part of T6.5 target training activities that fall under the broad topic of “Data Protection and the GDPR”. It was delivered as a follow up to [a webinar delivered on 13 October 2020](#) which showcased a previous release of the Consent Form Wizard.

The workshop focused on explaining the European legal framework for the processing of personal data which is one of the crucial aspects for ensuring sustainable and trustworthy research practice. Apart from developing the awareness linked to privacy in data management, the workshop also offered hands-on content by demonstrating a tool that helps researchers more easily comply with the relevant European legal framework. Therefore, the workshop successfully contributed to the SSHOC efforts linked to ensuring Open Science principles.

## Workshop overview and format

**Aim.** The General Data Protection Regulation (GDPR) provides a legal framework for data protection and processing of personal data and consequently extensively impacts the work of scientific community. The aim of the workshop was to help participants understand how their research is affected by the GDPR and what issues need to be considered. Researchers have to find a legitimate basis, for example by obtaining legal consent from their participants, in order to comply with the GDPR for many of their activities ranging from research to dissemination, which can often hinder their work. In order to help them overcome this obstacle, the DARIAH-EU working group ELDAH (Ethics and Legality in Digital Arts and Humanities) has developed a tool, the Consent Form Wizard, that provides standardized consent form templates for obtaining legal consent from data subjects, i.e. human participants of research endeavours and events. The participants of this workshop were able to test this tool which enables digital scholars and the wider research infrastructure community to quickly and easily obtain a standardized consent form that is legally valid in all of the European Union.

**Speakers.** The workshop was delivered by four speakers:

- Koraljka Kuzman Šlogar (Institute of Ethnology and Folklore Research/DARIAH-HR)
- Vanessa Hanneschläger (CLARIAH-AT)
- Walter Scholger (University of Graz/CLARIAH-AT)
- Paweł Kamocki (IDS Mannheim/CLIC)

**Organisers.** The workshop was organised in cooperation between T6.5 and the DARIAH-ELDAH group. The workshop was co-located with the [1st DARIAH-HR International Conference](#) and delivered at the University of Zadar. Appropriate technical settings were put in place in order to ensure good engagement of both, the onsite and online audiences.

**Participants.** There were 34 workshop participants: 25 online participants, and 9 onsite participants (excluding organisers and speakers).<sup>27</sup> The majority of viewers came from European countries (EU 66%, non-EU 31%), while the others represented countries outside Europe (India, the USA, Tunisia). The audience was mainly represented by “Researchers, Research Networks and Communities” (65%) and “Research Libraries and Archives” (20%), a couple of the participants represented “Research and e-Infrastructures” (15%).

**Brief summary of the event structure.** This was a full day workshop (9:30–17:00) delivered in a hybrid format (see Appendix 1 for full program). The programme was tailored to accommodate the needs of two audiences (onsite and online) and keep both informed and engaged. The workshop started with an introduction to the SSHOC project and DARIAH-EU ELDAH working group activities. It continued with a general introduction to the topics of data protection and the processing of personal data in the context of research, education and cultural heritage in order to summarize the pertinent articles of the GDPR. Afterwards, the speakers presented the most common scenarios in which humanities' and social sciences' researchers encounter data protection issues with examples such as surveys, interviews and other research activities involving human participants. They also presented common academic scenarios like communications (mailing-lists, newsletters) and events (e.g. regarding picture and video recordings). This introductory part was followed by a demonstration of the Consent Form Wizard. Afterwards, the participants had a chance to test the tool, apply it for their research purposes and explore the customization options while at the same time discussing issues with other participants and speakers (either in breakout groups online, or in small groups onsite). The workshop included two dedicated slots for discussions and ended with a lively exchange of feedback from all groups covering possible improvements, bugs in the tool and additional scenarios to be included in the tool in case of further funding and extension opportunities. The onsite participants also had a chance to network during breaks throughout the day.

---

<sup>27</sup> Overall, there were 59 workshop registrations. The number does not include the number of views of the workshop recording which is subject to change.

## Presentations & Discussions: Key Points

**First Session.** Presentation of the DARIAH-ELDAH working group and the SSHOC project

**Speakers.** Koraljka Kuzman Šlogar

**Main points.** The speaker shortly presented the history of ELDAH as well as its membership. She pointed out that ELDAH is closely collaborating with other groups working on ethical and legal issues, such as that of CLARIN and CESSDA. The expertise of ELDAH group covers intellectual property rights and licensing, data protection and privacy as well as research ethics and scholarly conduct. The speaker then presented the goals and the expected impact of the SSHOC project and placed the workshop in relation to them. In addition, the speaker also pointed out some of the SSHOC offerings, such as the SSH Open Market Place and the Discovery Toolkit for Trainers, and invited the participants to future SSHOC events.

**Links to materials.**

- [Presentation slides](#) (pp.: 1–9)
- 

**Second Session.** Introduction to data protection and the GDPR

**Speaker.** Paweł Kamocki and Walter Scholger

**Main points.** The speakers first underlined that data protection and privacy are not the same thing although they are related. GDPR concerns data protection and regulates the protection of personal information. It was passed in 2016 and enforced two years later, and is applicable all over EU. While it is obligatory for the Member states to comply with its contents, the Directive itself does not indicate the obligatory way of its application. GDPR is not retroactive, but in case of processing of data collected before 2018, the Directive nonetheless needs to be applied.

The speakers then presented the basic concepts of GDPR, i.e. personal data, processing, data subject, data controller, data processor, data protection officer, anonymisation and pseudonymisation. It was underlined that personal information is protected whether true or false, and that even information about a deceased person can be problematic in terms of data protection of their living relatives. While data subject and data controller always exist in the event of data processing, the role of data processor, defined as the entity who conducts operations with and on data on behalf of the data controller, is not necessarily always fulfilled. Given that personal data is any information that relates to a natural person in a way that this person can be identified via this particular information, anonymisation and pseudonymisation can be used as a way to ensure data protection but at the same time use this data for research purposes. However, breaking the link between the natural person and the data is becoming a very complex procedure, since with technological progress we experience that reasonable means of linking data to a person became considerably broader in comparison to half a decade ago. While anonymisation means breaking any link between the person and data, pseudonymisation means encrypting the personal data by replacing it with identifiers. It should be noted that pseudonymised data is still considered personal data, because the link still exists.

Afterwards, the speakers presented safeguards and data protection principles, namely lawfulness, fairness and transparency, purpose limitation, data minimization, accuracy or data quality, storage limitation, integrity and confidentiality or data security and accountability. With regard to safeguards and privileges for academic research, the speakers present Article 89 of GDPR. They presented the main

aspects and mentioned some good practices, i.e. the importance of the code of conduct, offering GDPR trainings to researchers. The next topic were data protection principles where the speakers among others most extensively referred to the Articles 5 and 6 of GDPR. The most commonly used legal basis in research are consent and legitimate/public interest. Although there are exceptions, consent is taken as the rule basis for the academic community. The core element of the consent is informing the data subject about the purpose and means of data collection and processing which is where Consent Form Wizard can greatly help. In principle, personal data can only be processed for a clearly defined purpose, but if proper safeguards exist, personal data can be reused for other research/archiving purposes. It is also important to remember that personal data can only be stored for a limited period of time. However, for research purposes, personal data can be stored for longer periods of time if proper safeguards are put in place. This does not mean that data necessarily needs to be destroyed after the agreed period of time, but that it needs to be thoroughly reviewed in order to ensure that it still complies with data protection standards which constantly evolve.

Finally, the speakers elaborated on the topic of rights of data subjects according to GDPR. The rights of the data subject include the right to being informed, have access to the information about data processing, to rectification, data portability, erasure, restriction of processing and to withdrawal of consent. However, these rights are heavily limited in research, especially if this means inhibiting the research process. At the end, the speakers also commented on all the necessary information that must be provided to data subjects in order for the process to be lawful

**Links to materials.**

- [Presentation slides](#) (pp.: 10–14)
  - [Workshop recording](#) (min: 00:00–49:54)
- 

**Third Session.** Introduction to the Consent Form Wizard

**Speaker.** Vanessa Hanneschläger

**Main points.** The speaker talked about most common scenarios in which humanities' and social sciences' researchers encounter data protection issues and presented available consent form templates. Three scenarios that would require a consent form in order to comply with the GDPR, were considered during the workshop, namely (1) gather data about living people for research purposes, (2) communicate through mailing lists or other digital media, and (3) gather data/consent from participants as a host of an academic event. The most often used template in the tool is the template for gathering data for research purposes. The speaker also demonstrated the process of creating a consent form and pointed to different wizard functionalities. It has been underlined, however, that the consent form does not replace an official opinion from legal experts. In fact, in more complex situations, especially for third persons, it is not advised to use the consent Form Wizard, because there are too many open legal issues. Although this is frustrating advice, the speaker believes it is the only responsible way of operating. In addition, the speaker highlighted the importance of Consent Form Wizard localization, and invited the participants to join their efforts in translating the tool in more European languages, since the tool is currently available only in English, German, Italian and Croatian.

**Links to materials.**

- [Workshop recording](#) (min: 49:55–1:15:00)
-



**Fourth Session.** Discussion in small groups/breakout rooms and Q&A time

**Speaker.** All speakers

**Main points.** Two discussions, one in the morning and one in the afternoon, were organised in small groups, either online or onsite, and were followed by a plenary discussion to share the most interesting insights. These exchanges offered an opportunity to discuss several topics. The participants, for example, discussed the difficulties related to data protection when using social media data for research. In fact, obtaining large-scale consent from social media data subjects is in practice almost impossible. Some pointers for participants interested in this topic were shared, esp. the CLARIN Café on Text and Data Mining Exceptions in the Directive on Copyright in the Digital Single Market.

Another topic discussed to some length referred to the difficulties linked to delimiting the border between private and public data, since it is not always easy to state where Personal Data starts, given that an institutional address might sometimes be quite personal. Additional problem is linked to the danger of indirect identification which can be, as indicated by some participants, so broad that the anonymization/pseudonymisation can alter the data to a great extent which might hinder research.

Among other discussion points, the participants also raised concerns related to the issue of the (too) many different consents that data subjects have to give and how the overflow of GDPR consent forms is sometimes disqualifying GDPR protection efforts. Furthermore, the participants discussed the data protection issues linked to audio data and audio consent, the difference between what is ethically right and what is legally necessary, the issues related to the overlap between data protection and other legal frameworks, such as intellectual property, the problems with legally defining the location of multinational institutions, like UNESCO, and addressed questions such as at what point does data processing start and others.

#### Outcomes & feedback

In this full-day workshop, the participants learned about most common data protection issues in humanities research and their possible GDPR-compliant solutions. Given the new technical developments and the ever changing digital landscape it is not surprising that data protection regulations and ethical principles in research constantly evolve. In order to support researchers in this numerous requirements, the ELDAH working group developed a tool that provides standardized consent form templates for obtaining legal consent from human participants. Through a demonstration of the ELDAH Consent Form Wizard, the participants not only learned how to use the tool that provides consent form templates that can be adapted to individual needs, but also gained understanding of the crucial elements of obtaining any consent from data subjects. In the workshop, the participants closely investigated three most common scenarios that, according to GDPR, require consent from participants, and were given ample time to share their thoughts with other participants, get answers to their questions from experts in the field, and suggest improvements to the tool. Given its hybrid format, the workshop was also an important event for the research community, since it provided a long awaited opportunity for onsite networking.



**Participant satisfaction.**<sup>28</sup> The feedback from participants shows that the workshop was well organised and delivered valuable information for their future work. The onsite participants greatly enjoyed the opportunity to be able to attend a face-to-face event. The online participants were very satisfied with the engagement in the break-out rooms which was very well stirred by the speakers/moderators. Overall, the participants like the fact that the speakers linked GDPR provisions to real-life situations. The morning and afternoon discussions proved to be a great opportunity to connect the two audiences, and the lively debate showed that the topic of data protection and GDPR-compliance is still a very timely issue. Participants stated that the workshop will have a direct positive impact on their work, especially thanks to the Consent Form Wizard which they will be able to use themselves and recommend to their colleagues.

## Appendix

### Appendix 1: Workshop program

9:30 - 10:30	Welcome by the organizers Participant introduction round Introduction to the workshop goals Introduction to SSHOC and ELDAH Introduction to data protection and the GDPR
10:30 - 11:00	Coffee break
11:00 - 11:45	Group discussion: most common data protection issues in humanities research
11:45 - 12:30	Plenary presentations of breakout results, common denominators
12:30 - 14:30	Lunch break
14:30 - 15:00	Introduction of the CFW: Scenarios, consent form templates, wizard functionality
15:00 - 16:00	CFW testing by participants: Bring your own use-case
16:00 - 16:15	Coffee break
16:15 - 17:00	Feedback by participants & End of the Workshop

<sup>28</sup> The post-event survey was only filled-in by three participants. This report, however, includes also the feedback collected during the workshop either orally or in writing via chat.

---

# ANNEX 5: WORKSHOP REPORT: CARING FOR SHARING – DATA MANAGEMENT AND FAIRNESS OF MIGRATION DATA

## Background

The workshop—[Caring for Sharing – Data Management and FAIRness of Migration Data](#)— was held on 9 March 2020 at the [COST Action 16111 - ETHMIGSURVEYDATA Work group \(WG\) plenary meetings and 2nd Annual Policy Dialogue Conference](#) in Brussels, Belgium. The workshop was organised by T9.2 and T6.4 in cooperation with T6.5 of the SSHOC project on the pre-defined topic of “Data Stewardship and RDM in theory and practice”.

SSHOC was involved in the event design and implementation in two specific ways. First, SSHOC facilitated a workshop—*Caring for Sharing – Data Management and FAIRness of Migration Data*. The workshop addressed how Social Sciences and Humanities (SSH) data, in particular within the ethnic and migration studies domain, can comply with and promote the objectives of Open Science and the FAIR principles<sup>29</sup> by leveraging existing SSH data management tools and resources. Second, as a thematically relevant follow-up to the workshop, SSHOC members from T9.2 Laura Morales and Ami Saji held a demonstration session of the EMM Survey Registry. The workshop “Caring for Sharing” as well as the event as a whole, therefore, supported SSHOC’s vision in making SSH data open and FAIR-compliant.

## Workshop Overview & Format

**Aim.** The targeted training SSHOC workshop (T6.5 activity) “Caring for Sharing” was dedicated to data sharing and ways to achieve it for migration survey data while ensuring data FAIRness. This workshop addressed the following topics: (1) how migration survey data can be prepared and processed before dissemination, (2) how transparency, availability, accessibility and replicability of data can be improved and (3) what resources and services offered through SSHOC could be useful in this sense. To this end, it made use of examples of data available from the EMM Survey Registry and the UK Data Service catalogue.

The aim of the joint event – SSHOC and COST Action 16111: ETHMIGSURVEYDATA Work group (WG) plenary meetings and [2nd Annual Policy Dialogue Conference](#) – was to illustrate how SSHOC and the COST Action are helping to make EMM survey data open and FAIR and to facilitate information/knowledge sharing of EMM survey research. In particular, the 2-day event focused on: (1) highlighting the efforts and strides made by the SSHOC T9.2 and COST Action and in improving the production, access, usability, and dissemination of survey data on ethnic and migrant minority (EMM)

---

<sup>29</sup> The FAIR acronym stands for principles that help make data Findable, Accessible, Interoperable and Reusable.

integration; and (2) showcasing interesting and innovative survey research being conducted on EMM integration by SSHOC and the COST Action members.

**Speakers.** The workshop was delivered by one speaker.

Anca Vlad is a data repository administrator at UK Data Service. She manages data deposits through UK Data Service self-deposit repository (ReShare), advises on data management, data deposit and ethical considerations when archiving data, and conducts data disclosure checks on deposited data.

**Organisers.** The workshop was organised in cooperation with partners from T6.5, T9.2 and T6.4 and the [COST Action 16111 - ETHMIGSURVEYDATA](#).

**Participants.** The workshop “Caring for Sharing” was attended by 33 participants from 17 European countries. The participants represent various scientific domains. Most of them are professionals in the field of political science and sociology. Others work in the fields of psychology, demography, social anthropology and ethnology, methods and statistics, and economics and finance. Four different stakeholder categories – *Researchers, Research Networks and Communities, Universities and research performing institutions, Research and e-infrastructures and Policy making organisations* – were represented. The first category was represented most prominently by 21 participants.

The entire event was attended by over 60 different individuals from the ETHMIGSURVEYDATA community: researchers and policy-oriented professionals from academic institutions, think tanks, local/national/regional governments, international organisations, civil society organisations, private companies, etc. The participants work with quantitative survey data and/or ethnic and migrant minority (EMM) integration.

**Brief summary of the event structure.** The workshop “Caring for Sharing” was divided into two parts. The first, theoretical part was delivered in the form of a lecture which also included an open discussion between participants and the speaker. The second, practical part included hands-on activities in the form of individualized consultations which were offered throughout the day of the workshop.

The event as a whole lasted for 2 days (9–10 March) and included Cost Action’s Working Group meetings and ETHMIGSURVEYDATA Policy Dialogue Conference. On the second day, Laura Morales and Ami Saji (both T9.2) delivered a follow-up session to the workshop “Caring for Sharing”. They presented the EMM Survey Registry, a tool developed in the framework of SSHOC, which is a concrete example of how the FAIR principles are being applied to EMM data. Detailed breakdown of activities can be seen in the [agenda](#).

## Presentations & Discussions: Key Points

**First SSHOC Session.** Data Management and FAIRness of Migration Data

**Speaker.** [Anca Vlad](#)

**Main points.** The main purpose of the workshop was to provide the audience with information, useful resources and skills to ensure the data they produce is findable, accessible, interoperable and findable (FAIR). The workshop was split into four sections: (1) detailed information about each of the principles,

(2) examples of FAIR datasets available from multiple sources, (3) useful resources that researchers can make use of to ensure, improve or check the FAIRness of their data, and (4) interactive activities.

To introduce the need for and importance of data sharing, the workshop began with a short summary of lessons learned during a migration research seminar organised by the UK Data Service in November 2019. This brought together various stakeholders, from researchers and research participants to third sector – independent charities as well as data publishers and other archives. The workshop focused on the complexities and opportunities of sharing, archiving and reusing migration data.

The speaker made a short presentation of the FAIR principles: what the essential aspects are and what boxes to tick when it comes to data FAIRness.

- **Findability:** researchers should look for a globally unique and persistent digital identifier (such as a DOI). A short video was played (available [here](#)) on why this digital identifier is so important. It is essential, of course, that data is published with an accredited archive or data repository that would produce this unique identifier. The researchers should also make sure to check whether the archive or repository chosen is [CoreTrustSeal](#) accredited.
- **Accessibility:** (meta)data should be retrievable using an open, universally applicable (i.e., to all datasets in an archive), standardized protocol. Depending on the content of the data, this protocol would provide information about the access level and terms to obtain the data. Although access restrictions are almost the same across data providers (open, safeguarded, or restricted), the process and conditions for obtaining the data can vary across archives/repositories/data publishers. Note that data does not need to be open to ensure accessibility of the data. What is essential is that this process of obtaining data is made clear, in plain language for each dataset.
- **Interoperability:** the metadata needs to be in specific formats and respecting community agreed standards and vocabularies/ontologies, such as the [DDI Schema](#). The data should be in specific formats because digital data is substantially software-dependent, therefore endangered by obsolescence of software/hardware. The [EMM Survey Registry](#) was used as an example, as it uses a multitude of machine readable metadata fields such as scope, region(s), start date, end date of survey, target population etc. to describe all its datasets.
- **Reusability:** the reusability of data is achieved with quality supporting documentation and metadata, as well as the applicable license. The license should allow the data to be available to the widest possible audience with the widest possible range of uses.

After looking in detail at the FAIR principles, the presentation covered some useful resources of which researchers can make use. These include:

- [SSH Open Marketplace](#) can be used: (1) to find relevant software and services, (2) to check whether your project aligns with standards and open science principles, (3) to find links to tutorials and other training materials and (4) as a forum where peers can comment on tools/software.
- [CESSDA Data Management Guide](#) was designed by European experts to help social science researchers make their research data FAIR. It supports the entire research data lifecycle from planning, organising, documenting, processing, storing and protecting the data to sharing and publishing.
- [DMP Online](#), provided by the Digital Curation Centre (DCC), helps with creation, review, and sharing of data management plans that meet institutional and funder requirements. Writing a data management plan at the beginning of a research project is an important step that can help

throughout the project with various aspects: collection, storage, documentation, formatting, ethics, copyright, transfer, de-identification/anonymization and sharing.

- [Go FAIR starter kit](#) offers guidelines for research data management and an open and inclusive ecosystem for individuals, institutions and organisations working together.
- [UK Data Service](#) offers [data management guidance](#) and training events.

Throughout the day of the workshop, the participants were offered individualized information about the application of FAIR principles to their data as well as a demonstration of the [QAMyData](#) tool. This is a free, easy-to-use open source tool that provides a health check for numeric data. The tool uses automated methods to detect and report some of the most common problems in survey data: missingness, duplication, outliers and direct identifiers (information that can point to one person in particular, such as name, address etc.). Outliers and direct identifiers are of particular concern when sharing data as researchers need to uphold confidentiality agreements. The tool offers a number of configurable tests, categorized by type: file, metadata, data integrity, and identifiers; it can run popular file formats, including SPSS, Stata, SAS and CSV.

**Links to materials.** [Presentation slides](#)

---

**Second SSHOC Session.** Demonstration of the EMM Survey Metadata Registry.

**Speakers.** [Laura Morales](#) and [Ami Saji](#)

**Main points.** Laura Morales and Ami Saji (both T9.2) hosted a workshop on [the EthmigSurveyData Hub](#), which is a free to access online space that will:

- facilitate knowledge sharing and transfer about existing quantitative surveys on EMM integration [in the 35 COST Action-participating countries](#); and
- empower users from all different sectors to make quantitative survey data on EMM integration FAIR, as well as better harness such data for their own research and policy-oriented work.

Morales and Saji focused their workshop on the first-to-be-developed and launched component of the EthmigSurveyData Hub: [the EMM Survey Metadata Registry](#) (currently in alpha version). The EMM Survey Metadata Registry, which is one of the main outputs of ETHMIGSURVEYDATA and T9.2 of SSHOC (namely D9.4 of WP9), is being designed to be a user-friendly and intuitive online tool that will allow individuals to search for and learn about existing quantitative surveys to EMM populations through the survey-level metadata. Individuals will also be able to contribute to the registry (and therefore make it a sustainable and relevant tool for the short and long-term) by adding information (i.e. metadata) about quantitative surveys they have conducted or will be conducting to EMM populations.

The participants of this workshop were excited to learn about the EMM Survey Metadata Registry, as it is a tool that can easily be leveraged by the diverse research community working with quantitative survey data to EMMs.

**Links to materials.**

- Presentation slides
  - [Showcase of the Ethnic and Migrant Minorities \(EMM\) Survey Metadata Registry](#)



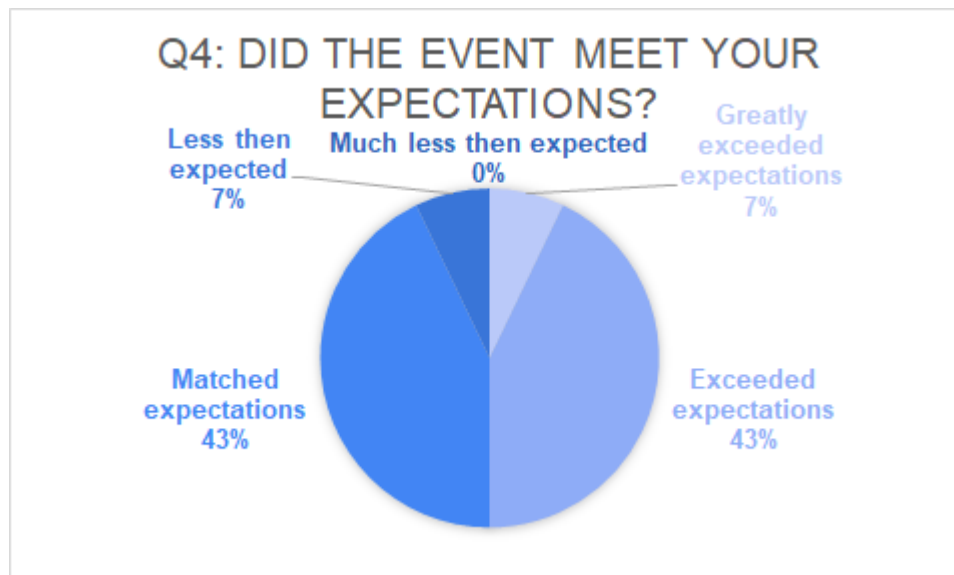
- [Sustainability and updating of the EMM Survey Metadata Registry](#)

## Outcomes & Feedback

One of the main takeaways of the “Caring for Sharing” workshop was the importance of data sharing for future research, in particular for future historians. It brought to light the lack of data currently available as well as how inaccessible and fragmented the small amount of available data is. Historians present at the event especially called upon migration researchers to contribute to reference archives.

Overall, the workshop “Caring for Sharing” as well as the follow-up session with the EMM Survey Registry demo were particularly instrumental in helping the participants understand why EMM survey data should be made FAIR and how this could be achieved through SSHOC and ETHMIGSURVEYDATA COST Action collaborative efforts. The feedback from the participants and their interest expressed through numerous questions during the sessions showed that services such as the [EMM Survey Metadata Registry](#) are in demand, and the presentations of good practices, such as the workshop “Caring for Sharing”, are crucial if we want to ensure FAIR data processing in the future.

**Participant satisfaction.** In general, the participants rated the workshop “Caring for Sharing” as Excellent (6) or Very good (6). Two participants marked the event as Good or Fair (cf. Q1). Based on the survey results, the workshop exceeded or matched expectations of the majority of the participants (cf. Q4).



The participants felt that the workshop will have a positive impact on their work. They were introduced to new tools and initiatives, and were able to learn about opportunities in cross-disciplinary work. They also made new contacts (cf. Q5). The added value of the workshop for researchers was nicely expressed by the following two participants:

- “/.../ It provided me valuable information on data access opportunities.” (Respondent No. 1, Q5).
- “Positive impact. I learnt about ongoing research, funding opportunities, and methodological tools.” (Respondent No. 5, Q5).

Overall, the participants underlined the quality of the workshop on the level of content and presentation:

- “/.../ It was comprehensive and elaborated with many details.” (Respondent No. 4, Q6).
- “/.../ Presentation was dynamic and informative.” (Respondent No. 9, Q6).

Some of them, however, were expecting more information about the SSHOC project and would like to have more time for discussion as put by Respondent No. 13, Q7:

“General discussion of some of the technical aspects/ restrictions now and in the near future, though it might be more relevant to a different type of audience.”

These are relevant remarks for future targeted training activities that will be organised by T6.5. In this particular case, however, we can state that due to last-minute venue change and preventative measures undertaken for the safety of the participants (both a consequence of Covid-19 pandemic) as well as due to the workshop format, group hands-on activities and extended discussions about specific aspects were not possible during the workshop. This limitation was predicted and mitigated with individualized consultations offered to the participants after the workshop “Caring for Sharing” throughout the whole day. Furthermore, as a work-in-progress, the majority of SSHOC services is currently not yet developed which is why specific information about tools is not available at this point in time and the focus is rather on raising awareness of the expected outcomes of the SSHOC project.

**Feedback regarding the organisation.** The great majority of the participants felt that the workshop was Excellently organised (cf. Q2). In terms of organisation, participants reported that it would be an improvement to the workshop if they could receive presentation slides in advance and if questions from the audience would be limited to the end of the workshop (cf. Q9).

**Future work.** The topic of data stewardship and RDM will be further elaborated in a webinar which will be organised as a follow-up T6.5 activity on 18 May 2020. The EMM Survey Metadata Registry will be launched in summer 2020 and participants of the workshop expressed excitement in testing out and using the registry once it goes live. T9.2 also plans to prepare workshops and demonstrations of the registry for the SSHOC community, EMM and survey researchers, policymakers and other policy-oriented professionals who could leverage EMM survey data (e.g., DG Home).

---

# ANNEX 6: WORKSHOP REPORT: COPYRIGHT ISSUES IN SECONDARY DATA USE

## Background

The online workshop — [Copyright Issues in Secondary Data Use](#) — was held on 24<sup>th</sup> and 25<sup>th</sup> January 2022 and was organised by SSHOC T6.5. The workshop focused on what needs to be taken into account when creating, using and sharing secondary data. Therefore, the workshop importantly contributed to the SSHOC objective of promoting best practices to produce FAIR data and maximize its re-use. The workshop offered a broad overview of the Copyright and its importance, but also explained the specific linked to Copyright for secondary data. The workshop thus enabled participants to expand their knowledge on Copyright and its main issues in secondary data use and gain a better appreciation of fair dealing with resources.

## Workshop Overview & Format

**Aim.** There is a misconception in the research community that the use of secondary data does not raise legal concerns. Secondary data is an intellectual property of the data owner, so any use of that data may require the data owner's permission or may be subject to certain terms and conditions. This workshop thus aimed to highlight the important aspects of using secondary data and to help potential users overcome the common issues. The content of the workshop was intended for anyone who was interested in learning the basic principles of Copyright as well as the main aspects of data creation, exchange and use. Apart from addressing general considerations concerning Copyright, special attention was given to the available licences together with guidelines how to choose between them when publishing data. The workshop was designed to encourage active participation from the audience, and the hands-on exercises were designed to help the participants remember the key aspects when sharing secondary data.

**Speakers.** The workshop was delivered by 5 speakers:

- **Hina Zahid** (UKDS) is the Senior Research Data Services Officer working in the UK Data Service's Research Data Management section based at the UK Data Archive, University of Essex. She leads on the ethical aspects of research data management, sharing and reuse. She works proactively with researchers, and research centres to provide guidance, advice and training to achieve the implementation of good data management practices, and to optimise the sharing and archiving of data for research.



- **Hannah Pyman** (University of Essex) joined the University of Essex Library team in September 2016 as a Graduate Trainee, going on to become Interlending and Document Delivery Assistant in April 2017. From September 2019 until July 2020 she was an Information Literacy Co-ordinator. She began her new role as Scholarly Communications Co-ordinator in July 2020. Hannah's role covers support across the research lifecycle, ranging from literature searching and evaluating resources, to publishing (including open access publishing), copyright, and citations.
- **Cristina Magder** (UKDS) manages the UK Data Service's Data Collections Development and Research Data Management teams at the UK Data Archive. She leads the research data management portfolio of support and training for the UK Data Service. Her main teaching interests are data management planning, sharing and archiving data with a specific focus on licencing, data quality assurance, disclosure risk assessment, and reproducibility.
- **Maureen Haaker** (UKDS) is a scholar in the field of qualitative methodology. She has worked with the UK Data Service for 10 years as a senior qualitative training officer and is a lecturer in the School of Social Sciences and Humanities at University of Suffolk. Her work has involved the ingest and digitisation of qualitative data, development of training for the management, deposit, and re-use of qualitative data, and the re-use of qualitative datasets for teaching and learning. She is a key contributor to Sage's Managing and Sharing Research Data (2nd edition) and Qualitative Secondary Analysis. She is currently completing a PhD in Sociology at University of Essex.
- **Anca Vlad** has been part of the collections development and data publishing team at the UK Data Archive for the past 6 years. She has been assisting researchers and data depositors with preparing their data deposits and delivering training on data management, archiving and open access publishing for different audiences.

**Organisers.** The workshop was organised as a stand-alone online event in cooperation between T6.5 members CESSDA/UKDS and CLARIN/UL-FF.

**Participation.** There were 38 participants at the workshop.<sup>30</sup> Two thirds of the participants came from European countries, both EU and non-EU, while others came from Australia, Brazil, Canada, Oman, Uganda and the USA. The audience of the livestream covered six stakeholder categories identified in D6.1. A great majority (80%) of the participants identified as belonging to the categories "Researchers, Research Networks and Communities", "Research Libraries and Archives" and "Universities and Research Performing Institutions". The remaining 10% of the audience represented "Policy Making Organisations" and "Private Sector and Industry Players". Considering the main scientific domain of the participants, the audience was very diverse. The most well represented domains were Management and Communication Studies (20% of the audience each), and Education and Sociology (10% of the audience each). Other

---

<sup>30</sup> There were 38 participants on Day 1 and 31 on Day 2. The number consists only of viewers of the livestream. It should be noted that the overall reach is higher since this number does not include the views of the workshop recording which is available on the SSHOC YouTube channel and promoted by SSHOC and its partners.

domains belonging to the SSH, such as Archaeology, Arts, Environmental studies, Linguistics, Psychology, etc., were also represented.

**Brief summary of the event structure.** The workshop was a virtual, two-day event (see programme below). The online format was chosen because of the restrictions linked to Covid-19 pandemic and to allow participants from multiple countries and backgrounds to join easily given the topics covered. The programme was split into two days in order to ensure engaging experience for the participants knowing that attention-wise, the online format is more demanding than an in-person event. Workshops are intended to let the audience connect with the presenters and enable an environment that allows attendees to speak freely. All the presentations delivered by the speakers have been recorded, but the Q&A sessions and hands on exercise have not been recorded to allow a better report.

The first day of the workshop was divided into four parts: an introductory session dedicated to the presentation of the SSHOC project; the main session in which the speakers presented the topic of the Copyright and Secondary data including a general overview of both concepts and historical background; the concluding session which was dedicated to the topic of secondary data; and two Q&A sessions to engage the audience and spark discussions. The second day of the workshop was again divided into four parts: a short introductory session dedicated to the recap of the first day; a longer session in which the speakers presented the specifics of Copyright including data licensing, international and national copyright requirements as well as some special cases, such as copyright for social media data; and a Q&A session to further engage the audience and allow discussions on questions covered in the Padlet. Both days included various prompts to engage the audience (live polls via Mentimeter, dedicated discussion space at Padlet) as well as hands-on exercises to encourage successful transfer of knowledge. Both days also had a dedicated time for a discussion, but the speakers answered questions also throughout the workshop.

## **Programme**

### **Day 1: 24<sup>th</sup> January 2022**

9:30 – 9:45	Welcome and Introduction
9:45 – 10:10	Presentation: Overview of copyright
10:10 – 10:45	Presentation: Introduction to copyright
10:45 – 11:00	Exercise
11:00 – 11:10	Break
11:10 – 11:30	Q&A Session 1
11:30 – 11:45	Presentation: What is secondary data?
11:45 – 12:10	Presentation: Copyright issues in secondary data use
12:10 – 12:25	Q&A Session 2
12:25 – 12:30	Close and briefing for the next session

### **Day 2: 25<sup>th</sup> of January 2022**

9:30 – 9:40	Welcome and recap of day 1
9:40 – 9:55	Exercise
9:55 – 10:15	Data licencing and copyright
10:15 – 10:35	Copyright exceptions and infringements
10:35 – 10:45	Break
10:45 – 10:55	Copyright in the international context
10:55 – 11:05	Copyright and social media
11:05 – 11:20	Q&A
11:20 – 11:30	Break
11:30 – 12:00	Activity and group discussion
12:00 – 12:05	Resources/checklists and showcases
12:15 – 12:30	Close and feedback

## Presentations & Discussions: Key Points

### DAY 1

**First Session.** Introduction to SSHOC and the workshop's programme

**Speaker.** Cristina Magder

**Main points.** The workshop started with an interactive introductory session (via Mentimeter), which showed that the attendees were keen on learning about potential copyright issues in secondary data use and how these can be overcome. Since the workshop was given in the framework of the SSHOC project, the speaker presented the goals and the expected impacts of the SSHOC project and situated the workshop in relation to them.

**Links to materials.**

- [Presentation slides](#) (pp. 1-14)
- [Workshop recording](#) (min: 00:00–11:20)

---

**Second Session.** Overview and introduction to Copyright.

**Speakers.** Hina Zahid, Hannah Pyman

**Main points.** Hina Zahid gave an overview of Copyright: its connection to Intellectual Property rights (IPR), its history and key terms in order to provide the context for the workshop. In general, IPRs are granted to creators and owners of works that are the result of human intellectual creativity. Copyright is thus one type of IPRs covering creative works such as books, music, etc. as opposed to IPRs covering other types of works, such as patents, trademarks and registered designs.

Hannah Pyman explained that creative work covered by Copyright can also be of several types. She introduced key concept related to Copyright in the context of academic publishing and teaching. She has covered several important copyright issues that researchers need to consider when planning to publish their work, including manuscripts, in subscription based, hybrid or open access journals, such as *do I own the rights to my PhD, if it is only for classroom use, do I need to consider Copyright, can I use images I don't own in my research, etc.*

**Links to materials.**

- [Presentation slides](#) (pp.16-67)
  - [Workshop recording](#) (min: 14:40–1:26:35)
- 

**Third Session.** Overview of secondary data and copyright issues in secondary data use

**Speakers.** Maureen Haaker, Hina Zahid

**Main points.** Maureen Haaker explained what secondary data is, what are its sources and how to best discover secondary data. She has talked through some excellent examples of such data, for example microdata, census data, administrative data, business microdata etc. and concluded her talk pointing to a published [paper that emphasizes how terms 'primary' and 'secondary' data should be used carefully](#).

Building on the previous session, Hina Zahid highlighted the most common copyright issues in secondary data use. She has covered key aspects such as who owns the copyright; university or an employee, university or student, research funder or researcher? What protection does copyright offers? What are database rights? And what happens when the copyright expires? Attention was also given to the best practices to deal with the so-called orphan works for which the rights holders cannot be identified.

**Links to materials.**

- [Presentation slides](#) (pp.68-97)
  - [Workshop recording](#) (min: 1:28:36 –1:01:56)
- 

**Q&A Sessions.** Discussions facilitated by all presenters based on Padlet and Chat questions from the audience

**Facilitators.** Cristina Magder, Hina Zahid, Hannah Pyman, Anca Vlad

**Main points.** The topics covered by the presentations on the first day sparked a variety of discussions covering niche topics such as copyright implications for online conference papers/presentations, paradata and consent and text and data mining exemption. The audience has engaged well with the Padlet and having separate sessions to discuss more in depth and to providing further examples have proved beneficial and widely appreciated by the audience.

---

## **DAY 2**

**First Session.** Recap of the first day of the workshop and a short exercise

**Speaker.** Cristina Magder, Hina Zahid

**Main points.** Cristina Magder started Day 2 of the workshop with welcome notes and a recap of the first day, followed by an overview of the day. Hina Zahid facilitated an exercise consisting of questions and copyright scenarios in order to reactivate the knowledge acquired on the first day.

### **Links to materials.**

- [Presentation slides](#) (pp.1-9)
- [Workshop recording](#) (min: 0:00–6:15; the hands-on part was not recorded)

**Second Session.** Data licensing and copyright (exceptions and infringements, international context and social media)

**Speaker.** Cristina Magder, Hina Zahid

**Main points.** Cristina Magder gave a brief overview of common data licences and how copyright is inherently linked to licencing. The presentation included information on Open Licences such as Creative Commons, Open Data Commons and Government Licences to bespoke data licences used by national repositories and data archives. As data sharing is becoming more and more prevalent in the European research community, with some funders now mandating data sharing, the speaker also addressed the considerations when licencing data such as ownership (copyright does not equal ownership), contents of data (requirements linked to anonymization), type of users (commercial or non-commercial) and permission to share (sharing enable without or only with permission).

Hina Zahid followed with an overview of copyright exceptions and limitations including the exception of ‘fair dealing’ and orphan works and discussed the Copyright in the international context. The speaker underlined that copying the whole work would generally not be considered fair dealing, and while researchers are allowed to copy the whole work for their personal use, in most cases, they are not allowed to also share it. In addition, the speakers discussed possible reactions to potential infringements. Hina Zahid also presented the ‘Berne Convention’ which is a treaty that defines minimum standards of protection regarding the types of works protected, duration of protection, scope of exceptions, limitations and principles such as “national treatment” and “automatic protection”. The speaker concluded her presentation by discussing copyright and social media data: how the social media data is acquired and what are the copyright implications in using the data from these platforms, especially when it comes to data sharing and archiving for future use. The session ended with an interactive activity.

### **Links to materials.**

- [Presentation slides](#) (pp.10-54)
- [Workshop recording](#) (min: 6:20–1:15:20)

**Third Session.** Concluding presentation

**Speaker.** Hina Zahid, Cristina Magder

**Main points.** The concluding presentations from Cristina Magder included a recap of the workshops, presentation of a list of resources that can be useful in the context of copyright and a discussion about some cases on overcoming copyright issues in practice. Hina Zahid led the hands-on activity in which six data sources were included such as World Bank Data, Penn World Table, Microsoft Academics. She presented the sources and their terms and conditions with regard to data sharing. The workshop ended with a final Q&A session which addressed the remaining and new questions that were added to the Padlet as well as enabled the attendees to contribute to the discussions throughout the workshop.

**Links to materials.**

- [Presentation slides](#) (pp.55-69)
- [Workshop recording](#) (min: 1:15:25–1:31:40; the hands-on part was not recorded)

---

**Q&A Sessions.** Discussions facilitated by all presenters based on Padlet and Chat questions from the audience

**Facilitators.** Cristina Magder, Hina Zahid, Maureen Haaker

**Main points.** Following the day 1 programme the dedicated Q&A session allowed the facilitators to address questions from the audience in more detail and open the session for discussions. Several questions were around the use of licences for different type of data. Some attendees were specifically interested in the difference between social media data use and distribution. Other questions covered cross-national legislation, attribution statements and API data.

---

## Outcomes & Feedback

Copyright and other intellectual property rights (IPR) are challenging aspects that researchers have to face frequently. From publications and teaching to using published data researchers must always consider IPR implications on their work. While a one tier approach is not possible, the workshop introduced key concepts, and showcased examples to enable participants to better understand and manage IPR implications. Several open source tools have been presented to enable the participants to make better informed decisions. Since the topics covered were complex, information has been recapped via interactive hands-on exercises. Participants were also presented with concrete situations where IPR challenges have been overcome and presented with standardized approaches for their own work.

**Viewers' satisfaction.**<sup>31</sup> All of the viewers rated the workshop as excellent or very good (cf. Q1). For half of the respondents, the workshop matched their expectations, while for the second half it (greatly) exceeded their expectations (cf. Q4). Respondents commented that Copyright gained a lot of attention in the last years both with the funders as well as the scientific community (cf. Q3). They attended the workshop to get up-to-date information about rights and obligations that follow from the copyright laws. The respondents believe that the workshop will have immediate positive impact on their work since it provided a lot of practical information and links that empower them in their own work, but also provide a very helpful resource that can be easily transferred between colleagues in their organisation (cf. Q5):

Yes, I feel better equipped to deal with some of the situations we meet in our work. (Respondent No. 9, Q5).

Yes, I will use the acquired knowledge in developing a strategy to avoid problems when collecting research data and developing our database system. (Respondent No. 10, Q5).

The respondents appreciated the workshop because the content was presented in an accessible way, the exercises included real-life examples and the speakers provided highly informative answers to questions (cf. Q6):

The extent of knowledge in presentations was very vast, the Q&A was very helpful and there was no shame and no stupid questions, the event-holders were not in a hurry, enough time was allocated to each section, the presentations and recording is easily made available, the event-holders were responsive. No question was left unanswered. (Respondent No. 4, Q6)

**Feedback regarding the organisation.** The respondents felt that the workshop was excellently organised (cf. Q9). They learned about the event in different ways, but mainly from SSHOC Newsletter, social media and colleagues (cf. Q2). The suggestions for improvements included having more or slightly longer breaks, and possibly using the same code for all Mentimeter activities and trying to deliver all presentations in free speech instead of reading (cf. Q7). The appreciation of the audience regarding the organisation is maybe best conveyed via the following two citations:

It couldn't be more comfortable in my opinion. It was online and if you like you could actively participate or you could just sit and watch/listen. They stuck to the schedule nicely (one item was shortened) so you didn't feel rushed or anything. (Respondent No. 10, Q10)

I did like fact that they used some very pragmatic examples to show us how to deal with copyright ownership. Oh and the way they subtly repeated information was nice (ask what people think about something (without knowledge), talk about the subject, ask questions to check if it was understood, answer and explain why answer was correct or not). (Respondent No. 10, Q10)

---

<sup>31</sup> The post-event survey was filled out by 10 participants.

**Future work.** There are no immediate future events planned as a follow-up to this workshop since the project is coming to an end. However, the respondents' feedback shows that there exists an interest for similar future events.

I hope there will be follow-up sessions with more advanced topics. It would be useful if participants before the webinar could send in suggestions on what topics/issues they would like to discuss. (Respondent No. 5, Q8)

Despite the end of the SSHOC project, the UKDS will continue to provide trainings and support to the scientific community on the topic of copyright. For more information, please see [UKDS homepage](#).



---

# ANNEX 7: WORKSHOP REPORT: DATA MANAGEMENT PLANNING AND OVERCOMING CHALLENGES IN SOCIAL SCIENCES DATA SHARING

## Background

The online workshop — [Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing](#) — was held on 14<sup>th</sup> and 15<sup>th</sup> February 2022 and was organized by SSHOC T6.5. The workshop focused on data management planning and how it helps to achieve optimal handling, organising, documenting and enhancing of research data. Therefore, the workshop importantly contributed to the SSHOC objective of encouraging data re-use and promoting FAIR principles and their role in Data Management Planning (DMP).

The workshop provided a comprehensive overview of DMP and its importance, explaining the main challenges of data sharing in the Social Sciences and how these can be overcome by implementing best practises in DMP. The workshop enabled participants to gain a better general understanding of DMP and its main challenges and enable them to design a comprehensive, accurate DMP for their specific case that enables data sharing.

## Workshop Overview & Format

**Aim.** Data management planning has become a central concern for publicly funded research, with both funders and academic institutions acutely aware of the ethical and legal responsibilities that come with conducting rigorous and transparent research. This workshop therefore focused specifically on how to write data management plans to help potential users overcome the common problems associated with data sharing. The content of the workshop was intended for early career researchers and support staff. In addition to general considerations concerning DMP and the design of data management plans, special attention was given to the challenges of data sharing and techniques to overcome them. The workshop was designed to encourage active audience participation with interactive sessions where participants critically discussed case studies and personal experiences of data management in social sciences.

**Speakers.** The workshop was delivered by 3 speakers:

- **Cristina Magder** (UKDS) manages the UK Data Service's Data Collections Development and Research Data Management teams at the UK Data Archive. She leads the research data

management portfolio of support and training for the UK Data Service. Her main teaching interests are data management planning, sharing and archiving data with a specific focus on licencing, data quality assurance, disclosure risk assessment, and reproducibility.

- **Maureen Haaker** (UKDS) is a scholar in the field of qualitative methodology. She has worked with the UK Data Service for 10 years as a senior qualitative training officer and is a lecturer in the School of Social Sciences and Humanities at University of Suffolk. Her work has involved the ingest and digitisation of qualitative data, development of training for the management, deposit, and re-use of qualitative data, and the re-use of qualitative datasets for teaching and learning. She is a key contributor to Sage's *Managing and Sharing Research Data* (2nd edition) and *Qualitative Secondary Analysis*. She is currently completing a PhD in Sociology at University of Essex.
- **Anca Vlad** has been part of the collections development and data publishing team at the UK Data Archive for the past 6 years. She has been assisting researchers and data depositors with preparing their data deposits and delivering training on data management, archiving and open access publishing for different audiences.

**Organisers.** The workshop was organised as a stand-alone online event in cooperation between T6.5 members CESSDA/UKDS and CLARIN/UL-FF.

**Participation.** There were 36 participants at the workshop.<sup>32</sup> Almost all participants (90%) came from EU countries, both EU and non-EU, while others (10%) came from other continents (Brazil, Canada, India, etc.). The audience of the livestream covered six stakeholder categories identified in D6.1. The audience (95%) mainly identified as belonging to the following stakeholder categories: "Researchers, Research Networks and Communities" (50%), "Research Libraries and Archives" and "Universities and Research Performing Institutions" (20% each). The remaining 10% of the audience represented "Research and e-Infrastructures", "Policy Making Organisations" and "Private Sector and Industry Players". Considering the main scientific domain of the participants, the audience was very diverse. Sociology was the most prominently represented field of expertise in the audience (30%), while around 10% of the audience worked either in Education or in Management. The other half of the audience represented various SSH fields in smaller shares, e.g., Communication Sciences, Cultural Heritage, Geography, History, Linguistics, Political Science or Psychology.

---

<sup>32</sup> There were 36 participants on Day 1 and 31 on Day 2. The number consists only of viewers of the livestream. It should be noted that the overall reach is higher since this number does not include the views of the workshop recording which is available on the SSHOC YouTube channel and promoted by SSHOC and its partners.

**Brief summary of the event structure.** The workshop was a virtual, two-day event. The online format was chosen due to the restrictions linked to Covid- 19 pandemic, while the programme was split over two days to provide an engaging experience for the participants, as the online format is more demanding than an in-person event.

The first day of the workshop was divided into five parts: an introductory session dedicated to the SSHOC project; a session where the speaker presented the standards of data management plans and ethical and legal considerations of data management; a session on how to prepare data for publishing and how to curate it; and finally an unrecorded Q&A session. The concluding session summarised the day and gave an overview of workshop programme for the next day. The second day was divided into four parts: a short introductory session dedicated to the recap of the first day; a longer session in which the speakers presented on the topic of data collections and evaluation of existing sources; a short breakout room exercise on research scenarios; a short unrecorded Q&A session and a concluding session where the speakers talked about common challenges in data sharing and how to overcome them. On both days there were various prompts to engage the audience (live polls via Mentimeter, Poll Everywhere) as well as practical exercises to encourage successful knowledge transfer. There was also time for discussion on both days, while the speakers also answered questions throughout the workshop.

## **PROGRAMME**

### **Day 1**

10.00-10.15 Welcome and Intro

10.15-10.45 Data Management Planning in Social Sciences

10.45-11.00 Ethical and Legal Considerations

11.00-11.15 Break

11.15-12.00 Management and Curation of Data

12.00-12.20 Q&A Session

12.20-12.30 Close of day

### **Day 2**

10.00-10.10 Welcome Day 2

10.10-10.30 Data collections and assessment of existing sources

10.30-11.00 Exercise (Research Scenarios)

11.00-11.15 Group Discussion

11.15-11.30 Break

11.30-12.00 Common challenges in data sharing: lessons learned

12.00-12.20 Q&A Session

12.20-12.30 Close of workshop

## Presentations & Discussions: Key Points

### DAY 1

**First Session.** Introduction to SSHOC and overview of the workshop

**Speaker.** Cristina Magder

**Main points.** The workshop started with a presentation of the SSHOC project in which the speaker introduced the objectives and the expected impact of the SSHOC project and situated the workshop in relation to them. This was followed by a short introductory session (via Poll Everywhere) which showed that participants were familiar with secondary data analysis but were keen to learn more about it. A brief overview of the workshop programme was also given.

**Links to materials.**

- [Presentation slides](#) (pp. 1-14)
  - [Workshop recording](#) (min: 00:00–13:00)
- 

**Second Session.** Structure and standards of data management plans and ethical and legal considerations of data management

**Speakers.** Cristina Magder

**Main points.** Cristina Magder opened the workshop with a comprehensive look at the structure and standards of data management plans (DPMs). In particular, a practical checklist was presented to help strategically address the most important issues within DPMs. Specific attention was also given to the FAIR (Findable, Accessible, Interoperable, Reusable) principles and their role in data management planning, with the speaker presenting the Horizon Europe DMP template as an example to illustrate the emphasis on FAIR and how good data management planning enables researchers to meet these requirements.

With regard to the ethical and legal aspects of data management, the speaker referred to the obligation of confidentiality and the standards set by GDPR and national data protection laws, which can impact how consent is obtained and how personal data are stored and accessed.

**Links to materials.**

- [Presentation slides](#) (pp.15-66)
  - [Workshop recording](#) (min: 13:10–58:00)
- 

**Third Session.** Management and Curation of Data**Speakers.** Anca Vlad

**Main points.** The speaker started with presenting a three-step strategy to protect research participants. This includes thoroughly obtaining consent before collecting data, having a clear plan for anonymising and further handling the data, and controlling who can access the data. These three complementary strategies form a comprehensive plan to protect identities and meet data management expectations. The speaker also presented strategies to ensure that data is kept secure, such as a digital backup strategy that's securely shared and disposed of in a safe manner (e.g. in collaborative environments). Finally, the speaker addressed costing for research projects and introduced the [UK Data Service Data Management Costing Tool and Checklist](#), an open-source resource created to help formulate the costs of managing research data before research begins. The session ended with a Menti exercise covering the concepts and information on data management and curation, as well as the ethical and legal context.

**Links to materials.**

- [Presentation slides](#) (pp.67-105)
  - [Workshop recording](#) (min: 58:01 –1:51:00)
- 

**Fourth Session.** Q&A Session**Speakers.** Cristina Magder, Anca Vlad, Maureen Haaker

**Main points.** The audience was encouraged to include questions either in the Zoom chat or on the prepared Padlet. The discussions covered best practices for sharing sensitive data sharing, consent as lawful basis for processing vs ethical requirement, responsibilities for DMPs, and ontologies, metadata schemas and controlled vocabularies.

**Fifth Session.** Recap of the day**Speakers.** Cristina Magder

**Main points.** The speaker summarised the content of the day and gave an overview of the programme following the next day.

**Links to materials.**

- [Presentation slides](#) (pp.106-111)
- [Workshop recording](#) (min: 1:51:00–1:54:54)

## **DAY 2**

**First Session.** Overview of the important points from Day 1 and a short exercise

**Speaker.** Cristina Magder

**Main points.** Cristina Magder started Day 2 of the workshop with welcome notes and a recap of the crucial points from first day. This was followed by an overview of the day and a short exercise to better engage with the audience and understand their challenges in data sharing.

**Links to materials.**

- [Presentation slides](#) (pp.1-14)
- [Workshop recording](#) (min: 0:00–7:20)

**Second Session.** Data Collections and Evaluation of Data

**Speaker.** Maureen Haaker

**Main points.** The speaker presented what archived datasets look like, how to check them for their suitability for one's own research and how to find data. Emphasis was placed on the ethical responsibility to use existing data, i.e. repositories should be thoroughly checked for relevant resources that go hand in hand with the aims of the research project before starting a creation process for a new resource. When searching for suitable data, the speaker encouraged participants to use the [CESSDA Data Catalogue](#) or search for data papers in data journals, such as the [Research Data Journal for Humanities and Social Sciences](#). The exercise covered the key issues to be addressed in a DMP for three research scenarios. The discussion revolved around consent for certain types of data, proprietary formats and sharing among research teams, and concerns about sharing.

**Links to materials.**

- [Presentation slides](#) (pp.15-42)
- [Workshop recording](#) (min: 7:20–37:52; the hands-on part was not recorded)

### **Third Session.** Research Scenarios Exercise

**Speakers.** Cristina Magder, Anca Vlad, Maureen Haaker

**Main points.** A hands-on worksheet was prepared for the audience. The aim of the exercise was to identify challenges in sharing the presented data and techniques to overcome these as presented during the workshop in small groups, or if more comfortable individually. Participants raised core issues, including consent for specific types of data, proprietary formats and sharing among research teams, and concerns related to sharing. The worksheet also provides example text for core DMP sections which have been very appreciated by the audience.

#### **Links to materials.**

- [Presentation slides](#) (pp.36-42)
- 

### **Fourth Session.** Challenges to sharing social science data

**Speaker.** Cristina Magder

**Main points.** The speaker encouraged participants to share (via Mentimeter) their concerns about data sharing. The speaker noted that the reported difficulties were very similar to the challenges raised in [2018 Springer Nature Whitepaper: Practical challenges for researchers in data sharing](#). These challenges include organising data in a presentable and useful way, copyright and licensing uncertainties, not knowing which repository to use, lack of time to deposit data and costs of sharing data. The speaker noted several strategies of identifying appropriate and responsible repositories, the range of licenses available for depositing data, and the extent to which participants want their data shared. The speaker demonstrated that the barriers to data sharing could often have quite straightforward solutions, especially at the level of publishing data. The workshop ended with a Q&A session and the concluding remarks from the speaker.

#### **Links to materials.**

- [Presentation slides](#) (pp.55-69)
- [Workshop recording](#) (min: 37:55–1:13:56)

## Outcomes & Feedback

In a data driven world, with the Open Science Agenda leading more and more research, Data Management Planning has become a core skill needed by researchers. Drawing attention to the importance of good DMPs and presenting to the research community how to implement FAIR principles throughout the research data lifecycle is crucial for scientific community since this enables better, safer and more informed data sharing. Key concepts have been included in the workshop such as data and

metadata standards, legal and ethical considerations, management and curation of data, security of data and assessing existing data resources. Given the variety and complexity of the information covered, issues have been recapped via interactive exercises and time has been given to answer questions from the audience. The workshop provided participants with templates and examples and signposted to various open resources that further facilitate the development of good DMPs and the sharing of FAIR data. Finally, participants were presented with real life example where data sharing challenges have been overcome by implementing good DMP practices.

**Viewers' satisfaction.**<sup>33</sup> All of the viewers rated the workshop as excellent or very good (cf. Q1). For all except one of the respondents, the workshop met their expectations (cf. Q4). Respondents attend the event mainly to learn about the topic of Data Management Planning and to improve their practical skills regarding DMP (cf. Q3). The respondents stated that the speakers explained well the importance of DMP which had an encouraging effect on those who felt that DMPs are too complicated to embark on this road by themselves (cf. Q5). Furthermore, the respondents felt that the workshop will have a positive impact on their work, because it provided a lot of practical information that are helpful when drafting DMPs, as well as many useful resources that can be easily shared among colleagues:

Yes. I can use many practical information, the links shared are going to be of great use to me and my institution, and I got many additional resources to read through. (Respondent No. 5, Q5).

Respondents appreciated the workshop because the exercises provided good practical examples, but also because of the speakers who were described as extremely knowledgeable and friendly with excellent communication skills (cf. Q6). After day one, some informal feedback was gathered via Mentimeter (see image below) – the participants found the workshop enjoyable because it gave a useful overview of the topic, was engaging due to exercises via Mentimeter and the Q&A sessions and it was practical with examples provided.

---

<sup>33</sup> The post-event survey was filled out by 9 participants.



## What have you enjoyed most of today?

 Mentimeter


**Feedback regarding the organization.** All respondents felt that the workshop was excellently organized (cf. Q9). They learned about the event in different ways, but mainly from newsletters (CESSDA and SSHOC newsletter), colleagues and browsing the web (cf. Q2). Suggestions for improvement included that speakers could give even more real-life examples and that some presentations should be less rushed (cf. Q7). Some even expressed a wish that the workshop was longer (cf. Q7). The appreciation of the audience regarding the organisation is best conveyed via the following citation:

No. Workshop was well time-managed, had a good mix between listening to presentations and actively doing exercises, responses to chat and Padlet questions provided opportunity for review, considerate turning off of recording when appropriate, and when people inadvertently left their mic on (myself included!) organizers were able to turn off. (Respondent No. 6, Q10)

**Future work.** As the project is coming to an end, no immediate follow-up events to this workshop are planned. However, feedback from respondents indicates that there is interest in similar events:

I know this workshop was targeted at Social Sciences, but I joined anyway because I needed some pointers in where to start. I would love to join a workshop for Digital Humanities with a focus on whether it is possible to share your research when it is derived from copyrighted material. (Respondent No. 6, Q7)

Despite the end of the SSHOC project, the UKDS will continue to provide trainings and support to the scientific community on the topic of copyright. For more information, please see [UKDS homepage](#).

---

# ANNEX 8: WORKSHOP REPORT: DATA CITATION IN PRACTICE

## Background

The online workshop — [Data Citation in Practice](#) — was held on 15<sup>th</sup> June 2021 and was organised by SSHOC T6.5 in cooperation with SSHOC WP3 and WP7. The workshop focused on different citation practices and discussed current issues and different solutions for efficient data citation. The workshop also presented the work done in the SSHOC Task 3.4 which is especially focused on how to make data citations actionable. The content of the workshop was aligned with the SSHOC objective to promote Open science by addressing one of the crucial points in ensuring FAIR data. The workshop enabled participants to expand their knowledge about the importance of appropriate citation practices and helped them acquire necessary skills to transfer best practices to their workflows.

## Workshop Overview & Format

**Aim.** FAIR data are the pillar of Open Science which is at the core of the SSHOC project. In order to ensure findability of data and other resources, it is crucial to provide easy-to-use recommendations for data citation in the SSH domains. However, there are different needs and practices with regard to citation in the broad range of SSH domains. This is why the workshop aimed to present solutions for efficient data citation from three different perspectives, namely in the context of CLARIN Virtual Collection Registry, Czech LINDAT Center and French CoCoon Center. The content of the workshop was adapted for professionals from a wide range of disciplines with a focus on researchers and repository managers. Apart from addressing general considerations with regards to citing practices, the speakers discussed challenges linked to producing citations, and suggested some approaches to overcome them. The hands-on exercises were designed in such a way that participants could familiarise themselves with good and bad citation protocols.

**Speakers.** The workshop was delivered by 4 speakers:

- [Dieter Van Uytvanck](#) (CLARIN ERIC) graduated in Informatics (2002, Ghent University) and Language and Speech Technology (2007, Radboud University Nijmegen) and has been involved in technical infrastructure building since CLARIN's preparatory phase in 2008. He currently serves as the Technical Director and Vice Executive Director of [CLARIN ERIC](#).
- [Pavel Stranak](#) (LINDAT) holds a degree in Czech Philology and Computational Linguistics and currently works at the Institute of Formal and Applied Linguistics of Charles University. He is also

scientific secretary of [LINDAT/CLARIAH-CZ](#). He specializes in lexical semantics and computational lexicography, but is also interested in the reliability of annotations, machine translation and application of NLP technology in everyday life.

- [Nicolas Larrousse](#) (Huma-Num/CNRS) is the Vice Director of [TGIR Huma-Num](#) and head of its Coordination of National and International User Communities group. He is an IT-specialist particularly interested in interoperability issues and long term preservation. He is involved in several European infrastructures and projects.
- [Tatsiana Yankelevich](#) (LIBER, SSHOC T6.5): is the SSHOC representative who moderated the workshop. Tatsiana currently serves as Training Coordinator at LIBER. With BA in Business Administration (LCC International University, Lithuania) and an MSc in International Relations and Diplomacy (Leiden University and the Netherlands Institute of International Relations Clingendael, the Netherlands), she has previously worked for a number of human rights non-governmental organisations, the United Nations and the Organisation for the Prohibition of Chemical Weapons.

**Organisers.** The workshop was organised as a stand-alone online event in cooperation with partners from T6.5 and WP3 and WP7 of the SSHOC project.

**Participation.** There were 36 viewers of the workshop.<sup>34</sup> The majority of the viewers (77%) came from European countries, both EU and non-EU, while the remaining participants came from the USA. The audience of the livestream covered five stakeholder categories identified in D6.1. Almost all (92%) participants represented the scientific and research community (“Universities and Research Performing Institutions” – 31%, “Research and e-infrastructures” – 23%, “Research Libraries and Archives” – 19% and “Researchers, Research Networks and Communities” – 19%). The workshop was attended also by some participants (8%) belonging to the “Private Sector and Industry Players” category. The categories “Research Funding Organisations”, “Policy Making Organisations” and “Civil Society and Citizen Scientists” were not represented in the audience. There were many scientific disciplines represented in the audience from Arts and Humanities, Social Sciences and Education, to Law and Management. The most prominent group (30%) included participants from the field of Linguistics.

**Brief summary of the event structure.** The workshop Data Citation in Practice was organised as a virtual event due to the restrictions linked to Covid-19 pandemic. The workshop lasted for one hour and ten minutes and was divided into five parts: a short one dedicated to the presentation of the SSHOC project; a longer one where three main speakers presented on the topic of the workshop; and a hands-on session with discussion. The speakers answered questions throughout the workshop.

---

<sup>34</sup> The number consists only of viewers of the livestream. It should be noted that the overall reach is higher since this number does not include the views of the workshop recording which is available on the SSHOC YouTube channel and promoted by SSHOC and its partners.

## Presentations & Discussions: Key Points

**First Session.** Introduction to SSHOC

**Speakers.** [Tatsiana Yankelevich](#)

**Main points.** The workshop was given in the framework of the SSHOC project. The speaker presented the goals and the expected impacts of the SSHOC project and situated the workshop in relation to them.

**Links to materials.**

- [Workshop recording](#) (min: 2:05–5:05)
- 

**Second Session.** Data Citation with Virtual Collections.

**Speaker.** [Dieter Van Uytvanck](#)

**Main points.** The speaker introduced CLARIN Virtual Collection – a coherent set of links to digital objects (e.g., annotated text/video) that can be easily created, accessed and cited. These virtual collections are accessible through a service – the CLARIN Virtual Collection Registry – where it is possible to register and publish virtual collections. All the collections get a Persistent Identifier (DOI or handle) and are provided with metadata which allow citation. There exist several methods of citing a virtual collection in this CLARIN service, such as (1) using a cite button (this option provides the user with a BibTeX snippet that can easily be copy-pasted), (2) via the browser plugin Zotero, or (3) using the DOI look-up function in reference manager that have this function.

**Links to materials.**

- [Presentation slides](#)
  - [Workshop recording](#) (min: 6:10–26:25)
- 

**Third Session.** Data Citations at LINDAT/CLARIAH-CZ.

**Speaker.** [Pavel Stranak](#)

**Main points.** The speaker addressed the issues with citations of data sets in the Czech LINDAT Center. The repository follows the Force-11 data citation principles and it provides formatted citation in simple text (ready to copy and paste) and the option to create BibTeX snippets. Despite these user-friendly citing methods, certain issues with citations remain. One such problem is that the handle system used in LINDAT Repository doesn't support altmetrics, CrossRef and other services, which are widely used because they allow for the citation to be found and formatted more easily. Luckily, this issue could be resolved by improving and adding support for Citeproc JSON format – this would allow the repository system to work properly with reference management software (i.e. Zotero or Mendeley) and would also offer more citation styles.

**Links to materials.**

- [Presentation slides](#)

- [Workshop recording](#) (min: 26:50–46:30)
- 

**Fourth Session.** Cocoon: a French repository dedicated to oral resources.

**Speaker.** [Nicolas Larrousse](#)

**Main points.** Not all resources are created equal and there are some specialities when citing oral resources. These were presented by [Nicolas Larrousse](#) (Huma-Num/CNRS) who introduced the French CoCoon Center and the different types of data in their repository, such as recordings (audio and video) and associated annotations (transcriptions, translations and measures). CoCoon repository uses different types of PIDs (DOI, ARK and PURL) for different purposes like OAI-PMH identifiers, long term preservation etc. The user can cite the data in a variety of citation styles (APA, MLA, BibTeX and others). A very interesting and useful method of citing offered in the repository is the citing of the specific part of the recording. This is done by using the specifications from the [Media Fragments URI](#) which is a recommendation from the World Wide Web Consortium (W3C). In practice, the researcher citing a part of the recording just enters the timestamp (e.g., #t=4,9) directly after the DOI, and the link automatically redirects to the exact part of the recording (e.g., the 4th second of the cited recording).

**Links to materials.**

- [Presentation slides](#)
  - [Workshop recording](#) (min: 46:50–1:01:56)
- 

**Fifth Session.** Hands-on session and discussion.

**Speaker.** All speakers

**Main points.** This part of the workshop was dedicated to hands-on exercises allowing participants to observe and discuss some consequences of using good and bad citation protocols on the examples from Zenodo, LINDAT/CLARIAH-CZ and Parlaritaliano.it. Participants were divided into three groups, where they had to perform three tasks related to citation. The first two tasks explored the pros and cons of automatic generation of citations, while the third task looked at best practices when automatic generation is not available. The workshop concluded with a short evaluation of the hands-on exercises and useful tips about using reference management software, such as Zotero, and repositories, such as Zenodo.

**Links to materials.**

- [Workshop recording](#) (min: 1:02:08–1:11:21; the hands-on part was not recorded)

## Outcomes & Feedback

Data citation is an important aspect of ensuring FAIR data, but it can often be challenging, which is why easy-to-use recommendations as well as discussion about possible issues and solutions are crucial. Recognising that there is no one-size-fits-all solution, this workshop provided participants with information on various citation techniques and tools offered by different repositories. Participants were presented with several possibilities for effective data citation, which they were able to apply in the practical session. Participants were also able to observe and address some of the consequences of using good or bad citation methods on specific examples.

**Viewers' satisfaction.**<sup>35</sup> All of the viewers rated the workshop as Excellent or Very good (cf. Q1). For most, the workshop matched their expectations and for some, their expectations were even exceeded (cf. Q4). The respondents stated that the workshop will have a positive impact on their work since it underlined the importance of citation which is the last step of sharing data that is often overlooked, but also introduced some new concepts and solutions which will help improve the citation practices in their organisations (cf. Q5). They appreciated the workshop because the presentations were very practical and easy to follow, but also because they could link theory and practice through exercises and discuss problems and solutions with the experts (cf. Q6).

**Feedback regarding the organisation.** The respondents felt that the workshop was well or even excellently organised (cf. Q9). They learned about the event in different ways, but mainly from SSHOC Newsletter, social media and colleagues (cf. Q2). The only suggestion for improvements pertained to the time allocated for exercises which could be longer (cf. Q7).

**Future work.** The respondents highlighted that data citation practices are constantly evolving which is why such events carry great value for the community since it enables the promotion of best practices across the scientific fields (cf. Q3 & Q7). In light of this, one respondent suggested that the community could benefit from an event which would deliver a general overview of data citation methods (cf. Q7). Since this workshop will be followed by a webinar on a similar topic, we see it as an opportunity to accommodate this wish. In conclusion, the widely adopted FAIR approach to data as well as the positive feedback following this event clearly show the pertinence of the topic and underlines the need for similar events in the future.

---

<sup>35</sup> The post-event survey was filled out by 5 participants.

---

# ANNEX 9: WORKSHOP REPORT: USING CORPORA FOR IMPLEMENTING VALIDATION

## Background

This report concerns the second workshop organised by T6.5 of the SSHOC project in cooperation with external lecturers on the topic of Data Science for the Social Science and Humanities.

The workshop/masterclass — [Using Corpora for Implementing Validation. Workflows that combine quantity and quality](#) — was held on 30 September 2019 at the [CLARIN Annual Conference 2019 in Leipzig](#), Germany and was designed for political and social scientists who are interested in using large text collections in their research. This event aimed to contribute to two major SSHOC objectives: developing relevant and applicable tools for specific user communities and empowering them to actively use the provided tools. The masterclass addressed the challenges that political and social scientists encounter when confronted with the need to validate their findings obtained with quantitative analysis of corpora.

## Masterclass Overview & Format

**Aim.** Despite the potential of quantitative corpus analysis, it is clear that findings need to be validated. However, validation is inhibited by technical restrictions. Usually, the solution includes costly and time-consuming development of dedicated software which can then only be used by a small group of scientists. In order to promote and implement validation on a larger scale, researchers in Social Sciences and Humanities working on text corpora need open-access and easy-to-use workflows that combine quantitative and qualitative approaches. Providing researchers with accessible tools is also central to the SSHOC project and the [polmineR R-package](#) answers exactly this challenge.

Therefore, the aim of this masterclass was to start a discussion around the need for validation and the potential approaches to validation, as well as to demonstrate the use of the polmineR R-package using the UN General Assembly corpus.

**Speakers.** The masterclass was taught by [Andreas Blätte](#), head of the [PolMine](#) project and developer of the [polmineR](#) R-package, and [Christoph Leonhardt](#) (both University of Duisburg-Essen). Kristina Pahor de Maiti (Faculty of Arts, University of Ljubljana; T6.5 member) gave a short introductory presentation about the SSHOC project.

**Organisers.** Due to the co-location of the event at the CLARIN Annual conference, the event was organised by the SSHOC Workshops Team from T6.5 in cooperation with the Organizing Committee of the CLARIN Annual conference 2019.

**Participants.** Political and social scientists interested in using text corpora in their research formed the main target audience. The 4 participants, however, represented also other disciplines, namely methods and statistics, computer science, history and linguistics. Three different stakeholder groups were covered: researchers, research networks and communities; research and e-infrastructures; and research libraries and archives.

**Brief summary of the event structure.** The masterclass was an all-day event prepared as a combination of a lecture by Andreas Blätte and Christoph Leonhardt, practical presentation of the PolmineR R-package, a hands-on session and a discussion.

The demonstration of the polmineR R-package was based on the following three scenarios:

1. Validating the results obtained from dictionary-based sentiment analysis and classification.
2. Validating the results of Latent Dirichlet Allocation (LDA) topic modelling.
3. Giving substantial meaning to the results of co-occurrence analyses.

## Presentations & Discussions: Key Points

**Introductory Session.** Presentation of the SSHOC project.

**Speakers.** Kristina Pahor de Maiti (Faculty of Arts, University of Ljubljana; T6.5 member)



**Main points.** The participants were informed about the SSH Open Cloud: why it is being developed, its main objectives and the broader context. The work and aims of WP6 were also presented and the participants were invited to engage directly with SSHOC (e. g., signing-up for the newsletter, joining the training community).

---

**Main Session.** A combination of lecture, showcase of polmineR features, hands-on activities and open discussion.

**Speakers.** [Prof. Dr. Andreas Blätte](#) and [Christoph Leonhardt](#)

**Main points.** The speakers presented common research strategies, talked about why implementing validation remains a technological frontier, mapped out various validation requirements and offered suggestions on how to satisfy the need for validation.

Andreas Blätte suggested the term *quanlification* for approaches that combine quantitative and qualitative approaches to corpus analysis, in order to achieve valid and sound research results. Due to technical restrictions, validation by *quanlification* is challenging to achieve and therefore a set of scenarios and workflows implemented using the polmineR package (developed by Andreas Blätte) were presented as a potential way forward. Topics covered by these workflows were counting, co-occurrence analysis, sentiment analysis, text classification, and of Latent Dirichlet Allocation (LDA) Topic Modelling.

Given the various disciplinary backgrounds of the attendees – ranging from computer and social science to humanities – these workflows were introduced with a focus on the validation of the output rather than on the production of code. However, participants were given ample opportunity to experiment with the PolmineR R-package in order to develop experience with the implementation of validation strategies.

In the course of the day, the participants intensively discussed the possibilities and limits of validation. A shared understanding emerged that the need to integrate quantitative and qualitative approaches to corpus analysis is central to these endeavours. Validating algorithmically derived findings of quantitative

approaches based on the initial text is necessary for a more complete insight into both the data and what a method actually measures, ensuring intersubjective and valid research.

counting words, contexts have to be taken into account. When calculating co-occurrences, outputs should be filtered by their actual semantic meaningfulness. Sentiment analyses should take into account both the complex nature and ambiguity of human speech and hence be evaluated carefully. And machine learning approaches need to be checked by looking back at the initial data.

**Links to materials:** [Presentation slides](#).

## Outcomes & Feedback

**Participant satisfaction.** All participants rated the masterclass as very good and the great majority indicated that it matched their expectations (cf. Q1 and Q4). It is interesting that one of the participants joined the masterclass to get to know the SSHOC project, while the others wanted to learn more about validation and the R programming language. In general, participants were eager to know more about the SSHOC project and were enthusiastic about following its developments. Most participants indicated that they will use the acquired knowledge in their future work, which is represented also in the answer of the Respondent No. 1 (cf. Q5):

» Yes, future work will focus more on implementing ways to validate quantitative results. «

The participants enjoyed the discussion and the interactive part of the masterclass, although some of them wished for more hands-on activities (cf. Q6 and Q7).

**Feedback regarding the organisation.** All participants thought that the event was well organised (cf. Q9) but pointed out some possible improvements (cf. Q10, Respondent No. 3):

» The room could be prepared in advance, otherwise all was fine. It was nice to have the opportunity to meet other participants of the conference. «

Respondents' answers also show that SSHOC channels could be more heavily used for dissemination purposes. Participants either heard for the event through the CLARIN mailing list or stumbled upon it on the internet. Another challenge that was identified during this masterclass (and which needs to be addressed in the organisation of future events) is the number of people who registered but did not

attend. The number of participants was half of those who applied to join. We suggest a pre-workshop survey asking for the background and motivation of the applicant as one possible solution to the identified problem. This information will also help organisers and lecturers prepare an event that better suits the needs of the participants.

**Future work.** The [polmineR R-package](#) provides a tool which offers both qualitative and quantitative approaches to corpus analysis, always allowing to reconstruct the full text. The discussion at the end of the session offered a great opportunity to elaborate on the package's design by presenting workflows which live up to these standards. Based on the encouraging feedback from the participants, a SSHOC webinar is planned in spring 2020, where Andreas Blätte will present the potentials of polmineR for validation purposes to a wider audience from the Social Sciences and Humanities domain.

---

# ANNEX 10: WORKSHOP REPORT: EXPLORATION OF SOCIETY THROUGH THE LENS OF LABOUR MARKET RELATED DOCUMENTATION – CBAQUEST AND PARLAMINT – EXPLORING SOCIETAL ISSUES THROUGH COMPARABLE CORPORA OF PARLIAMENTARY DEBATES

## Background

The two workshops — [Exploration of Society Through the Lens of Labour Market Related Documentation – CBAQuest](#) and [ParlaMint – exploring societal issues through comparable corpora of parliamentary debates](#) — were held from 19 to 28 May 2021 as part of the [Helsinki Digital Humanities Hackathon](#) organised by SSHOC T6.5 and T3.2, CLARIN, DARIAH and the University of Helsinki and Aalto University. The 10-day hackathon enabled dynamic knowledge transfer in interdisciplinary teams and resulted in research projects with concrete outcomes.

The workshop contributed to SSHOC objectives and its commitment to further Open Science by reusing tools and resources developed or enriched in the SSHOC project in order to produce new knowledge. In this way, the workshops also promoted the FAIR principles and helped spread good practices with regard to acknowledging previous work. But maybe most importantly, due to their interdisciplinary nature, the workshops offered unique opportunity to bridge the interdisciplinary gap, and highlighted the importance of group work.

## Workshop Overview & Format

**Aim.** The main aim of the Helsinki Digital humanities Hackathon, an annual event, is to offer a space to SSH researchers and computer/data scientists to experience intensive, interdisciplinary work in small groups with the aim of solving concrete research questions. This enables the participants from different backgrounds to improve their ability to work in groups, gather new skills with regard to inter-disciplinary communication and explore the possibilities for novel research problems that can only be solved in such interdisciplinary teams.

In the CBAQuest workshop, the participants used text mining tools to explore employment negotiations and their outcomes in order to ultimately help workers, trade unions, employers and policy makers to know more about labour rights. Experts from SSHOC helped the participants dig into the WageIndicator Collective Agreements Database – a resource enriched in the SSHOC project – and explore how different aspects of working conditions are addressed around the world.

In the ParlaMint workshop, the group of researchers focused on the comparison of parliamentary debates before and during COVID-19 pandemic across Europe – the ParlaMint corpus developed by SSHOC partners – from a linguistic, sociological, politological and computational perspective. The group's objective was to learn how to use comparable parliamentary corpora from various European countries that are annotated with metadata, such as speaker and session information, and linguistic annotations, such as morphosyntactic and named entity tags, for studying societal issues caused by the COVID-19 pandemic.

**Group leaders.** The two SSHOC-related workshops were led by:

- [Daniela Ceccon](#) and Stefano Ceccon (the CBAQuest group), and
- [Ajda Pretnar](#) and [Matej Klemen](#) (the ParlaMint group).

**Organisers.** The workshop was organised in cooperation with partners in SSHOC T6.5 and T3.2, CLARIN, DARIAH and members of the University of Helsinki and Aalto University: [Mikko Tolonen](#), [Eetu Mäkelä](#), [Jukka Suomela](#), [Jouni Tuominen](#).

**Participation.** There were 65 participants of the hackathon. Given the hackathon vision, the working groups are kept small to ensure stimulating environment for everyone. Therefore, the participants were divided into 7 groups according to their preferred topic. Those groups were led by 20 experts. The ParlaMint group consisted of 6 members and the CBAQuest of 13 participants. The great majority of the participants represented students and post-doctoral researchers. The main stakeholder category represented is thus "Researchers, Research Networks and Communities". Participants came from EU and non-EU countries (65%) and from countries outside Europe (35%), such as the USA, Canada, China, India, Pakistan, Nigeria, etc.

**Brief summary of the event structure.** Helsinki Digital Humanities Hackathon consisted of three introductory meetings prior to the beginning of the hackathon and eight days of intensive group work during the actual hackathon. The intensive period included work in small groups and regular updates which were shared with the entire hackathon community. The format of the group work was flexible which means that the group members were working either individually, in pairs or smaller groups according to the task at hand. Continuous communication was achieved through a communication platform. The hackathon ended with a final 15-minute presentation of the results by each group which was followed by questions and discussion.

## Presentations & Discussions: Key Points

**Workshop.** CBAQuest: Exploration of society through the lens of labour market related documentation

**Leaders.** [Daniela Ceccon](#) and [Stefano Ceccon](#)

**Main points.** The CBAQuest team explored the feasibility of assessing the ‘worker-friendliness’ of collective bargaining agreements in order to find new ways of understanding agreements and to contribute to improving global labour market transparency. To this end, the team produced [a prototype of a digital tool](#) that offers visualized information about CBAs to anyone interested in the documents governing the lives of workers.

The worker friendliness of collective labour agreements was rated by considering the following measurements:

- Equality is evaluated through 4 indicators that fall under gender equality trigger: gender equality, discrimination, sexual harassment and grievance procedure;
- Overtime and annual leave is evaluated by checking whether there are regulations on overtime, whether there is travel allowance provided, and whether the number of days of annual leave after 1 year of working is above the international standard of 15 working days.
- Text accessibility is evaluated through 3 indicators: concreteness, readability, lexical density which provide the information about how easy it is for the workers to understand the contract.

The project identified a number of ways that the ‘worker-friendliness’ of agreements might be measured, making use of text mining methods to analyse and score agreements on various indicators. By using and visualising these scores, the team has been able to find new ways of evaluating agreements at a glance, in ways that might facilitate understanding of these agreements for labour market researchers and workers in general. However, a number of challenges and limitations have been identified that invite further research into the secrets of collective bargaining agreements.

### Links to materials.

- [Final output report](#)

---

**Workshop.** Parliamentary debates in the COVID-19 times

**Leaders.** [Ajda Pretnar](#) and [Matej Klemen](#)

**Main points.** The research questions of the ParlaMint group focused on the identification of differences and similarities in parliamentary debates on the COVID-19 pandemic across Italy, Poland, Slovenia and the UK. To this end, the group first analysed the country-specific data, and then compared the results across countries as well as the data from the pre-COVID-19 period. They also mapped the COVID-19-related debates in time and compared them to the epidemiological situation in each country.

For each country, they asked:

- How do speeches on COVID-19 differ from regular debates?
- Which topics arise in COVID-19 debates? Which topics are shared between the countries and which are country-specific?

- Do the debates highlight any major shifts in topics or priorities over time?
- What is the frequency of COVID-19-related debates over time, and is there any connection between debates and COVID-19 cases reported?

Given the force with which the pandemic swept through the countries, it is not surprising that the datasets exhibit high similarity when looking at the top 20 COVID-19-related keywords with respect to the pre-COVID-19 period for each country and multiple overarching themes can be observed. Broadly speaking, two different concerns reflect through parliamentary debates: the pandemic itself and its consequences, and reaction to the pandemic and adoption of mitigation measures (section on the right). The comparison of COVID-19 cases and COVID-19-related debates shows that Italian and UK MPs started discussing this topic at the time of first confirmed cases, while the debates in the Slovene and Polish parliament slightly preceded the first confirmed cases. At the time of the second wave, however, only the debates by the Italian and Polish MPs coincided with the second wave, while there was no clear increase in the Slovenian and UK parliament.

#### Links to materials.

- [Final output report](#)

## Outcomes & Feedback

The Helsinki Digital Humanities Hackathon offered a unique opportunity to raise-awareness about digital humanities and SSHOC-related activities among highly motivated individuals, and to offer them intensive training experience in an interdisciplinary environment. Computer and data scientists were coupled with researchers from SSH in order to answer complex real-life research questions and to discover new problem areas that can only be answered in an interdisciplinary team. The participants were expected to formulate research questions with respect to their particular dataset, develop and apply methods and tools to answer them and to present as well as disseminate their work throughout the event to the hackathon community as well as wider audience via social media.

**Participants' satisfaction.**<sup>36</sup> Overall, the respondents were very satisfied with the hackathon and found the introductory lectures excellent. Moreover, one respondent suggested that the future hackathon could include additional time to discuss group members' strengths and weaknesses as well as the individuals' learning objectives. Only a small share (10%) were less satisfied with the event and found the introductory lectures only satisfactory.

The answers show that the participants most cherished the selflessness of the entire hackathon community, the team spirit, respect for the diversity and the possibility to have lively exchanges and discussion about research ideas and possible approaches.

---

<sup>36</sup> The post-event survey was filled out by 19 respondents.

People, teamwork, the challenge. We're convening with the group after the event. (Respondent No.1)

The best thing was that I learned a ton of things. (Respondent No.6)

I felt everyone was dedicated and tried their best, and it made me feel inspired. (Respondent No.7)

DHH21 was an amazing opportunity to work in an interesting project. This was an excellent learning experience, in both skills and working with a diverse group. (Respondent No. 18)

**Feedback regarding the organisation.** The respondent felt that the organisation, including the overall structure of the hackathon, pre-event communication, organisers' and leaders' support during the hackathon, was mostly excellent. They liked the structure of the groups and the fact that there was more than one team leader (as in the CBAQuest and ParlaMint groups). They also thought that the interim presentations gave good structure to the event.

I think you did a really great job in selecting the people taking part. A diverse environment. (Respondent No. 12)

Of course having multiple leaders means additional organisational effort, time commitment from the extra people, which may not always be possible, but I think it was a huge plus. (Respondent No. 13)

As the main downside, multiple respondents reported the virtual nature of the event. However, many also stated that it would be good to keep it virtual or hybrid (some groups virtual, some groups f2f) in the future, since this expands the possibilities to participate for those that do not have the means for travel.

The online nature of the event. Although it brings opportunities to people who otherwise could not participate, I couldn't help but think it that live event could have been much better for building team spirit, networking and just socializing in general, not to mention how tiring it is to stare at your screen for hours on hours every day. (Respondent No. 8)

Another suggestion touched upon the online facilities. Many shared the opinion that it would be good to reduce the number of platforms in the future, since they experienced frequent confusion and technical problems due to switching between the platforms (Gather, Zoom, GitHub, GDrive, Slack).

**Future work.** Several respondents underlined that this was their best experience of the hackathon so far and that they hope the tradition will continue. Apart from the ideas outlined above, one respondent also suggested that it could be interesting for the future hackathons to include a short conference where the teams could present their work.



From personal experience, it is not easy, but it is a great way to engage people even more and can be an incentive for the teams to continue working after the end of the hackathon to publish the work done. (Respondent No.17)