SSHOC Workshop

Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing

Day 1: 14 February 2022

SSHOC
social sciences & humanities open cloud

HOUSEKEEPING NOTES

- **The presentations are being recorded.** All participants will receive a link to the recording shortly after the event
  - Please keep your camera and microphone off, if you do not wish to appear on the recording – Q&A and exercises sessions will not be recorded

- **Slides are available** **https://tinyurl.com/5abb5fyb**

- **Questions:** Write them in the chatbox, ask them during the Q&A sessions or use our Questions Padlet https://padlet.com/dcmagd/rne46mrue7nfup14

- **Post-event feedback:** https://forms.gle/fPwXscqBcUHHyP7J9

SSHOC
social sciences & humanities open cloud

SPEAKERS

Hina Zahid

Cristina Magder

Anca Vlad

Maureen Haaker

SSHOC
social sciences & humanities open cloud
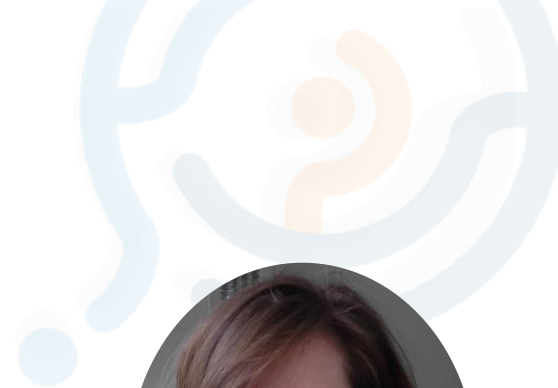
ABOUT US

- <u>UK Data Archive</u> is the lead partner of the <u>UK Data Service</u>

- funded by the <u>Economic and Social Research Council</u>, part of <u>UK Research and Innovation</u>

- centre of excellence in acquiring, curating and providing access to the largest collection of social science and population data in the United Kingdom for over 50 years

- provide a Trusted Digital Repository

- sharing best practice, we help to lead international best practice in <u>data management</u>

- based at the <u>University of Essex</u>

Project:

SSHOC
social sciences & humanities open cloud

Horizon 2020
European Union Funding
for Research & Innovation

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

Partners: 47
(20 beneficiaries + 27 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Project budget:
€ 14,455,594.08

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

Project website:
www.SSHOpenCloud.eu

Objectives:

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

SSHOC offerings

On- and offline trainings and training materials and an international cross-disciplinary trainer network.

Availability of an EU-wide, easy-to-use SSH Open Marketplace, where tools and data are openly accessible.

SSHOC
social sciences & humanities open cloud

https://www.sshopencloud.eu/training

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

ICEBREAKER

Please visit
https://tinyurl.com/ymhmnn56

SSHOC
social sciences & humanities open cloud
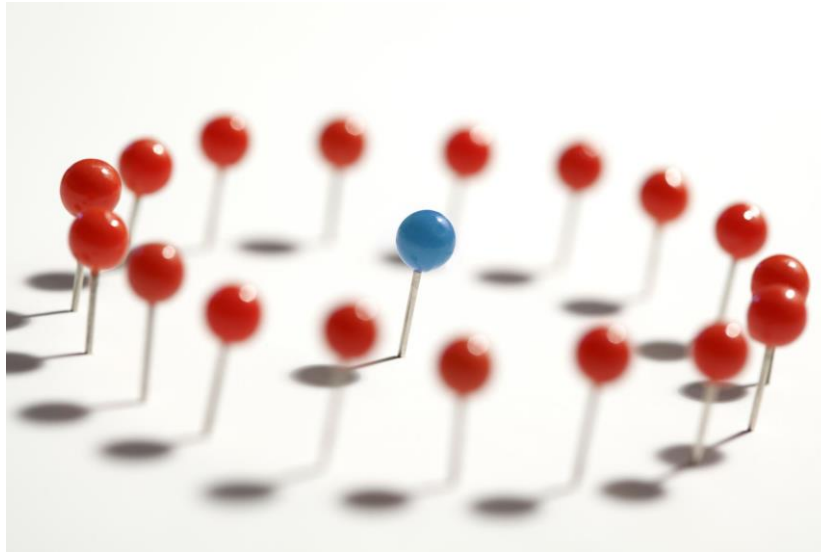
Workshop Learning Objectives

- Be able to recognise challenges in sharing data and identify techniques to overcome these

- Gain a better understanding of Data Management Planning requirements in Social Sciences

- Be able to design a comprehensive, accurate Data Management Plan which enables data sharing

SSHOC
social sciences & humanities open cloud

Day 1 Programme

- 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

Thank you for your attention!

Please share your thoughts about the event:
https://forms.gle/fPwXscqBcUHHyP7J9

www.sshopencloud.eu

@SSHOpenCloud

/in/sshopencloud

info@sshopencloud.eu

SSHOC Workshop

Data Management Planning in Social Sciences

Cristina Magder

UK Data Service

SSHOC

social sciences & humanities open cloud

Contents

Objectives of Data Management Planning

FAIR Principles

Data Management Plan Templates

Standards in social sciences

Checklist

Research Data Lifecycle
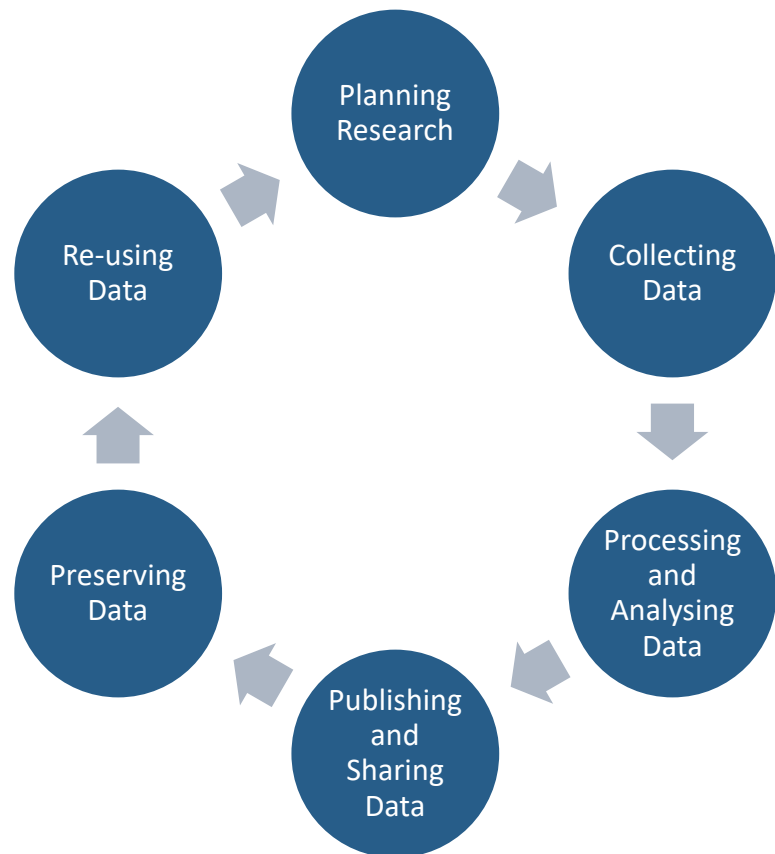
Data Management Planning

A data management plan (DMP) describes how the data will be:

- Collected

- Organised

- Managed

- Preserved

- Shared

SSHOC
social sciences & humanities open cloud

Why do we need data management planning?

- Plan ahead what to do with research data

- Keep track of research data (e.g. staff leaving)

- Identify support, resources, budget, services needed

- Plan storage, short & long-term

- Plan security, consider ethical and legal aspects

- Plan long-term sharing

- Show accountability (funder, institution, partners)

- To fulfil funders requirement

- Makes data Fairer: Findable, Accessible, Interoperable, Reusable

SSHOC
social sciences & humanities open cloud

FAIR Principles for publishing data

Findable

Accessible

Interoperable

Re-usable

Force 11 FAIR Principles

SSHOC
social sciences & humanities open cloud

Data Management Templates

Data management plans can differ from one funder/institution to another

Always consult your funder/host institution requirements

DMP: Horizon Europe (European Commission)

1. Data Summary

- Purpose of the data collection/generation

- Types, formats and size

- Existing data and sources (if any)

- Data utility

SSHOC
social sciences & humanities open cloud

DMP: Horizon Europe (European Commission)

2. FAIR Data

2.1 Making data findable

- discoverability (metadata), indexing and standards

- identifiability (persistent identifiers),

- versioning and naming conventions used

SSHOC
social sciences & humanities open cloud

DMP: Horizon Europe (European Commission)

2. FAIR Data

2.2 Making data openly accessible:

- how the data will be made available (including restrictions)

- methods/tools to access the data

- archiving of collection

DMP: Horizon Europe (European Commission)

2. FAIR Data

2.3 Making data interoperable:

- data and metadata vocabularies

- data and metadata standards

- data and metadata methodologies

DMP: Horizon Europe (European Commission)

2. FAIR Data

2.4 Increase data re-use (through clarifying licenses):

- data licencing

- dissemination embargo (if any)

- data quality assurance processes

- preservation

SSHOC
social sciences & humanities open cloud

DMP: Horizon Europe (European Commission)

3. Other Research Outputs

- e.g. software, workflows, protocols, models, antibodies, reagents, samples, etc.

- application of FAIR to other outputs

SSHOC
social sciences & humanities open cloud

DMP: Horizon Europe (European Commission)

4. Allocation of resources

- cost to making data FAIR

- responsibilities for data management

- potential value of long term preservation

SSHOC
social sciences & humanities open cloud

DMP: Horizon Europe (European Commission)

5. Data Security

- storage

- recovery

- transfer of sensitive data

- archving

SSHOC
social sciences & humanities open cloud

DMP: Horizon Europe (European Commission)

6. Ethics

- ethical and legal considerations for data sharing

- "will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?"

SSHOC
social sciences & humanities open cloud

DMP: Horizon 2020 (European Commission)

7. Other issues

- national/funder/sectorial/departmental procedures for data management

DMP: ESRC

Following aspects need to be considered:
Assessment of existing data
Information on new data
Quality assurance of data
Backup and security of data
Management and curation
Difficulties in data sharing and measures to overcome these
Consent, anonymization and re-use strategies
Copyright/Intellectual property rights
Responsibilities
Preparation of data for sharing and archiving

SSHOC
social sciences & humanities open cloud

DMP Sections at a Glance

Description of data (new & existing)

Management and Curation

Storage and backup

Legal and ethical aspects

Data sharing

Responsibility and resources

SSHOC
social sciences & humanities open cloud

Metadata Standards

- Dublin Core

- ISO 19115 for geographic information

- Data Documentation Initiative (DDI)

- Statistical Data and Metadata eXchange (SDMX)

- Metadata Encoding and Transmission Standard (METS)

- General International Standard Archival Description (ISAD(G))

- DataCite metadata schema for the publication and citation of digital datasets with a persistent identifier

SSHOC
social sciences & humanities open cloud

Data Documentation Initiative

- a rich and detailed metadata standard originally designed for describing social, behavioural and economic sciences data

- used by most social science data archives in the world.

- DDI catalogue records contain mandatory and optional metadata elements relating to study description, data file description and variable description

- data file description indicates data format, file type, file structure, missing data, weighting variables and software used

- variable descriptions indicate the variable labels and codes.

SSHOC
social sciences & humanities open cloud

Controlled Vocabularies: HASSET & ELSST

Humanities and Social Science Electronic Thesaurus (HASSET) and multilingual version, the European Language Social Science Thesaurus (ELSST)

- cover the core social science disciplines: politics, sociology, economics, education, law, crime, demography, health, employment, information and communication technology and, increasingly, environmental science

- follow ISO 25964: Thesauri and interoperability with other vocabularies as far as possible

- are available in the Linked Open Data format, SKOS

SSHOC
social sciences & humanities open cloud

Persistent identifiers

Ensure outputs can be:

- reliably referenced
- linked

Open Researcher and Contributor ID (ORCID) is a persistent digital identifier that distinguishes you as an individual from every other researcher.

Digital Object Identifiers (DOIs) for data collections

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

A checklist, can help with writing a data management plan, as it helps you identify what to put in place for good data practices and which actions to take to optimise data sharing

Planning
Who is responsible for which part of data management?
Are new skills required for any activities?
Do you need extra resources to manage data, such as people, time or hardware?
Have you accounted for costs associated with depositing data for longer-term preservation and access?

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

A checklist, can help with writing a data management plan, as it helps you identify what to put in place for good data practices and which actions to take to optimise data sharing

Planning
Who is responsible for which part of data management?
Are new skills required for any activities?
Do you need extra resources to manage data, such as people, time or hardware?
Have you accounted for costs associated with depositing data for longer-term preservation and access?

Data Management Plan Checklist

Documenting data
Will others be able to understand your data and use them properly?
Are your structured data self-explanatory in terms of variable names, codes and abbreviations used?
Which descriptions and contextual documentation explain what your data mean, how they were collected and the methods used to create them?
How will you label and organise data, records and files?
Will you be consistent in how data are catalogued?

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

Formatting
Are you using standardised and consistent procedures to collect, process, transcribe, check, validate and verify data, such as standard protocols, templates or input forms?
Which data formats will you use? Do the formats and software you use enable sharing and the long-term sustainability of data, such as non-proprietary software and software based on open standards?
When converting data across formats, do you check that no data, annotation or internal metadata has been lost or changed?

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

Storing
Are your digital and non-digital data, and any copies, held in multiple safe and secure locations?
Do you need to securely store personal or sensitive data? If so, are they properly protected?
If data are collected with mobile devices, how will you transfer and store the data?
If data are held in multiple places, how will you keep track of versions?
Do you know which version of your data files is the master?
Who has access to which data during and after research? Is there a need for access restrictions? How will these be managed after you are dead?
How long will you store your data for and do you need to select which data to keep and which to destroy?

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

Confidentiality, ethics & consent
Do your data contain confidential or sensitive information? If so, have you discussed data sharing with the respondents from whom you collected the data?
Are you gaining written consent from respondents to share data beyond your research?
Do you need to anonymise data, for example, to remove identifying information or personal data, during research or in preparation for sharing?

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

Copyright
Have you established who owns the copyright in your data? Might there be joint copyright?
Have you considered which kind of licence is appropriate for sharing your data and what, if any, restrictions there might be on reuse?
If you purchase or reuse someone else's data sources, have you considered how that data might be shareable? For example have you thought about negotiating a new licence with the original supplier?
Can you preserve, for the long term, personal information, so that it can be used in the future?

SSHOC
social sciences & humanities open cloud

Data Management Plan Checklist

Sharing
Do you intend to make all of your data available for sharing? Or do you know how you will select which data to preserve and share?
How and where will you preserve your research data for the longer term?
How will you make your data accessible to future users?

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

www.sshopencloud.eu

@SSHOpenCloud

/in/sshopencloud

info@sshopencloud.eu

SSHOC
social sciences & humanities open cloud

SSHOC Workshop

Ethical and Legal Considerations

Cristina Magder

UK Data Service

SSHOC
social sciences & humanities open cloud

Contents

Key principles of ethical research

Duty of confidentiality

Data protection legislation

Consent in research

Copyright considerations

SSHOC
social sciences & humanities open cloud

Key principles for ethical research

- To maximise benefit for individuals and society & minimise risk and harm

- The rights and dignity of individuals and groups should be respected

- Voluntary and appropriately informed participation

- Research should be conducted with integrity and transparency

- Clearly defined lines of responsibility and accountability

- Independence of research should be maintained and where conflicts of interest cannot be avoided they should be made explicit

SSHOC
social sciences & humanities open cloud

Best practices for ethical sharing of research data

- Avoid social and personal harm

- Ethical obligations should be considered throughout the research lifecycle; from planning and research design stage, data collection stage to the future use including publications, archiving, sharing and linking of data

- Comply with relevant laws

- Be knowledgeable about relevant research organisations own standards and requirements

- Data centres facilitate ethical and legal re-use of research data, protection of participants and safeguarding of personal data

SSHOC
social sciences & humanities open cloud

Duty of confidentiality and data sharing

Exists in common law and may apply to research data

Disclosure of confidential information is lawful when:

- the individual to whom the information relates has consented

- disclosure is necessary to safeguard the individual, or others, or is in the public interest

- there is a legal duty to do so, for example a court order

Best practice is to avoid very specific promises in consent forms

SSHOC

social sciences & humanities open cloud

Data protection considerations

If personal information about people is collected or used in research data protection regulations apply

EU GDPR (2018) & country specific laws

- a researcher based in a country from the European Union (country that adheres to EU GDPR) collects personal data about people living in a country from the European Union

e.g. DPA (2018) & the UK GDPR applies when

- a researcher based in the UK collects personal data about people anywhere in the world

- a researcher outside the UK collects personal data on UK citizens

e.g. DPA (2018), EU GDPR (2018) & the UK GDPR applies when

- a researcher based in the UK collects personal data about people across Europe

SSHOC
social sciences & humanities open cloud

Consent in research

Consent is the process by which a researcher discloses appropriate information about the research so that a participant may make a voluntary, informed choice to accept or refuse to cooperate

Why to seek consent?

- Participants understand what they're signing up to making participation and research more effective

- Research conducted is ethical

- Compliance with data protection regulation

SSHOC
social sciences & humanities open cloud

Types of consent

Consent for research ethics: provide information regarding study purpose, risks, benefits, voluntary participation

Consent can also be used as a legal basis for the processing of personal data under GDPR

Information Sheet

An information sheet should cover the following topics:

- Purpose of the research

- What is involved in participating

- Benefits and risks of participating

- Procedures for withdrawal

- Usage of the data during research, dissemination, storage, publishing and archiving

- Details of the research: funding source, sponsoring institution, name of project, contact details for researchers, how to file a complaint

SSHOC
social sciences & humanities open cloud

Consent Forms

Use simple language and free from jargon

Allow the participant to clearly respond to points such as:

- The participant has read and understood information about the project

- The participant has been given the opportunity to ask questions

- The participant voluntarily agrees to participate in the project

- The participant understands that they can withdraw at any time without giving reasons and without penalty

- Future uses (e.g. publications, share and reuse)

- Signatures and dates of signing for the participant and the researcher

SSHOC
social sciences & humanities open cloud

UKDS Model Consent Form

The template can be used by researchers to gain informed consent to conduct research that collects data from people using questionnaires, observations, interviews, diaries, focus groups, video recordings, etc.

It pays particular attention to ensure that research data can be curated and made available for future use, as well as addressing all standard requirements of a consent form.

UKDS Model Consent Form

SSHOC
social sciences & humanities open cloud

Consent Forms: Taking Part

Informed Consent for [name of study]

Please tick the appropriate boxes

	Yes	No

1. Taking part in the study

I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction. ☐ ☐

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. ☐ ☐

I understand that taking part in the study involves [...]. ☐ ☐

> Describe in a few words how information is captured, using the same terms as you used in the information sheet, for example: an audio-recorded interview, a video-recorded focus group, a survey questionnaire completed by the enumerator, an experiment, etc.].
>
> For interviews, focus groups and observations, specify how the information is recorded (audio, video, written notes).
>
> For questionnaires, specify whether participant or enumerator completes the form.
>
> For audio or video recordings, indicate whether these will be transcribed as text, and whether the recording will be destroyed.

If there is a potential risk of participating in the study, then provide an additional statement: I understand that taking part in the study has [.....................................] as potential risk. ☐ ☐

SSHOC
social sciences & humanities open cloud

Consent Forms: Use of Data in Study

2. Use of the information in the study

I understand that information I provide will be used for [...]. ☐ ☐

> List the planned outputs, e.g. reports, publications, website, video channel etc., using the same terms as you used in the study information sheet.
> Consider whether knowledge sharing and benefits sharing needs to be considered, e.g. for indigenous knowledge.

I understand that personal information collected about me that can identify me, such as my name or where I live, will not be shared beyond the study team. ☐ ☐

> At times this should be restricted to the researcher only.

SSHOC
social sciences & humanities open cloud

Consent Forms: Future Use

3. Future use and reuse of the information by others

I give permission for the [specify the data] that I provide to be deposited in [name of data repository] ☐ ☐ so it can be used for future research and learning.

Specify in which form the data will be deposited, e.g. de-identified (anonymised) transcripts, audio recording, survey database, etc.; and if needed repeat the statement for each form of data you plan to deposit.

Specify whether deposited data will be de-identified (anonymised), and how. Make sure to describe this in detail in the information sheet.

Specify whether use or access restrictions will apply to the data in future, e.g. exclude commercial use, apply safeguarded access, etc.; and discuss these restrictions with the repository in advance.

SSHOC
social sciences & humanities open cloud

Best practices for personal data

- Do not collect personal data if it is not essential

- Indicate clearly in a consent form where the participant's consent is being asked for processing their personal data and where consent is being asked for taking part in the research

- Keep consent forms under constant review

- Different forms of consent for different materials, e.g. audio recordings vs transcripts

SSHOC
social sciences & humanities open cloud

Copyright considerations

- Copyright is an intellectual property right assigned automatically to the creator

- Data owner (researcher) has copyright of research data

- Compiled datasets contain original copyright – seek permission to archive when collecting

- Data archives publish data – hold no copyright

If information is in the public domain (e.g. online) it does not mean copyright does not apply!

SSHOC
social sciences & humanities open cloud

Best practices when using secondary data

Question to ask:

- Who the copyright holder of the datasets is?

- Are you allowed to use them and in what way?

- Are you allowed to archive and publish them in a data repository?

If not, you may need to seek for further permission to distribute material you do not own - copyright clearance

If permission is not granted, need to remove copyrighted variables/material before publishing or sharing

SSHOC
social sciences & humanities open cloud

Further Resources

- [DARIAH ELDAH Consent Form Wizard | CFW](#)

- [UKDS Model Consent Form](#)

- [Example Consent Forms](#)

- [Example Information Sheet](#)

- [Consent for data sharing](#)

- [HRA Example Consent Form](#)

SSHOC
social sciences & humanities open cloud

Acknowledgements

- Dr. Hina Zahid, Senior Research Data Officer, UK Data Service, UK Data Archive

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

www.sshopencloud.eu

@SSHOpenCloud

/in/sshopencloud

info@sshopencloud.eu

SSHOC
social sciences & humanities open cloud

SSHOC Workshop

Management and Curation of Data

Day 1

Anca Vlad

SSHOC

social sciences & humanities open cloud

Summary

1. Anonymisation and Pseudonymisation
2. Documenting your data
3. Quality assurance and security
4. Responsibilities and costing

Menti exercise

SSHOC
social sciences & humanities open cloud

Part 1: Anonymisation and Pseudonymisation

Anonymisation and pseudonymisation

Summary

- About anonymisation and pseudonymisation

- 3 tier approach

- Plan to anonymise

SSHOC

social sciences & humanities open cloud

Anonymisation and Pseudonymisation

Anonymised data is '...information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable." (Recital 26, GDPR)

- even the data owner cannot re-identify data subjects.

Pseudonymised data "…the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person." (Article 4, GDPR)

- identifiable data has been removed or redacted so that cannot be traced back to the real values. Re-identification of data can only be achieved with knowledge of the de-identification key or by combination.

SSHOC
social sciences & humanities open cloud

3 Tier Approach

1. Consent
2. Anonymisation/Pseudonymisation
3. Access

Using this approach correctly should enable most data to be shared/archived for reuse.

SSHOC
social sciences & humanities open cloud

Plan to anonymise

Q: How do I know to what extent to pseudonymise or anonymise?

A: In line with consent form.

Q: Which one should go for?

A: Consider data utility and information loss. Value for future, secondary use.

The more information we strip out in the process of anonymisation, reusability value decreases.

Q: What access level will apply to my data when published?

A: "As open as possible, as closed as necessary."

SSHOC
social sciences & humanities open cloud

Resources

- [UKDS recurring anonymisation webinar](#)

- [Anonymising Research Data](#) - ESRC National Centre for Research Methods, Working Paper 7/06

- [Guide to Social Science Preparation and Archiving](#) from the InterUniversity Consortium for Political and Social Research

- [Anonymisation and Social Research](#), Ruth Geraghty

- [Timescapes anonymisation guidelines](#), University of Leeds

- [Anonymisation: managing data protection risk](#) - ICO code of practice

- [The Anonymisation Decision-Making Framework](#) - Mark Elliot, Elaine Mackey Kieron O'Hara and Caroline Tudor

- [Jisc guidance on anonymous data](#)

SSHOC
social sciences & humanities open cloud

Part 2: Documenting your data

Documenting your data

- What is documentation and why is useful/needed?

- Documentation types:
    - Study level documentation
    - Data level documentation

- Examples: User guide, data list

- Transcription

- File formats

SSHOC
social sciences & humanities open cloud

Documenting your data

- Enables you to understand the data if/when you return to it.

- Sufficient information for future researchers to understand and use the data

✓ If using your data for the first time, what would a new user need to know to make sense of it?

- The UK Data Archive uses data documentation to:
    - Supplement a data collection with documents and research instruments
    - Ensure accurate processing and archiving
    - Create a catalogue record for a published data collection

SSHOC
social sciences & humanities open cloud

Document your data

1. Study level documentation

- Data collection methodology and processes: sampling, sample size, fieldwork protocol, experiment protocol, interviewer instructions

- Codebook, user guide (for quantitative data)

- Information sheet, consent form (blank versions)

- Questionnaires, show cards, topic guides

- Data list: overview of key information about each interview, a map of the data collection (for qualitative data)

- Links to reports and publications (preferably DOIs where possible)

SSHOC
social sciences & humanities open cloud

Document your data

2. Data level documentation

- All structured, tabular data should have adequate variable names, variable and value labels

- Transcripts: header with context information: data and place of interview, interviewer, interviewee details (in line with consent form) etc.

- Variable names might include:
  - Question number system matching questions in the questionnaire used e.g. Q1a, Q1b, Q2, Q3b
  - Numerical order system e.g. V1, V2, V3
  - Meaningful abbreviations or combinations of abbreviations referring to meaning of the variable e.g. 'oz%=percentage ozone', 'GOR=Government Office Region', 'moocc=mother occupation'
  - For interoperability across platforms, variable names should not be longer than 8 characters and without spaces

SSHOC
social sciences & humanities open cloud

Document your data

2. Data level documentation (continued)

Similar principles for variable labels:

- Be brief and concise, maximum 80 characters
- Include unit of measurement where appropriate
- Reference the question number of a survey or questionnaire

> e.g. variable 'q11hexw' with label 'Q11b: hours spent taking physical
> exercise in a typical week' – the label gives the unit of measurement and a
> reference to the questions number (Q11b)

- Coding or classification schemes used, with a bibliographic reference

> e.g. Standard Occupational Classification 2000; ISO 3166 alpha-2
> country codes

For value labels:

- Codes of, and reasons for, missing data
- Avoid blanks, system missing or '0' values e.g. '99= not recorded', '98= not provided (no answer)', '97=not applicable(skipped)', '96= not known', '95=error'

SSHOC
social sciences & humanities open cloud

In practice: user guide

A user guide should contain variety of documents that provide context for your data: interview schedule, methodology, study findings, consent procedures, transcription notes, codebook etc.

User guide for *Mort, M. (2006). Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003. [data collection]. UK Data Service. SN: 5407* http://doi.org/10.5255/UKDA-SN-5407-1

SSHOC
social sciences & humanities open cloud

In practice: data list

Study Number 6377

Integrated Floodplain Management, 2006-2008

Morris, J.

Floodplain farm survey

Interview ID	Farmer code	Age	Farm scheme	Farm type	Size of farm (hectare)	Number of holdings	Date of interview	Interviewer name	No of pages	Text file name	Audio file name
1	Be1	35-45	Beckingham	Beef	360	1	04.12.2006	Helena	28	6377int001	6377int001
2	Be2	45-55	Beckingham	Arable	364	1	05.12.2006	Helena	21	6377int002	6377int002
3	Be3	45-55	Beckingham	Arable	372	2	06.12.2006	Helena	22	6377int003	6377int003
4	Be4	45-55	Beckingham	Arable	194	3	06.12.2006	Helena	18	6377int004	6377int004
5	Be5	55-65	Beckingham	Arable	108	1	07.12.2007	Helena	21	6377int005	6377int005
6	Be6	45-55	Beckingham	Arable	1254	2	01.02.2008	Helena	19	6377int006	
7	Bu1	55-65	Bushley	Mixed	101	2	13.02.2007	Quentin	29	6377int007	6377int007
8	Bu2	>65	Bushley	Mixed	97	1	15.02.2007	Quentin	15	6377int008	6377int008
9	Bu3	>65	Bushley	Arable	194	4	13.02.2007	Quentin	21	6377int009	6377int009
10	Bu4	55-65	Bushley	Mixed	202	1	15.03.2007	Helena	19	6377int010	6377int010
11	Cu1	35-45	Cuddyarch	Dairy	64	1	08.05.2007	Helena	19	6377int011	6377int011
12	Cu2	55-65	Cuddyarch	Dairy	189	2	08.05.2007	Helena	18	6377int012	6377int012
13	Cu3	55-65	Cuddyarch	Mixed livestock	76	1	08.05.2007	Helena	13	6377int013	6377int013
14	Cu5	45-55	Cuddyarch	Mixed livestock	198	1	09.05.2007	Helena	24	6377int014	6377int014
15	Cu6	55-65	Cuddyarch	Dairy	89	1	09.05.2007	Helena	14	6377int015	6377int015
16	Cu7	>65	Cuddyarch	Mixed livestock	190	4	11.05.2007	Helena	20	6377int016	6377int016
17	Cu8	55-65	Cuddyarch	Mixed livestock	109	2	11.05.2007	Helena	22	6377int017	6377int017
18	Id1	55-65	Idle	Arable	158	3	07.02.2007	Quentin	17	6377int018	6377int018a
18	Id1	55-65	Idle	Arable	158	3	07.02.2007	Quentin	17	6377int018	6377int018b
19	Id1b	55-65	Idle	Arable	158	3		Quentin	22	6377int019	
20	Id2	45-55	Idle	Dairy	150	1	08.02.2007	Quentin	17	6377int020	6377int020

SSHOC
social sciences & humanities open cloud

Transcription template

Should:

- ✓ Possess a unique identifier

- ✓ Adopt a uniform layout throughout the research project

- ✓ Make use of speaker tags – turn-taking

- ✓ Carry line breaks

- ✓ Be page numbered

- ✓ Include a document header giving brief details of the interview: data, place, interviewer name, interviewee details, etc.

Other considerations (optional):

- Cover page

- Short summary in English (for transcripts in foreign languages)

- Compatibility with import featured of Computer Assisted Qualitative Data Analysis Software (CAQDAS)

SSHOC
social sciences & humanities open cloud

In practice: transcript format

Study Name: **Interview ID:**
Depositor: **Date of Interview:**

Information about interviewee:
(e.g. Age, Gender, Occupation, Marital Status, Geographic region, etc. as relevant /appropriate)

R= Respondent/Interviewee *(if more than one respondent, use R1, R2, etc.)*
I=Interviewer

R: I came here in late 1968.

I: You came here in late 1968? Many years already.

R: 31 years already. 31 years already.

I: (laugh) It is really a long time. Why did you choose to come to England at that time?

R: I met my husband and after we got married in Hong Kong, I applied to come to England.

I: You met your husband in Hong Kong?

R: Yes.

I: He was working here [in England] already?

R: After he worked here for a few years -- in the past, it was quite common for them to go back to Hong Kong to get a wife. Someone introduced us and we both fancied each other. At that time, it was alright to me to get married like that as I wanted to leave Hong Kong. It was like a gamble. It was really like a gamble.

I: You were very brave to think about going abroad as you were so young at that time.

Model Interview Transcript:
https://www.ukdataservice.ac.uk/media/622380/ukdamodeltranscript.pdf

SSHOC
social sciences & humanities open cloud

File formats

Choice of software format for digital data:

- Planned data analyses

- Software availability / cost

- Hardware used – e.g. audio capture

- Discipline – specific standards and customs

Digital data is software dependent, so endangered by obsolescence of software/hardware.

Best formats for long-term preservation:

- standard, interchangeable and open

- UK Data Service optimal file formats for various data types

- Digital Preservation Coalition guidance on preservation formats
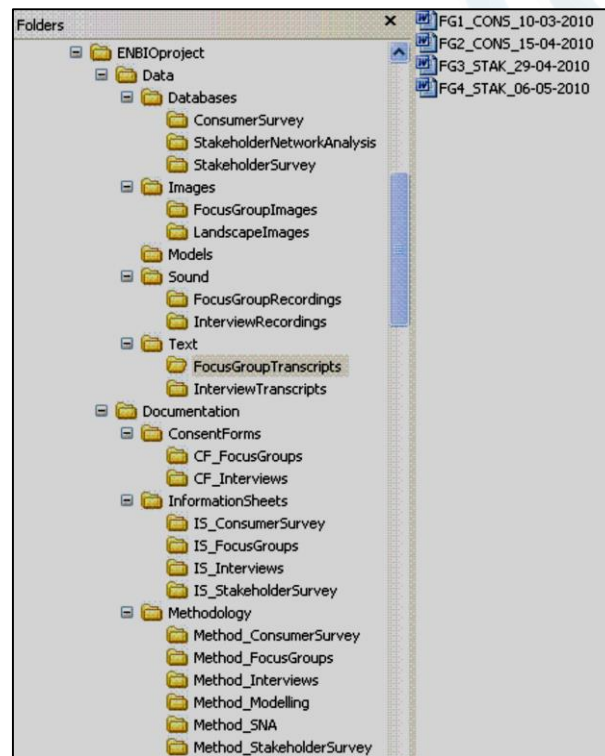
SSHOC
social sciences & humanities open cloud

Organising data

- Plan in advance how to best organise data (project specific)

- Use a logical structure and ensure collaborators understand

Examples

- Hierarchical structure of files, grouped in folders e.g. audio, transcripts and annotated transcripts

- Interview transcripts: individual well-named files

Resources

- [UKDS Data Management Guidance](#)

- [CESSDA Data Management Expert Guide](#)

- Managing and Sharing Research Data – a Guide to Good Practice (Sage Publications Ltd)

- [How and why you should manage your research data: a guide for researchers](#) – Jisc Guides

SSHOC

social sciences & humanities open cloud

Part 3: Quality assurance and security

Data security and storage

Protect data from unauthorised:

- Access

- Use

- Change

- Disclosure

- Destruction

Who knows who is watching, listening or attempting to access your data…

Data security strategy

- Control access to computers:
  - ✓ Use passwords and lock your machine when away from it
  - ✓ Run up-to-date anti-virus and firewall protection
  - ✓ Power surge protection
  - ✓ Restrict access to sensitive materials e.g. consent forms and patient records
  - ✓ Personal data need more protection – always keep them separate and secure
  - ✓ Utilise encryption
    - on all devices: desktops, laptops, memory sticks and mobile devices
    - at all locations: work, home and travel

- Control physical access to buildings, rooms and filing cabinets

- Properly dispose of data and equipment once the project is finished

SSHOC
social sciences & humanities open cloud

Encryption software

Encryption software can be easy to use and enables users to:
- ✓ Encrypt hard drives, partitions, files and folders
- ✓ Encrypt portable storage devices such as USB flash drives

- VeraCrypt

- Axcrypt

BitLocker

FileVault2

SSHOC
social sciences & humanities open cloud

Video tutorials

- VeraCrypt: https://www.youtube.com/watch?v=Ogm9QHQpFqU

- AxCrypt: https://www.youtube.com/watch?v=ACcRInsoYZg

- FileVault 2: https://www.youtube.com/watch?v=JIZ9EFMS0ic

- BitLocker: https://www.youtube.com/watch?v=y4Iosu-Yfsw

- Time Machine: https://www.youtube.com/watch?v=hIsQaVj7WtA

- MD5summer: https://www.youtube.com/watch?v=VcBfkB6N7-k

SSHOC
social sciences & humanities open cloud

Digital back-up strategy

- Consider:
- **What's backed-up?** - all, some or just the bits you change?
- **Where?** - original copy, external local and remote copies
- **What media?** - DVD, external hard drive, USB, Cloud?
- **How often?** - hourly, daily, weekly? Automate the process?
- **What method / software?** - duplicating, syncing or mirroring?
- **For how long is it kept?** - data retention policies that might apply?
- **Verify and recover** - never assume, regularly test and restore

- Backing-up need not be expensive
- 2Tb external drives are around
- 3 copies in 3 separate locations

- Consider non-digital storage too!

"We back up our data on sticky notes because sticky notes never crash."

SSHOC
social sciences & humanities open cloud

File sharing and collaborative environments

- Sharing data between researchers
- Too often sent as insecure email attachments
- Other options: Virtual Research Environments (MS SharePoint)
- Locally managed; ownCloud and ZendTo
- File transfer protocol (FTP)
- Physical media
- Cloud solutions
  - Google Drive, DropBox, Microsoft OneDrive and iCloud (insecure?)
  - More secure options? Mega.nz   SpiderOak   Tresorit

- Assess risks of using cloud storage (consult with your institution)

SSHOC
social sciences & humanities open cloud

Data disposal

Ensure proper disposal of equipment and media.

Even reformatting a hard drive is *not* sufficient.

If in doubt, physically destroy the drive.

BCWipe - uses 'military-grade procedures to surgically remove all
traces of any file

- can be applied to entire disk drives

AxCrypt - free open source file and folder shredding

- Integrates into Windows well, useful for single files

SSHOC
social sciences & humanities open cloud

Part 4: Responsibilities and costing

Who is responsible for data management?

- ✓ Ensure responsibilities for different aspects of data management are clearly defined early on and stated in your DMP.

- ✓ Often the Principal Investigator but not necessarily. Other members of the research team can be responsible for storage, others for anonymisation/pseudonymisation, others for archiving etc.

SSHOC
social sciences & humanities open cloud

Costing data management

- important to include costing in your DMP to ensure budgets are kept under control;

- plan and manage expenses;

- includes people's time, equipment, infrastructure and tools to manage, document, organise, store and provide access to data.

- on average: 2-3 weeks costed into an a typical two-year research grant application, to prepare and collate materials for deposit.

- UKDS costing tool

Costing data management

Examples of expenses:

- Data cleaning
- Formatting
- Anonymisation/Pseudonymisation
- Transcription & Digitisation
- Documentation
- Storage/Transfer/Access
- Backup
- Consent (retrospective – can be expensive)
- Copyright

SSHOC
social sciences & humanities open cloud

Mentimeter Exercise

Mentimeter Exercise

This exercise will cover concepts and information covered in:

✓ Management and Curation of Data

✓ Ethical and Legal Context

• 10 Minutes

• 13 Questions, 10 minutes

SSHOC
social sciences & humanities open cloud

Mentimeter Exercise

To participate, please go to www.menti.com

Use code: 3950 1107

You can open a separate tab in your browser or use another device (i.e phone, tablet etc.)

SSHOC
social sciences & humanities open cloud

Q&A

Resources and templates

Model consent form and survey consent statement:
https://ukdataservice.ac.uk/learning-hub/research-data-management/ethical-issues/consent-for-data-sharing/
Transcription template:
https://dam.ukdataservice.ac.uk/media/622380/ukdamodeltranscript.pdf
Transcription instructions:
https://ukdataservice.ac.uk//app/uploads/ukda-example-transcription-instructions.pdf
Transcription confidentiality agreement:
https://dam.ukdataservice.ac.uk/media/622354/ukda-transcriber-confidentiality-agreement.pdf
Data list template:
https://ukdataservice.ac.uk/uk_data_archive_data_listing_template

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

Please share your thoughts about the event:
https://forms.gle/pwPv3kbYJE3ASmPSA

www.sshopencloud.eu

@SSHOpenCloud

/in/sshopencloud

info@sshopencloud.eu

SSHOC
social sciences & humanities open cloud

SSHOC Workshop

Data Management Planning and Overcoming Challenges in Social Sciences Data Sharing

Day 1: 14 February 2022

Day 1

- Structure of DMPs
- Key principles and standards
- Ethical and legal considerations
- Anonymisation and documenting data
- Quality assurance and security
- Responsibilities and costing

SSHOC
social sciences & humanities open cloud

Day 2

- Existing data sources

- Secondary data use considerations

- Individual/group exercise + group discussion

- Common challenges in data sharing

- Showcases: Practices to enable data sharing

SSHOC
social sciences & humanities open cloud

Love Data Week

Love Data Week takes place in 2022 from 14 to 18 February

#LoveData22

SSHOC
social sciences & humanities open cloud

What have you enjoyed most about today?

Please visit
menti.com

use voting code
4131 7446

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

www.sshopencloud.eu

@SSHOpenCloud

/in/sshopencloud

info@sshopencloud.eu

SSHOC
social sciences & humanities open cloud