
UK Data Service Perspectives: Research Data Management

Dr Scott Summers

UK Data Service

University of Essex

WRDTP: Open Scholarship and Research Data
Management for Doctoral Researchers

5th December 2017

UK Data Service



Presentation overview

- UK Data Service
- Data sharing – background
- Consent
- Anonymisation
- Access
- Copyright
- Documentation and depositing shareable data
- Our tools and templates
- Your questions





UK Data Service

UK Data Service



What is the UK Data Service?

- A comprehensive resource funded by the ESRC
- A single point of access to a wide range of secondary social science data
- Support, training and guidance



What does the UK Data Service do?

- Put together a **collection of the most valuable data** and enhance that over time
- **Preserve data in the long term** for future research purposes
- Make the **data and documentation** available for reuse
- **Provide data management advice** to data creators
- **Provide support** for users of the service
- Provide information **about the use** to which data are put
- Easy access through a **website** - ukdataservice.ac.uk



Data Sharing



Data Sharing Background

- Data sharing is fast becoming a new paradigm in research across all disciplines, providing benefits to individual researchers, institutions, funders and more
- Many funders and academic publishers now specify requirements for data handling, including the formulation of a data management plan
- Good research data management habits are essential to creating data that are suitable for sharing



Why is it important to manage research data well?

- Data creation in research is often expensive
- Data is the cornerstone of research
- Good quality data leads to good quality research
- Data underpins published findings
- Enables compliance with ethical codes, data protection laws, journal requirements and funder policies
- To protect data from loss, destruction and potential exposure



ESRC Research Data Policy

- 9 Principles
- Expectation that ESRC data are made available for future reuse
- Deposit with the UK Data Service or with a data repository that complies with the FAIR principles
- FAIR Data Principles
- Make data – Findable, Accessible, Interoperable and Re-usable

<http://www.esrc.ac.uk/files/about-us/policies-and-standards/esrc-research-data-policy/>

<https://www.force11.org/group/fairgroup/fairprinciples>

UK Data Service



How to share data obtained from participants in a legal and ethical way

1. **Obtain informed consent**, for data sharing and long-term preservation/curation
 2. **Protect identities** through anonymisation and not collecting personal data for admin
 3. **Regulate access** where needed (all or part of data) e.g. by group, use or time period
- * Store personal or sensitive data **securely** and separately*
- (Utilise encryption and consider the storage location)

Consent



Consent is needed across the data life cycle

- Engagement in the research process
 - Decide who approves final versions of transcripts
- Dissemination in presentations, publications, the web
 - Decide who approves research outputs
- ***One more small step*** - Data sharing and archiving
 - Consider future uses of data



Informed consent for data sharing

- Consent needs to be **freely given, informed, unambiguous, specific** and by a **clear affirmative action** that signifies agreement to the processing of personal data
- The best way to achieve informed consent for data sharing is to **identify** and **explain** the **possible future uses of their data** and offer the participant the option to consent on a **granular level**
- For example, in a qualitative study, this may involve allowing the participant to consent to data sharing of the anonymised transcripts, the non-anonymised audio recordings and the photographs
- Consent should be documented

In practice: wording in consent forms / information sheets

We expect to use your contributed information in various outputs, including a report and content for a website. Extracts of interviews and some photographs may both be used. We will get your permission before using a quote from you or a photograph of you. After the project has ended, we intend to archive the interviews at Then the interview data can be disseminated for reuse by other researchers, for research and learning purposes.

The interviews will be archived at and disseminated so other researchers can reuse this information for research and learning purposes:

- I agree for the audio recording of my interview to be archived and disseminated for reuse
- I agree for the transcript of my interview to be archived and disseminated for reuse
- I agree for any photographs of me taken during interview to be archived and disseminated for reuse

In practice: wording in consent forms / information sheets

Use of the information I provide beyond this project		
I agree for the data I provide to be archived at the UK Data Archive. ²	<input type="checkbox"/>	<input type="checkbox"/>
I understand that other genuine researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	<input type="checkbox"/>	<input type="checkbox"/>
I understand that other genuine researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	<input type="checkbox"/>	<input type="checkbox"/>



In practice: wording in consent forms / information sheets

We ask you to consider the following points before agreeing to participate.

- Your contribution to the research will take the form of a focus group participant. This will be digitally video recorded and transcribed.
- Your name and any information which may directly or indirectly identify you will be altered to protect your anonymity.
- Any recordings of the discussions will be kept securely, and only authorised to other researchers on the condition they preserve your anonymity.
- The transcriptions (*excluding* names and other identifying details) will be retained by the researcher and analysed as part of the study. They will also be deposited with the UK Data Archive which has strict regulations about accessing data for research and protecting participant confidentiality.

Once a dataset is truly anonymised and individuals are no longer identifiable, European data protection law no longer applies – **Article 29 Working Party**

Question: In practice can we ever truly anonymise data?



Anonymisation



Anonymising quantitative data

- Direct and indirect identifiers
- Remove direct identifiers
e.g. names, address, institution, photo
- Reduce the precision/detail of a variable through aggregation
e.g. birth year instead of date of birth, occupational categories rather than jobs; and, area rather than village
- Generalise meaning of detailed text variable
e.g. occupational expertise
- Restrict upper lower ranges of a variable to hide outliers
e.g. income, age



Anonymising qualitative data

- Remove direct identifiers, or replace with pseudonyms
- Avoid blanking out; use pseudonyms or replacements
- Identify replacements with [brackets]
- Plan or apply editing at time of transcription
- Consistency throughout project
- Keep an anonymisation log of all replacements or removals made – keep separate from anonymised data files
- Avoid over-anonymising – removing information in text can distort data, make them unusable, unreliable or misleading; so balance anonymisation with the need to preserve context

www.ukdataservice.ac.uk/deposit-data/stories/gush



Anonymisation log

Example: Anonymisation log interview transcripts		
Interview / Page	Original	Changed to
Int1		
p1	Spain	European country
p1	E-print Ltd	Printing company
p2	20 th June	June
p2	Amy	Moira
Int2		
p1	Francis	my friend

P31. Joan → Mary

P97. Carol → {Mother}

P34. Colchester → {Town in S.E.England}

P65. Welshpool High School → @@##High School##@@

In practice: example anonymisation

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (SN 5407)

M. Mort, Lancaster University. Institute for Health Research

Date of Interview: 21/02/02

Interview with **Lucas Roberts**, DEFRA field officer

Date of birth: 2 May 1965

Comment [v1]: Replace: Ken

Comment [v2]: Delete

Gender: Male

Occupation: Frontline worker

Location: Plumpton, North Cumbria

Comment [v3]: Delete

Lucas was living at home with his parents, "but I'm hoping to move out soon" so we met at his parents' small neat house. We sat in a very comfortable sitting room with an open fire and **Lucas** made me coffee and offered shortbread. Although at first **Lucas** seemed a little nervous, quick to speech and very watchful he seemed to relax as we spoke and to forget about the tape.

Comment [v4]: Replace: Ken

Comment [v5]: Replace: Ken

Comment [v6]: Replace: Ken

I will just start by asking you to tell me a little bit about yourself and your background.

Access



Managing access to data

Open

- available for download/online access under open licence without any registration

Safeguarded

- available for download/online access to logged-in users who have registered and agreed to an End User Licence (*e.g. not identify any potentially identifiable individuals*)
- special agreements (depositor permission; approved researcher)
- embargo for fixed time period

Controlled

- available for remote or safe room access to authorised and authenticated users whose research proposal has been vetted and who have received training

In practice: data with access conditions

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

- Interviews (audio and transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users (**Safeguarded**)
- 3 interviews and 5 diaries were embargoed until 2015 (**Safeguarded – Embargoed**)
- Audio files archived and only available by permission from researchers (**Safeguarded – Special Agreement**)

discover.ukdataservice.ac.uk/catalogue/?sn=5407

doc.ukdataservice.ac.uk/doc/5407/mrdoc/pdf/q5407userguide.pdf



In practice: access conditions ReShare

Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance

Leonard, David 2014. Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance. [Data Collection]. Colchester, Essex: Economic and Social Research Council. <http://reshare.ukdataservice.ac.uk/id/eprint/851233>

The programme of research centres on how the various institutions responsible for the production of security and the management of conflict in Sub-Saharan African societies do, could and should evolve in response to the presence of violent conflict. These institutions include: the armed forces, police, courts (civil as well as criminal), 'traditional' institutions of local governance, elections (and other methods of choosing leadership), international organisations (such as the United Nations, African Union, UN High Commission for Refugees, the World Court,), and the community of international donor nations. The programme is built on the observation that all governance (especially in Africa) is multi-leveled and networked - from the village to the international organisation, and well beyond what is specified in formal government structures. Thus the focus will be not only on the ways in which key conflict-management institutions evolve themselves but also on the changing ways in which the networks in which they are embedded actually operate. This leading edge research challenges theories about the state, state formation, and sovereignty as well as the dynamics of violent conflict as presented in the dominant current work on Africa. The primary methods of research will be network analysis and qualitative interviewing.

Creators:	Creator Name	Email	Affiliation	ORCID
	Leonard, David	Unspecified	Institute of Development Studies	Unspecified
Research funders:	Economic and Social Research Council			
Grant reference:	RES-071-27-0048			
Subjects:	Law, crime and legal systems Politics			
Date deposited:	31 Jan 2014 19:00			
Last modified:	22 Oct 2014 19:27			

UK Data Service



In practice: access conditions ReShare

— Coverage and Methodology				
Collection period:	Date from:	Date to:		
	1 April 2009	31 March 2013		
Country:	Sierra Leone			
Data collection method:	Group interviews in 39 villages and towns in rural Sierra Leone			
Observation unit:	Groups			
Kind of data:	Alpha-numeric, Textual			
Type of data:	Qualitative and mixed methods data			
Resource language:	English			
— Access and Administration				
Data sourcing, processing and preparation:	Citizen interviews			
Copyright holders:	Name	Email	Affiliation	ORCID
	Leonard, David	Unspecified	Institute of Development Studies	Unspecified
	,	Unspecified	Unspecified	Unspecified
Contact:	Name	Email	Affiliation	ORCID
	Leonard, David	leonard@berkeley.edu	Institute of Development Studies	Unspecified
Notes on access:	The names of individual respondents have been removed from the data set.			
Publisher:	Economic and Social Research Council			
Last modified:	22 Oct 2014 19:27			



In practice: access conditions ReShare

AVAILABLE FILES

Data

– Security_%26_Networks.xlsx

Accessible to: Registered users only (safeguarded data)

File or bundle content: Data

File or bundle description: Security & Networks

File format: application/octet-stream

License: UK Data Service End User Licence

File size: 10Kb

+ Sierra_Leone_Security_%26_Networks__Coded.xls

+ Copy_of_Somalia_SC_peace_initiatives.xlsx

Documentation

– Sierra_Leone_methods.doc

Accessible to: Anyone (open data)

File or bundle content: Documentation

File or bundle description: Sierra Leone methods

File format: application/msword

License: UK Data Service End User Licence

File size: 64Kb



Copyright



IP/copyright

- Broadly speaking a researcher has copyright in the research data that they collect
- Compiled datasets contain original copyright – seek permission to archive when collecting
- Data archives publish data – hold no copyright *note*
 - Copyright holders give permission to preserve and disseminate data
- Data can be archived in various places, e.g. ReShare and local repository



Depositing Shareable and Reusable Data

UK Data Service



Include documentation

- Sufficient information for future researchers to understand and use the data
- Data collection methodology and processes: sampling, sampling size, fieldwork protocol and interviewer instructions
- Information sheets and consent forms
- Questionnaire, showcards and question lists
- Transcripts: header with context information: date & place interview, interviewee name, etc.
- Data list: overview of key information about each interview, as 'at-a-glance' summary of the data collection
- Links to reports and publications
- A README file

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
- carry a document header giving brief details of the interview: date, place, interviewer name, interviewee details, etc

In practice: transcript format

Study Name:
Depositor:
Interviewer:

Interview number:
Interview ID: Firstname Lastname
Date of interview:

Information about interviewee

Date of birth:
Gender:
Geographic region:

Marital status:
Occupation:

Y=Interviewee

I=Interviewer

Y: I came here in late 1968.

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

Y: 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?

Y: 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?

Y: Yes.

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

In practice: data list

- Data listing provides an at-a-glance summary of interview sets

Study Number 5407

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001

Mort, M.

The panel respondents for the study were divided into six population groups. The data list for the diary and interviews has been colour-coded accordingly for clarity, using the depositor's original colours:

Group 1: Farmers	Group 2: Rural Business	Group 3: Agricultural related occupations	Group 4: Frontline Workers	Group 5: Community	Group 6: Animal / Human Health Professionals
------------------	-------------------------	---	----------------------------	--------------------	--

1. Interviews

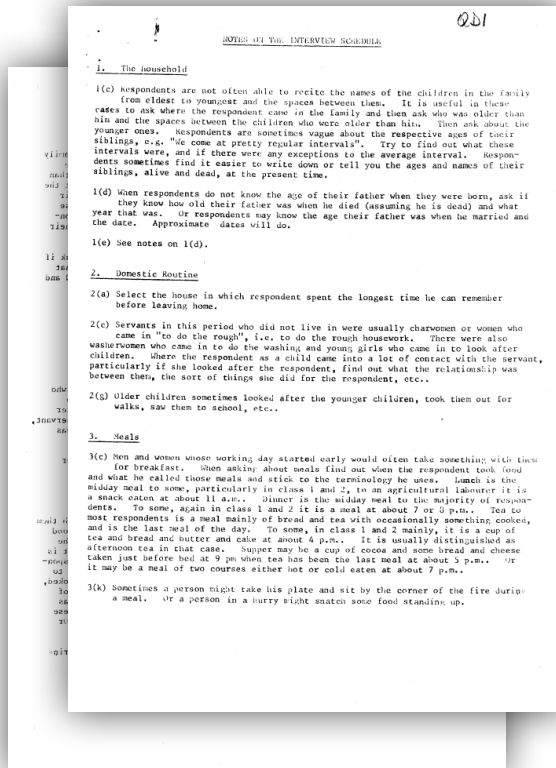
Respondent ID	Population Group	Date of Birth	Gender	Occupation	Interview summary	Place of Interview
PM02	Group 6: Animal / Human Health Professionals	1975	M	Veterinary Surgeon	Family and background, career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria, resp home
PM03	Group 6: Animal / Human Health Professionals	1966	F	Veterinary Surgeon	Family and background, career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria
PM07	Group 6: Animal / Human Health Professionals	1964	F	Veterinary practice manager	Family and background, career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria, resp home

UK Data Service

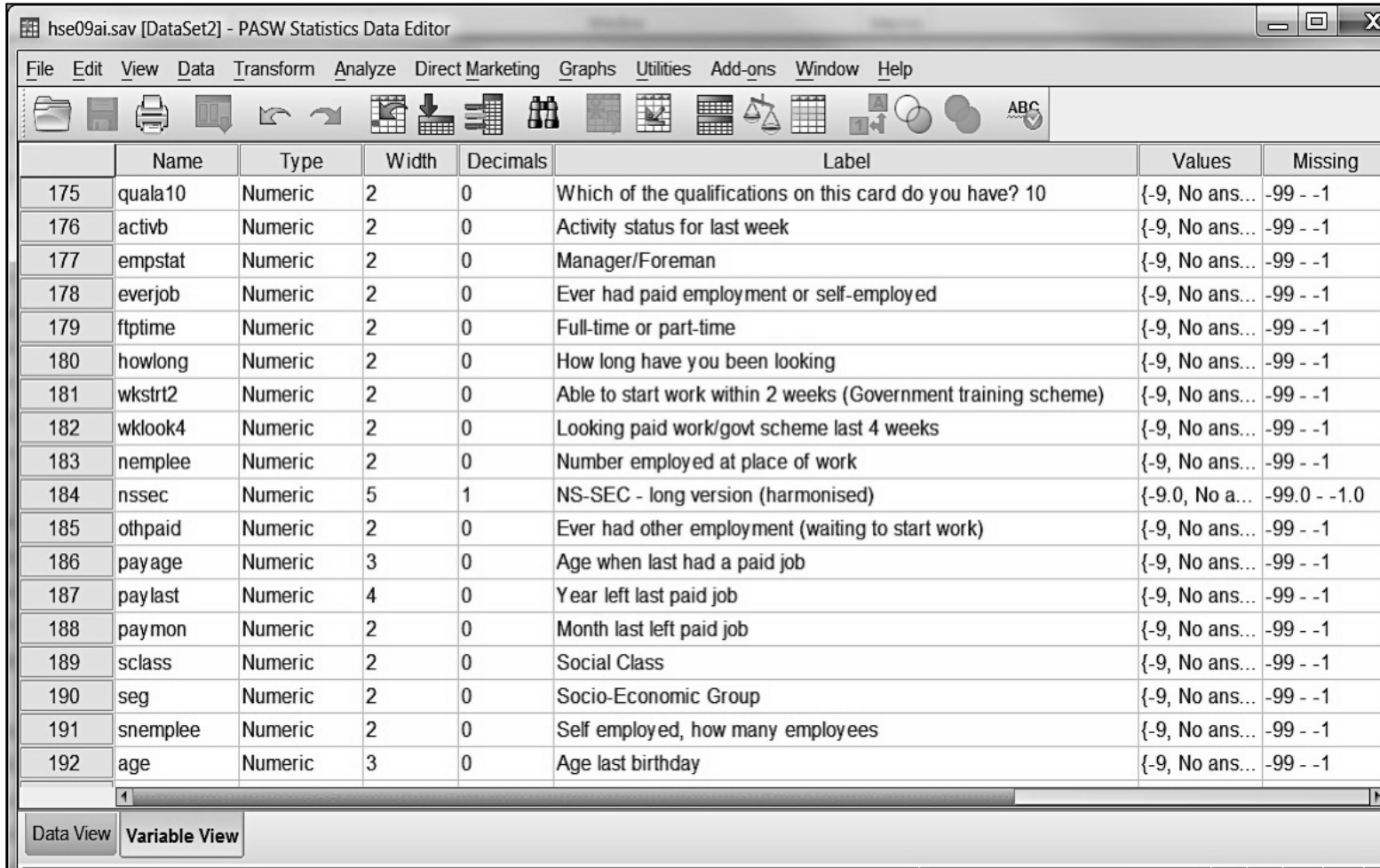


In practice: user guide and documentation

- A user guide could contain a variety of documents that provide context: interview schedule, transcription notes, even photos



Embedded data-level metadata in SPSS file



hse09ai.sav [DataSet2] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing
175	quala10	Numeric	2	0	Which of the qualifications on this card do you have? 10	{-9, No ans...	-99 - -1
176	activb	Numeric	2	0	Activity status for last week	{-9, No ans...	-99 - -1
177	empstat	Numeric	2	0	Manager/Foreman	{-9, No ans...	-99 - -1
178	everjob	Numeric	2	0	Ever had paid employment or self-employed	{-9, No ans...	-99 - -1
179	ftptime	Numeric	2	0	Full-time or part-time	{-9, No ans...	-99 - -1
180	howlong	Numeric	2	0	How long have you been looking	{-9, No ans...	-99 - -1
181	wkstrt2	Numeric	2	0	Able to start work within 2 weeks (Government training scheme)	{-9, No ans...	-99 - -1
182	wklook4	Numeric	2	0	Looking paid work/govt scheme last 4 weeks	{-9, No ans...	-99 - -1
183	nemplee	Numeric	2	0	Number employed at place of work	{-9, No ans...	-99 - -1
184	nssec	Numeric	5	1	NS-SEC - long version (harmonised)	{-9.0, No a...	-99.0 - -1.0
185	othpaid	Numeric	2	0	Ever had other employment (waiting to start work)	{-9, No ans...	-99 - -1
186	payage	Numeric	3	0	Age when last had a paid job	{-9, No ans...	-99 - -1
187	paylast	Numeric	4	0	Year left last paid job	{-9, No ans...	-99 - -1
188	paymon	Numeric	2	0	Month last left paid job	{-9, No ans...	-99 - -1
189	sclass	Numeric	2	0	Social Class	{-9, No ans...	-99 - -1
190	seg	Numeric	2	0	Socio-Economic Group	{-9, No ans...	-99 - -1
191	snemplee	Numeric	2	0	Self employed, how many employees	{-9, No ans...	-99 - -1
192	age	Numeric	3	0	Age last birthday	{-9, No ans...	-99 - -1

Data View Variable View



Preparing data files

- Consistent/meaningful file names, no spaces/special characters
- Self-explanatory variable names, codes, abbreviations
- Label all variables
- Remove temporary/dummy/admin variables
- Anonymise sufficiently (also look at file properties)
- Check textual variables are OK to disseminate
- Missing values – label and consistency
- Check [best file format for longevity](#)
- Check data for completeness
- Bundle multiple files into zip bundles



Depositing your research data in ReShare



reshare.ukdataservice.ac.uk
reshare@ukdataservice.ac.uk

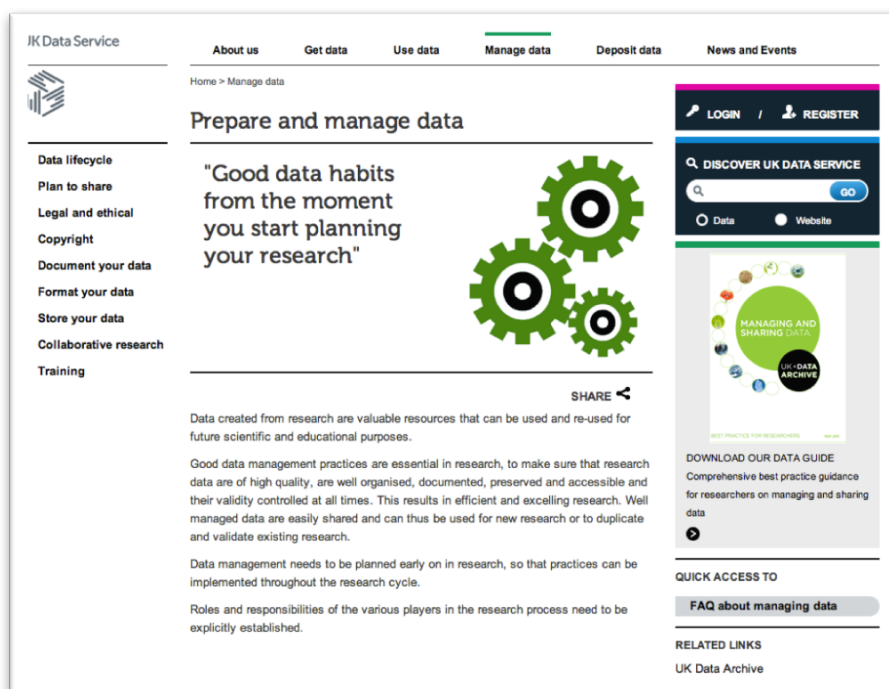


UK Data Service



Our data management guidance

- Online best practice guidance: ukdataservice.ac.uk/manage-data.aspx
- [Managing and Sharing Research Data – a Guide to Good Practice:\(Sage Publications Ltd\)](#)
- Helpdesk for queries: ukdataservice.ac.uk/help/get-in-touch.aspx
- Training: www.ukdataservice.ac.uk/news-and-events/events
- Twitter: [@UKDSRDM](https://twitter.com/UKDSRDM)



The screenshot shows the UK Data Service website interface. The main navigation bar includes 'About us', 'Get data', 'Use data', 'Manage data' (highlighted), 'Deposit data', and 'News and Events'. The page title is 'Prepare and manage data'. A central quote reads: "Good data habits from the moment you start planning your research". To the right of the quote are three green gears. Below the quote is a 'SHARE' button. The main content area contains three paragraphs of text about the value of research data, the importance of good data management practices, and the need for planning data management from the start of a research project. On the right side of the page, there is a sidebar with a search bar, a 'DISCOVER UK DATA SERVICE' button, and a 'GO' button. Below the search bar are radio buttons for 'Data' and 'Website'. There is also a 'LOGIN / REGISTER' link and a 'DOWNLOAD OUR DATA GUIDE' section with a 'QUICK ACCESS TO' button and a 'FAQ about managing data' link. At the bottom, there are 'RELATED LINKS' including 'UK Data Archive'.



UK Data Service



Tools & templates (1)

- Model consent form: <http://www.data-archive.ac.uk/media/112638/ukdamodelconsent.pdf>
- Survey consent statement: <http://data-archive.ac.uk/media/147338/ukdasurveyconsent.doc>
- Transcription template: <http://data-archive.ac.uk/media/136055/ukdamodeltranscript.pdf>
- Transcription instructions: <http://data-archive.ac.uk/media/285633/ukda-example-transcription-instructions.pdf>
- Transcription confidentiality agreement: <http://data-archive.ac.uk/media/285636/ukda-transcriber-confidentiality-agreement.pdf>
- Data list template: <http://data-archive.ac.uk/media/2989/UK%20Data%20Archive%20Example%20Data%20List.pdf>



Tools & templates (2)

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=hlsQaVj7WtA>
- MD5summer: <https://www.youtube.com/watch?v=VcBfkB6N7-k>



Questions?

