
Disseminating digital data

Anca Vlad & Veerle Van den Eynden

UK Data Service

University of Essex

Essex

3 Oct 2018

UK Data Service



Content summary

- How and where can audio-visual data be published?
- What activities are needed to publish data
 - Documenting data
 - Preparing to publish digital data
 - Licensing and access control
- Short demo: Creating a data collection in ReShare
- How are people re-using data?



UK Data Service

- Curate, preserve, provide access to social science data for reuse
- Funded by UKRI Economic and Social Research Council
- Data management advice for data creators
- Support for users of the service
- Information about the use to which data are put

ukdataservice.ac.uk



Some statistics about our UK Service

- **7,300** datasets in the collection
- **1050** qualitative and mixed methods collections
- **400** new datasets added each year
- **219** case studies of data reuse
- **25,000** registered users
- **60,000** downloads worldwide per year
- **4000+** user support queries per year



Discussion – challenges

- What are challenges and opportunities for disseminating and publishing your data?

Common data publishing routes

- Do it yourself: project website, Youtube, Vimeo
- Journal, as supplementary materials
- General repository, e.g. Figshare, Dryad
- Institutional repository
- Domain /subject repository, e.g. UK Data Service, ELAR



Over 1500 repositories



Pros and cons to each method

FAIR principles for repositories

Findable

Accessible

Interoperable

Re-usable

Persistent identification of collections

UK Data Service



Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Basic data publishing requisites

- ✓ Usable format
- ✓ Long-term preserved/backed-up
- ✓ Self-explanatory for users
- ✓ Non disclosive where promised
- ✓ Rights are in place to redistribute

Meeting the FAIR Principles



Considerations audiovisual data

- File sizes
- Formats, compression,
- Raw footage or processed
- Download or stream
- Annotated, transcribed or not
- Open or restricted access



ACTIVITIES NEEDED TO PUBLISH DATA

Documenting data

Quality documentation is key to enabling the secondary user to make informed use of the data. Depending on data type, the following pieces of documentation are normally produced/archived.

Qualitative data:

- Interview schedule (for structured and semi-structured interviews)
- Data list
- Annotations file
- Project Overview
- Methods file
- ReadMe file
- Consent form(s)
- Information Sheet(s)

Quantitative data:

- Variable & value labels (data file)
- Codebook (including how derived variables are created, if any);
- Questionnaire (blank)
- Scripts for secondary data;
- Project Overview
- Methods file
- ReadMe file
- Consent form
- Information sheet

Documentation for audio-visual data

- Annotations can be produced to document video data, however make sure to check with the receiving archive for acceptable formats (ELA only accepts annotations produced by ELAN and by Praat)
- Data Lists can be useful to have an overview of multiple similar files(videos) and their content.
- Methods document – provide information on components of research design and details about data collection.
- ReadMe files can be useful to provide future users with a summary of all files archived as well as the structure of the collection (folders and subfolders).

Methods document

Methods document – provide information on components of research design, timeline and details about data collection.

Plan of work: Months	1-3	4-9	10-12	13-16	17-20	21-24
Review and design	■					
Phase 1: film production		■				
Phase 2: Online interaction			■			
Phase 3: Evaluation			■	■		
Analysis, Writing, Dissemination			■	■	■	■

Source: **CHAsH_MethodsDescription.pdf** in *Butcher, Melissa (2015). Creating Hackney as Home: Films, transcripts and metadata, 2013-2015. [Data Collection]. Colchester, Essex: UK Data Archive. 10.5255/UKDA-SN-851888*

ReadMe file* example (qualitative collection)

*summary document containing a short description of content and any relationships between files

The following files have been archived:

File name	Description
<u>Kikole</u> PV Summary Report FINAL.docx	This is a summary report into the respondents and research for the ethnographic, participatory video data.
LIBRO LOMERIO BOOK FINAL DIGITAL FINAL (1).pdf	A .pdf book about local conceptions of forest conservation, markets and environmental justice. (In Spanish).
Guidelines for ethnographic studies.doc	A framework for the research's ethnography analysis.
Report UEA_InsightShare.pdf	A report detailing the training and professional support in Participatory Video (PV) for researchers associated with the ESRC-funded project Conservation, Markets and Justice Research Programme.
Participatory Video in Conservation, Markets and Justice - A <u>Photostory</u> low res.pdf	A photo story of the participatory video project.
Summary of <u>Ruhatwe</u> Participatory Video FINAL.docx	A summary report of the participatory video project in the village of <u>Ruhatwe</u> .
ConsentForm.docx	Consent form for participants in research methods.
EthicsApprovalForm.docx	Ethics approval form for the research project from the University of East Anglia.

Source: Martin, Adrian and Kebede, Bereket and Sikor, Thomas and Gross-Camp, Nicole and Rodriguez, Iokine (2017). Conservation, markets and justice - Part 2: Ethnographic participatory video data. [Data Collection]. Colchester, Essex: UK Data Archive. 10.5255/UKDA-SN-852476

UK Data Service



ReadMe file example (mixed methods collection)

The following files [have been archived](#):

File name	Description
FarmSurvey2013.sav	SPSS file containing a dataset generated by a questionnaire survey of small-scale sugarcane growers in Nkomazi, South Africa, in 2013. The dataset contains 104 cases (questionnaires) and 106 variables.
VariableDescriptions.docx	This file contains descriptions, including explanations of how some variables have been calculated from others, for the variables in file Farm survey2013.sav
SampleDesign.docx	This file contains an explanation of the design of the sample used in the questionnaire survey that generated the data in file Farm survey2013.sav
QuestionnairePrint.xps	This file shows the questionnaire form constructed using SPSS software and filled in by typing responses into a laptop. These data generated the dataset contained in file Farm survey2013.sav
InterviewTranscripts.docx	This file contains transcripts of 15 semi-structured interviews undertaken with small-scale sugarcane growers in Nkomazi in 2013. The interviews examine interviewees' life histories and explore the impacts of sugarcane cultivation on their livelihoods. Those interviewed include some of those in the sample used for the questionnaire survey.
ConsentForm.doc	Consent form for participants in questionnaire survey
ParticipantInformationSheet.doc	Project information sheet given to participants in questionnaire survey.
WorkingPaper1.pdf	Working paper that presents an analysis of the questionnaire survey dataset in file Farm survey2013.sav
WorkingPaper2.pdf	Working paper that combines the findings of Working Paper 1 with additional material from secondary sources and from the interview transcripts in file Interview Transcripts.docx
WorkingPaper3.pdf	Working Paper that summarises the land reform context on the Nkomazi area.

Source: Woodhouse, Philip (2015). *Survey of small-scale sugarcane growers in South Africa 2011-2015. [Data Collection]. Colchester, Essex: UK Data Archive. 10.5255/UKDA-SN-852082*

UK Data Service



Data list

Scope: for qualitative data collections, such as interview, video, audio or image collections.

Content:

- provides information for users that enables them to easily identify and locate relevant transcripts or items within a data collection
- provides key biographical characteristics and features of interviewees (age, gender, occupation, organisation, location, place of interview, date of interview, transcript file name, recording file name and any other details relevant to the project)
- should indicate where parts of the data are missing, such as partial or missing transcripts
- each item in the list should have a unique identifier (identifiers used should be consistent so links can be made with other related files)

Our data list template can be found [here](#).

Data list example (interview data)

Interview ID	Gender	Description	Date of Interview	No of Pages	Text File Name
x001	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/09/2005	36	6124int001
x002	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	41	6124int002
x003	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	39	6124int003
x004	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	36	6124int004
x005	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	34	6124int005
x006	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	14/09/2005	50	6124int006
x007	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	31	6124int007
x008	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	35	6124int008
x009	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	33	6124int009
x010	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	23	6124int010
x011	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	36	6124int011
x012	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	41	6124int012
x013	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	21	6124int013
x014	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	30/09/2005	20	6124int014
x015	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	19	6124int015
x016	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	27	6124int016
x017	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/10/2005	27	6124int017
x018	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	17/10/2005	11	6124int018
x019	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	19/10/2005	33	6124int019
x020	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/11/2005	21	6124int020
z001	Male	Interview with Hospital Doctor in Northern Tourist City	19/07/2005	50	6124int021
z002	Male	Interview with Hospital Doctor in Northern Tourist City	10/08/2005	45	6124int022
z003	Male	Interview with Hospital Doctor in Northern Tourist City	17/08/2005	56	6124int023
z004	Male	Interview with Hospital Doctor in Northern Tourist City	07/11/2005	21	6124int024
z005	Female	Interview with Hospital Doctor in Northern Tourist City	14/11/2005	32	6124int025
z006	Male	Interview with Hospital Doctor in Northern Tourist City	15/11/2005	23	6124int026
z007	Male	Interview with Hospital Doctor in Northern Tourist City	16/11/2005	23	6124int027
z008	Female	Interview with Hospital Doctor in Northern Tourist City	17/11/2005	18	6124int028
z009	Male	Interview with Hospital Doctor in Northern Tourist City	18/11/2005	20	6124int029

Source: Nettleton, S. (2009). *Being a Doctor: a Sociological Analysis, 2005-2006*. [data collection]. UK Data Service. SN: 6124, <http://doi.org/10.5255/UKDA-SN-6124-1>



File naming conventions and best practice

- file name/base name(not including file format extension) = principal identifier of file
- use logical naming i.e. easy to identify and retrieve the file
- naming provides organisation, context & consistency
- name elements: version number, date, content description, creator name

Best practice:

- name independent of location
- relevant to content
- no special characters, dots or spaces
- for separation use underscores _
- versioning via filename: ascending, decimal version numbers
- avoid very long file names

Licensing and access control

Archiving data in a repository does not mean this would be openly accessible. To protect data that cannot be anonymised without reducing reuse potential, access restrictions are used.

Open access – can be accessed by any user, no registration needed

Safeguarded- accessible only to registered users

Permission only – accessible with permission from depositor

Embargo – closed temporarily (up to 12 months to allow publication)

Note: Access restrictions can only be applied to data files. Documentation files and metadata is always openly available.

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 and who have received training

Access categories, choosing the right one

Choosing the appropriate access category to fit a particular data file or collection is always a case by case process, however the following are considered:

- Identifiable information in the data – quality and procedures of anonymization;
- Consent statements
- Linkage possible? (to other published data/information which might contribute to identifying participants)

The appropriate access category is chosen during the review process, in collaboration with the depositor; access is set at file level, therefore one collection can contain multiple data files/folders under different access restrictions. For an example, please see:

Simpson, Edward and Tina, Otten and Alice, Tilche and Tommaso, Sbriccoli and Patricia , Jeffery (2018). **Rural change and anthropological knowledge in post-colonial India: a comparative 'restudy' of F.G. Bailey, Adrian C. Mayer and David Pocock, 1950-2012**. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-852771](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-852771)



Licensing

Depending on the appropriate access level, the corresponding licence will apply.

For open access data, different versions of Creative Commons International:

[Creative Commons: Attribution 4.0 International](#)

[Creative Commons: Attribution-ShareAlike 4.0 International](#)

[Creative Commons: Attribution-NonCommercial-ShareAlike 4.0 International](#)

For restricted access data archived with UKDA, our [End User Licence](#) applies.

For data disseminated via alternative routes, other licences may apply; i.e. Youtube allows users to mark their videos with a Creative Commons [CC BY](#) licence; with Vimeo, you own content you submit to the Vimeo Service.

EXAMPLE DISSEMINATION OPTIONS



Dissemination via an archive: ReShare

[Creating Hackney as Home: films, transcripts and metadata, 2013-2015](#)

Collections of :

- films created as the central output of the project
- transcripts of flipcam reflections
- interviews made by the peer researchers
- transcripts of team meetings
- interviews with stakeholders
- transcripts of discussions following public screenings of the films

Short films:

- mp4 format
- <1GB each

Dissemination via an archive: ReShare

[Rural change and anthropological knowledge in post-colonial India: a comparative 'restudy' of F.G. Bailey, Adrian C. Mayer and David Pocock, 1950-2012](#)

Dataset results from an anthropological project that 'restudied' three villages in India (Sundarana in Gujarat, Jamgod in Madhya Pradesh and Bisipada in Orissa) that had originally been studied by anthropologists in the 1950s. The dataset consists of:

- maps and census data generated by the fieldwork in 2012
- digitised fieldnotes (some 4,500 pages) of anthropologist F.G. Bailey's fieldwork in Orissa in the 1950s, for which the high definition images are hosted by the SOAS Digital Collection
- interviews with Bailey and Mayer about their research (video and audio)
- photographic collection of Orissa in the 1950s
- images of Bisipada village, Odisha of 1950 and 2012
- ethnographic data collected during the 2012 resurvey of Jamgod, Madhya Pradesh
- ethnographic data collected during the 2012 resurvey of Sundarana, Gujarat

Formats: mp4, wav. Jpg

Size: bundles of 10-13 GB each, consisting of short video recordings (raw footage)

UK Data Service



Dissemination via an archive: Endangered Languages Archive (SOAS London)

- The Endangered Languages Archive (ELAR) is a digital repository preserving and publishing multimedia collections of endangered languages
- Collections contain audio and video recordings of every day language use, verbal art, songs, narratives, ritual and more.
- Access: All materials are digital and free of charge to access (registration needed)
- Collections can be browsed and accessed through the ELAR online catalogue:
<https://elar.soas.ac.uk/>
- Formats: mp4 (video), wav (audio)

(ELAR is committed to assist in finding the right archive for your materials and/or discuss a possible collaboration for archiving with ELAR.)



Dissemination via alternative routes

Project website: [Vegetative and minimally conscious states: Interviews with families](#)

In-depth video-recorded interviews with 22 family members of severely brain-injured patient, exploring people's experiences of having a son, daughter, sibling, partner, parent or other relative in a vegetative or minimally conscious state.

Youtube: [Conservation, markets and justice - Part 2: Ethnographic participatory video data](#)

This dataset includes a set of participatory videos exploring conceptions of environmental justice as a means to understand the tensions between biodiversity conservation and local livelihoods of people in the tropics. The archive also includes a participatory video photo story; a final report; a summary report of participatory video in the village of Ruhatwe and guidelines for ethnographic studies related to our project.

Vimeo: [Ethnographic documentary: De-Commodifying food ways](#)

Documentary offers an ethnographic eye on the joys, perils and challenges of those who live from sustainable agriculture and food production. Diving deep into the everyday rhythm of a couple of goat keepers, a collective of cheesemakers, a Marxist-inspired beekeeper and an herbal expert, the documentary takes the viewer on a journey through the difficulties of producing and selling products, loneliness and nature, unknown circuits of alternative micro-economies, as well as the incisiveness of contemporary neo-peasant activism as a source of inspiration.

EXERCISE



Exercise

- Evaluate example data collections and compare them against the FAIR principles
- Which considerations are important for YOUR data, to adhere to FAIR principles
- Consult re3data.org for suitable repositories for video data; evaluate:
 - Who can publish data via these systems?
 - How are datasets exposed / disseminated, so others can find them?
 - Can anyone download data or are there access controls in place?
 - Which licences and data use agreements are used?

Prepare to deposit with an archive

To prepare for archiving data:

- decide in advance if possible where the data will be archived in order to align with file format/documentation/metadata requirements
- ensure that nothing precludes data sharing in an archive (explicit consent)
- decide how to best organise data, using a logical, hierarchical structure that would be easy to understand
- bundle similar files:
 - by type of files: data/documentation/metadata
 - by type of data: quantitative/qualitative or transcripts/audio/video
- file naming - important to save time and avoid misunderstandings when multiple people collect/manage/use/re-use data at the same time.

How to prepare for archiving data in ReShare

- Familiarise yourself with:
 - Collections in our catalogue [Discover](#)
 - Video tutorial on [ReShare homepage](#)
 - Review procedures:
<http://reshare.ukdataservice.ac.uk/reshare-review-procedures/>
 - FAQs:
<https://www.ukdataservice.ac.uk/help/faq/reshare>

What is ReShare?

- UK Data Service's online data repository, where researchers can archive, publish and share research data: <http://reshare.ukdataservice.ac.uk>
- It allows depositors to create data collections themselves - upload data and documentation files;
- Once completed by depositors, collections go into a review area, where we check for disclosure risk, copyright breaches, validity of file formats and level of documentation;
- All numerical data files and at least a 10% random sample of textual data files are reviewed.

Short demo: Archiving data in ReShare

Logged in as mr user 01 Logout UK Data Service home Help About FAQ Contact

UK Data Service
ReShare

Home Legal Review procedures

Edit collection: [Data Collection #90](#)

My data

Terms and conditions → Grant details → People → **Data collection** → Upload → Deposit

* Data collection title ?

+ Alternative title

* Data description (abstract) ?

* Keywords ?

 Add

HOW RESEARCH DATA ARE REUSED

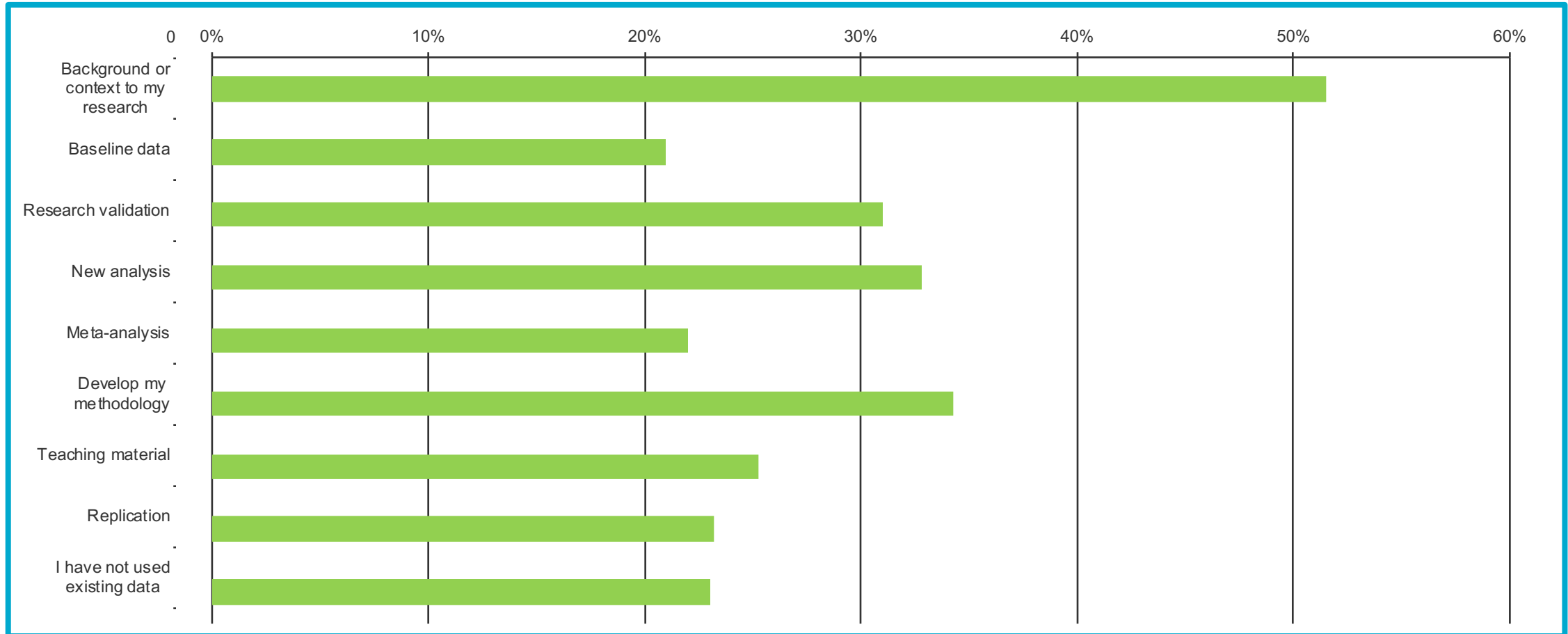


How are people re-using archived data?

- Descriptive material
- Comparative research, restudy or follow-up study
- Re-analysis/secondary analysis
- Replication
- Research design and methodological advancement
- Teaching and learning



Reuse of data (N=842)



Van den Eynden, Veerle et al. (2016) Towards Open Research: Practices, experiences, barriers and Opportunities. Wellcome Trust. <https://dx.doi.org/10.6084/m9.figshare.4055448>

UK Data Service



Analysing reuse of qualitative data

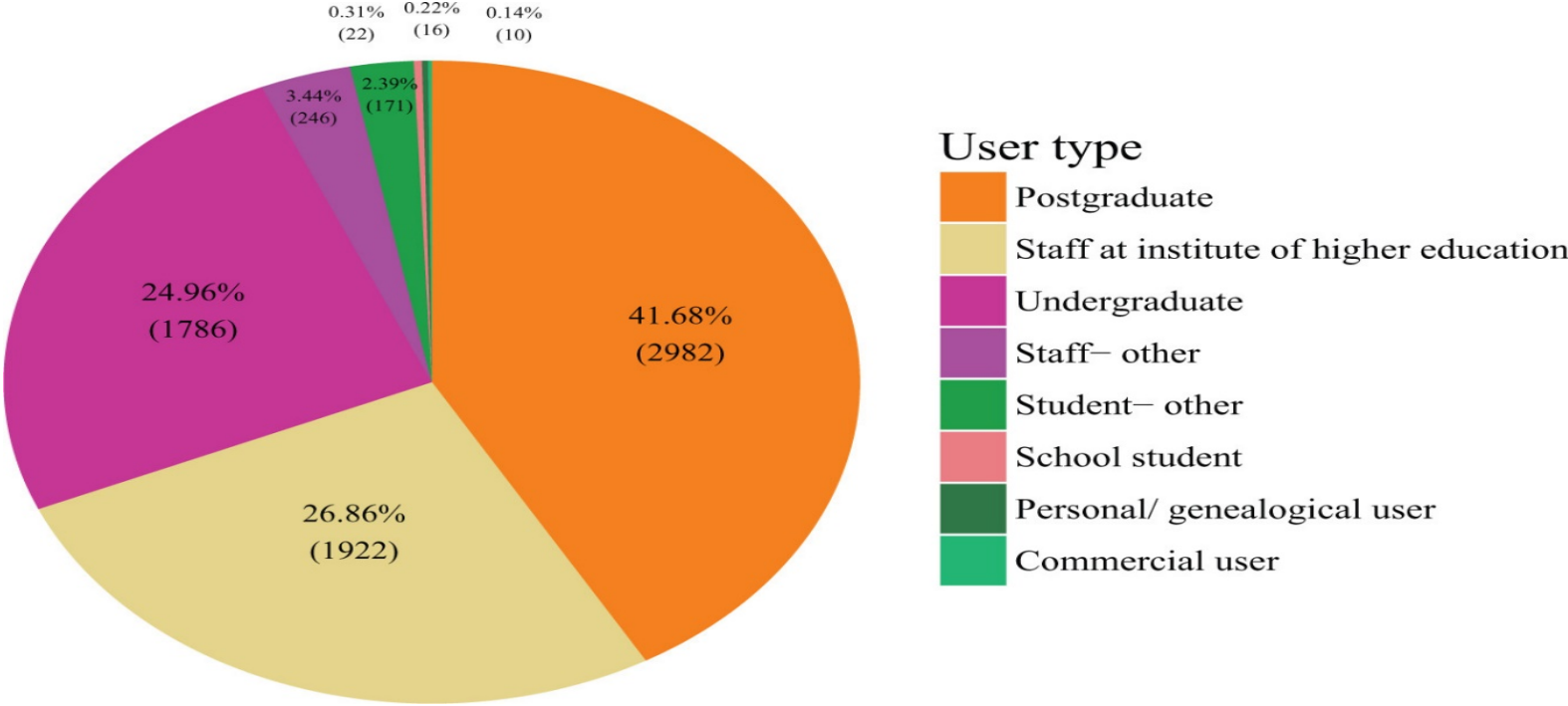
- UK: 444 datasets archived by UK Data Service (UKDS)
 - reuse metrics 2002-2016
- Finland: 177 datasets archived by Finnish Social Science Data Archive (FSD)
 - reuse metrics 2014-2016
- Majority of reusers: postgraduates, undergraduates
- Much reuse for learning and teaching
- 11-15% reuse in research

Bishop & Kuula-Luumi (2017) Revisiting Qualitative Data Reuse: A Decade On. SAGE Open 7 (1).
<https://doi.org/10.1177/2158244016685136>

UK Data Service



Who reuses qualitative data

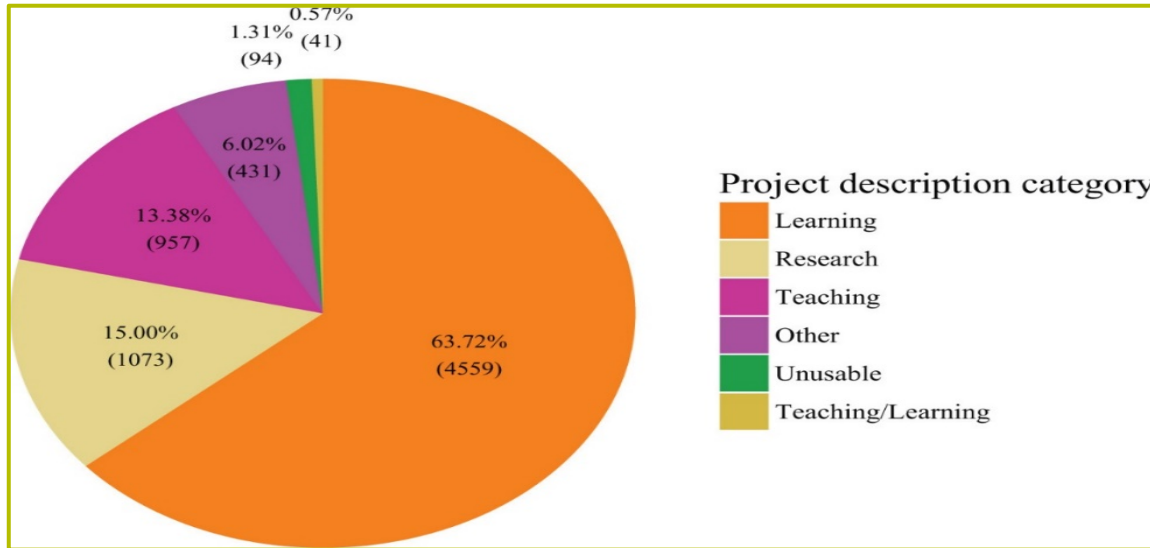


Downloaded data by type of user, 2002-2016, UK Data Service

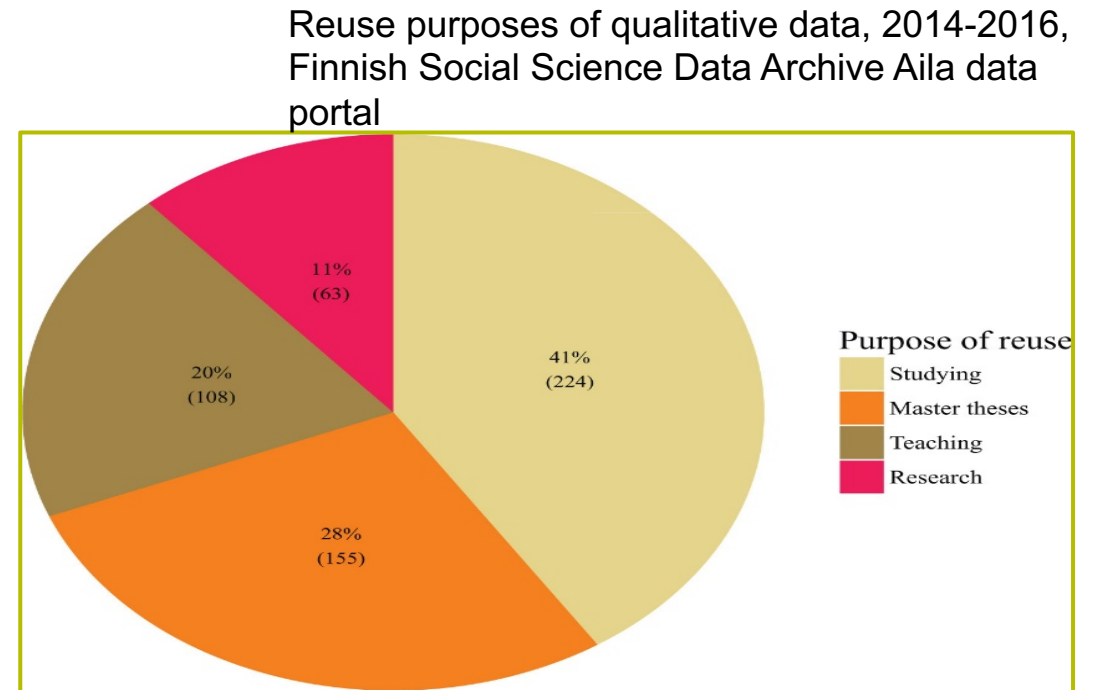
Published in: Libby Bishop; Arja Kuula-Luumi; *SAGE Open* 7,
DOI: 10.1177/2158244016685136
Copyright © 2017 SAGE Publications



How qualitative data are reused



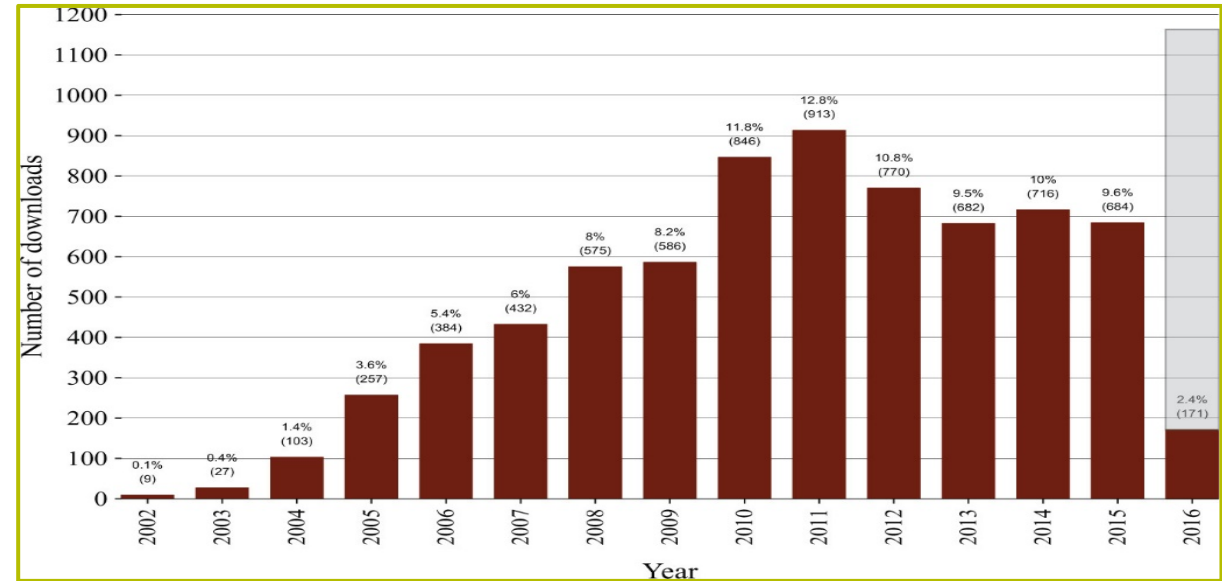
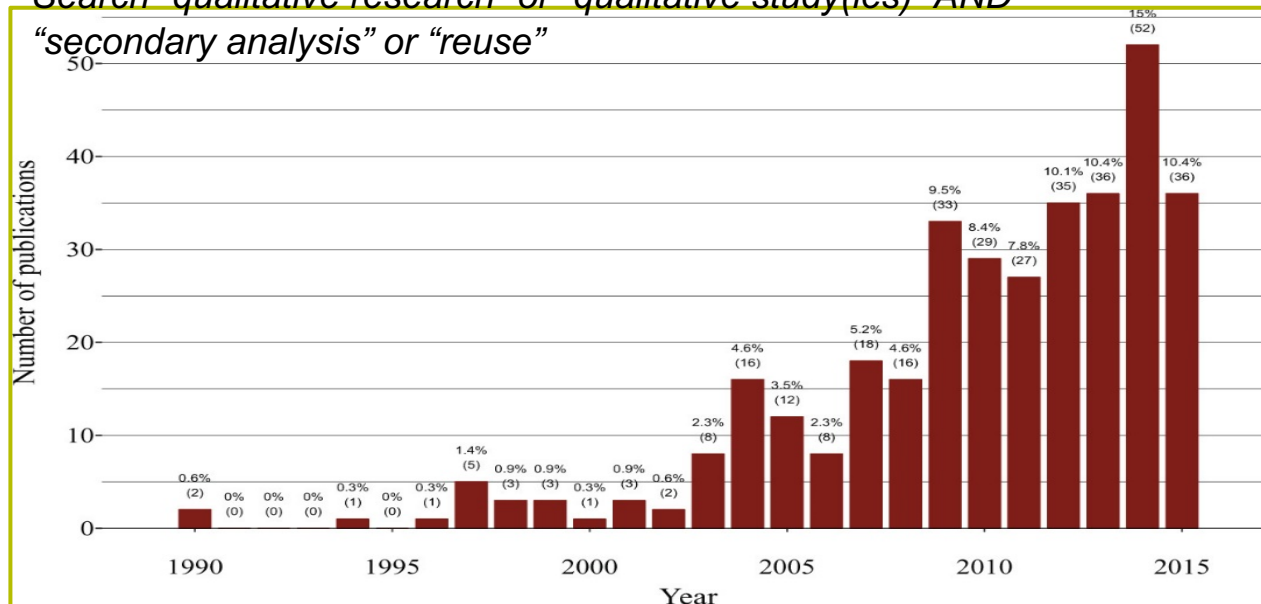
Reuse purposes of qualitative data, 2002-2016, UK Data Service



Reuse trends

Publications likely to have used qualitative data for secondary analysis—Web of Science, 1990-2015

Search “qualitative research” or “qualitative study(ies)” AND “secondary analysis” or “reuse”



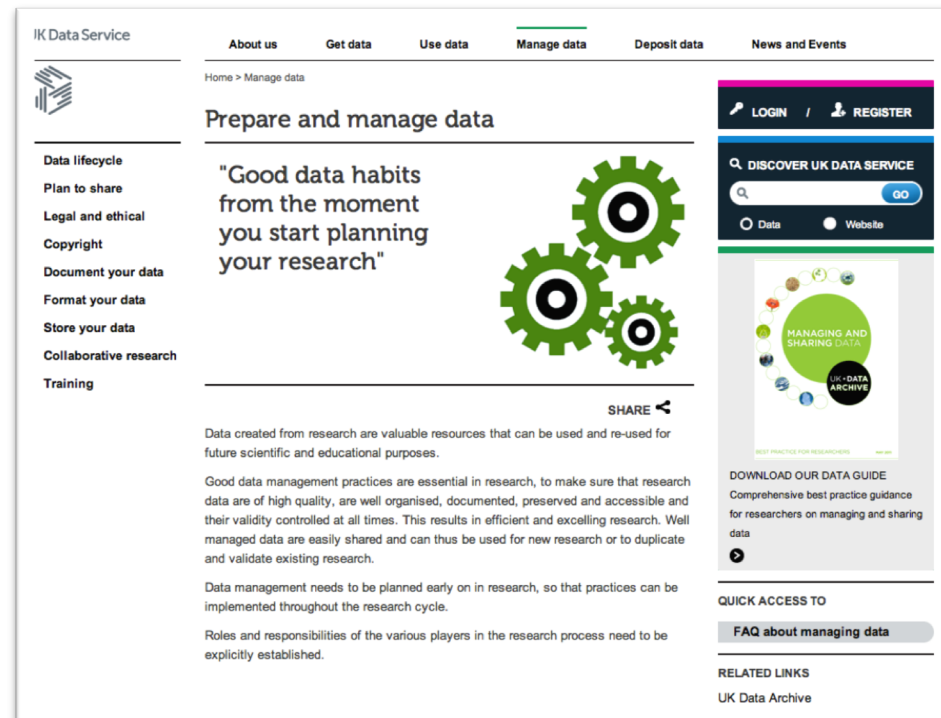
Number of data downloads per year, 2002-2016, UK Data Service

Published in: Libby Bishop; Arja Kuula-Luumi; *SAGE Open* 7, DOI: 10.1177/2158244016685136
Copyright © 2017 SAGE Publications



UKDS 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.data-archive.ac.uk/create-manage/advice-training/events



The screenshot displays the UK Data Service website's 'Manage data' page. The navigation bar includes 'About us', 'Get data', 'Use data', 'Manage data', 'Deposit data', and 'News and Events'. The main heading is 'Prepare and manage data'. A quote reads: "Good data habits from the moment you start planning your research". Below the quote are three interlocking green gears. A 'SHARE' button is visible. The page contains several paragraphs of text regarding data management practices and their importance in research. On the right side, there is a search bar, a 'LOGIN / REGISTER' button, and a 'DISCOVER UK DATA SERVICE' search box. Below the search bar, there are radio buttons for 'Data' and 'Website'. A 'QUICK ACCESS TO' section highlights 'FAQ about managing data'. At the bottom, there is a 'RELATED LINKS' section with a link to 'UK Data Archive'.



Questions?



<https://pbs.twimg.com/media/B7ZUntnCUAEQAgR.jpg>