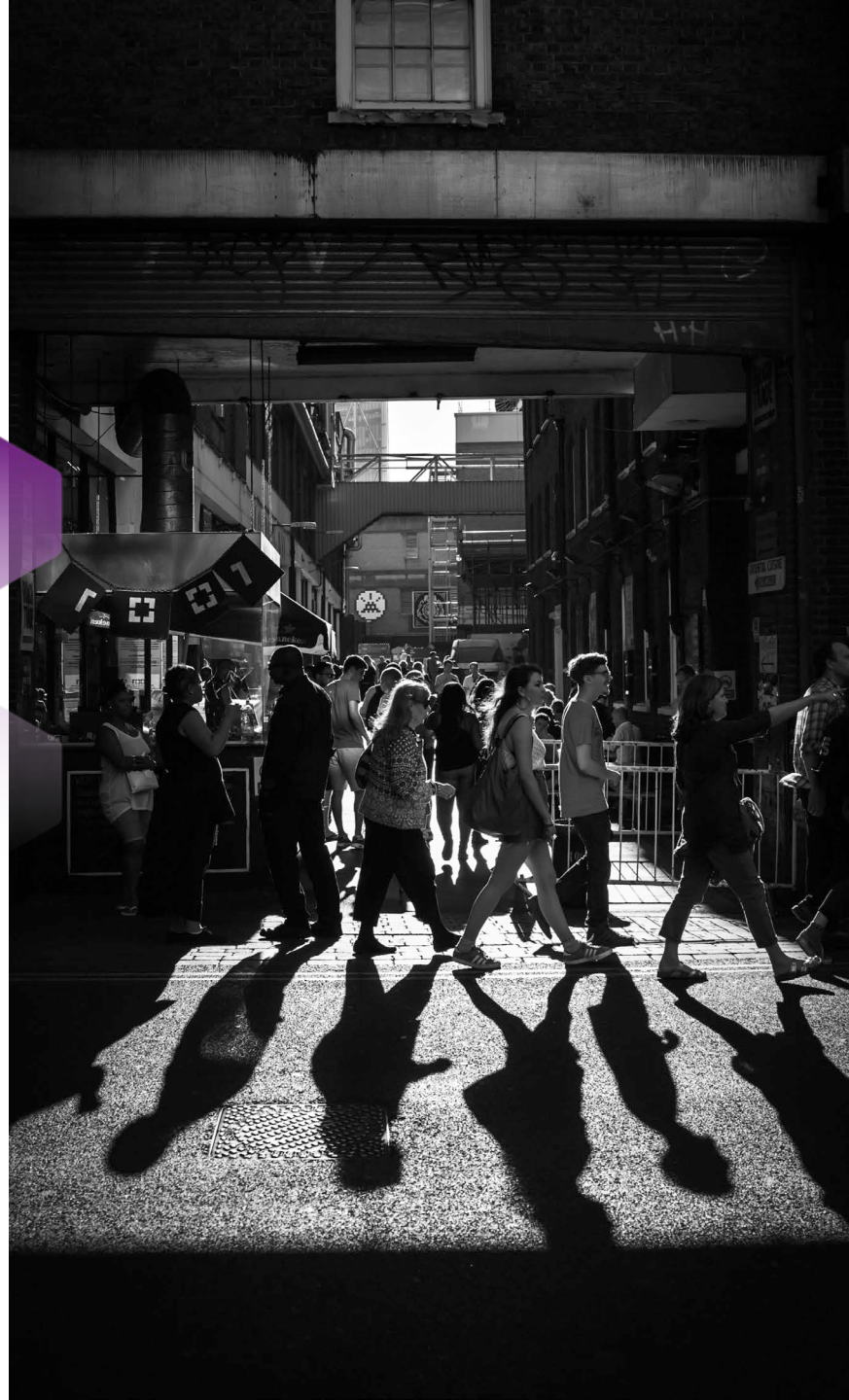
UK Data Service

How to Document Quantitative and Qualitative Data

Anca Vlad

Maureen Haaker



Economic and Social Research Council

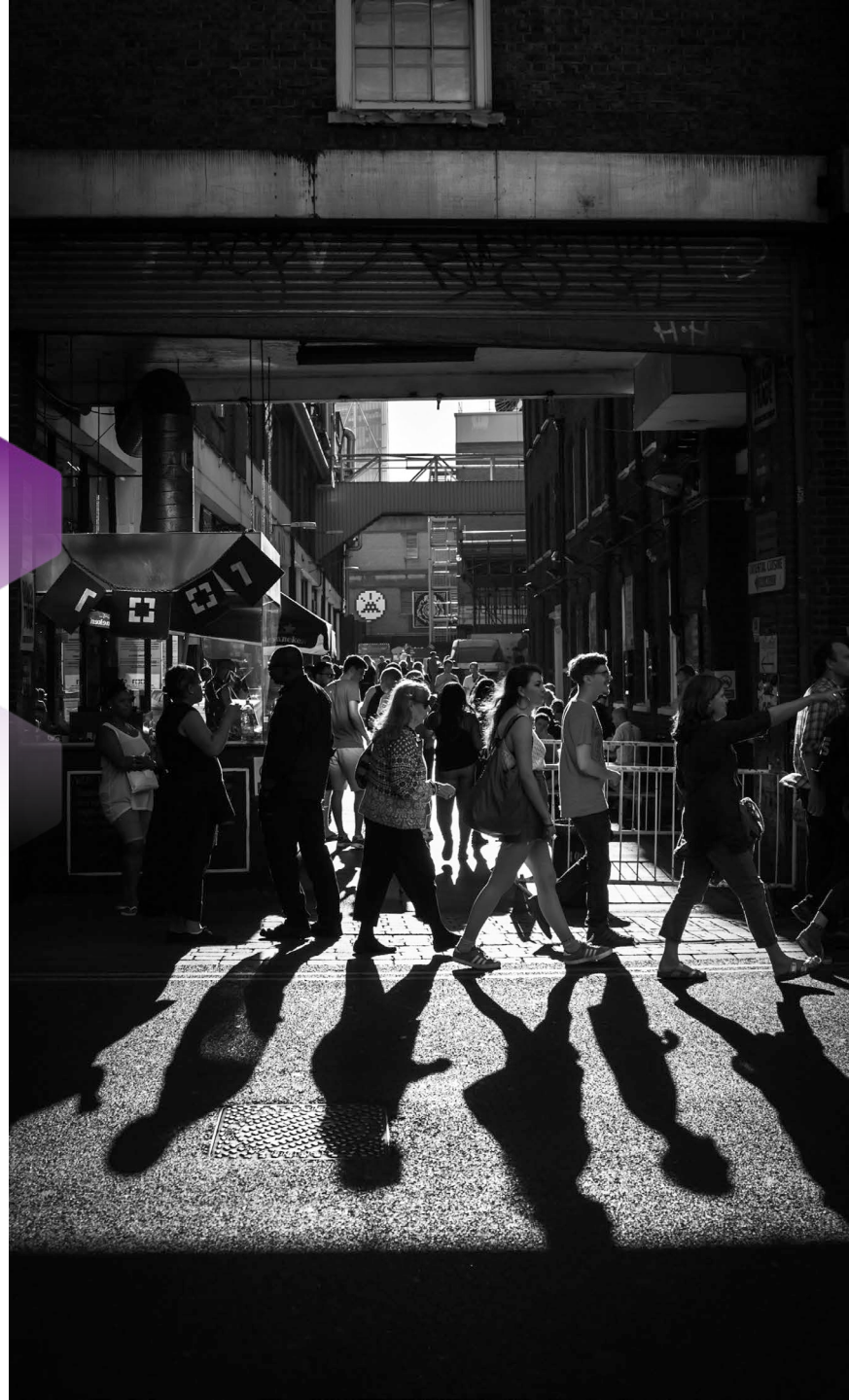
Copyright © 2023 University of Essex. Created by UK Data Archive, UK Data Service.

UK Data Service

How to Document Quantitative and Qualitative Data

Anca Vlad

Maureen Haaker



Economic and Social Research Council

Copyright © 2023 University of Essex. Created by UK Data Archive, UK Data Service.

Overview

- Introduction: Documenting Data and FAIR principles
- Documentation for Qualitative Data
- Documentation for Quantitative Data
- Metadata
- Exercise: Mentimeter
- Q&A
- Further resources and questions

Documenting Data – What is it?

Data documentation provides essential context for the data itself. There are two levels of documentation:

Project-level documentation or collection level documentation, is the information available about the research project as a whole, such as: abstract, keywords, methodology, who collected the data and when, topics, geography, how can the data be accessed.

Data-level documentation also referred to as object-level documentation, is information about the individual units/objects (i.e. interview transcript, focus group, variables), for example: interview/focus group header, variable and value labels.

Some potentially new terms?

- “User guide”
- “Data listing”
- “Data dictionary”
- “Readme file”
- “Research materials”
- "Data assets"
- "Metadata", including "DDI“

Visit CODATA’s pages for more information about terminology: https://terms4fairskills.github.io/.

Documenting Data – Why do it?

- Maximises re-use value of data
- Provides provenance for future, historical use of data
- Space to expand on methods and processes not normally covered in published work
- Teaches others on good research practice
- Demonstrates transparency and can facilitate replication where appropriate
- Enhances research outputs
- Allows data to be used correctly and efficiently in future research projects
- Essential for archives/repositories to be able to review and prepare data for publication
- Aids creation of metadata for FAIR research

Practical activity

Background to FAIR principles

- Data landscape is changing rapidly – large volumes of complex data is constantly being created and shared for a huge range of purposes (in academia and beyond). Discussions on accessibility have reflected on this for some time, noting the challenges of data which couldn't be easily processed with machines.
- Good data management is a catalyst for knowledge discovery, innovation, and constructive collaborations. But what is good data management?
- Published in 2016 in *Scientific Data*, FAIR principles are outlined as guidelines to help define what good data management looks like, emphasising machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention)
- In 2017, Germany, Netherlands, and France established an international office to support the FAIR initiative, the GO FAIR International Support and Coordination Office.
- FAIR principles have since been embedded into policies of Research Data Alliance, Association of European Research Libraries, as well as funders in the UK, including UKRI and its research councils.

FAIR Principles

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 right)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, they need to know how they can be accessed, *possibly including authentication and authorisation*.

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

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage and processing.

- 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

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- 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

UK Data Service

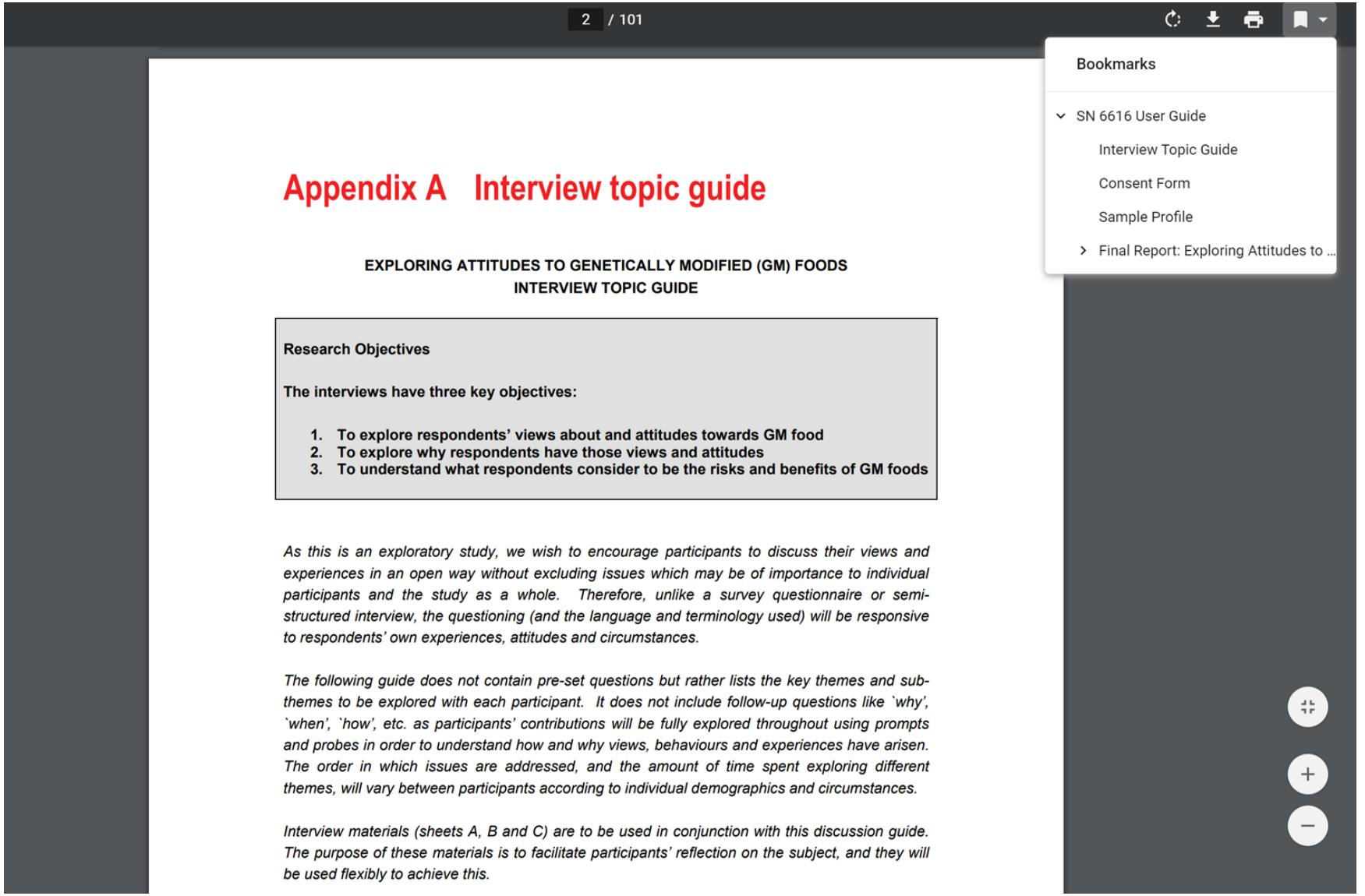
Documentation for Qualitative Data

- Good examples of documentation
- Utilising changing technology
- Re-using documentation

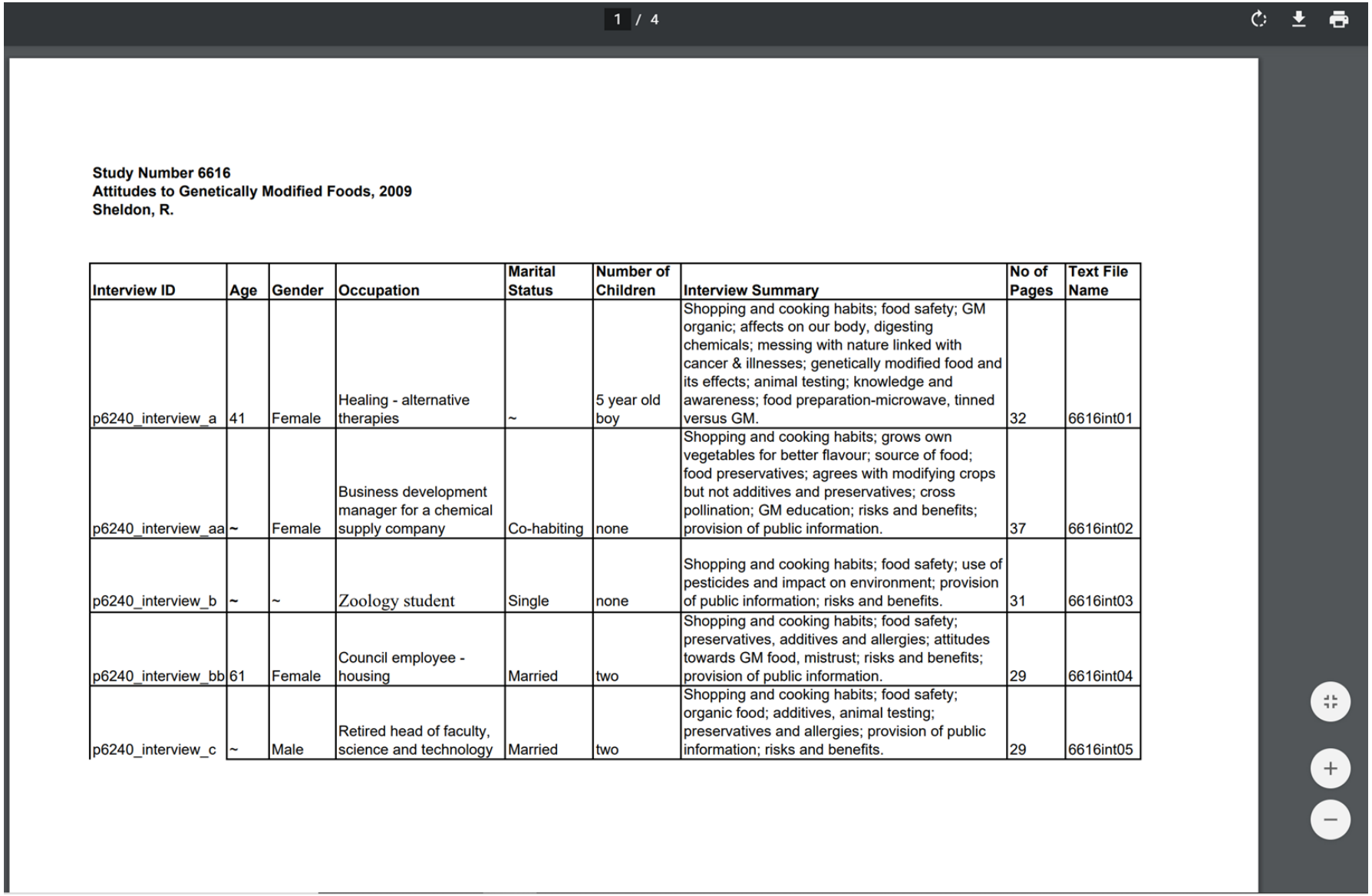
Examples of qualitative documentation

- Interview topics, guides, prompts, or cue cards
- Blank consent form, participant information sheet, or call for recruitment
- Sampling strategy
- Instructions for research team
- Extracts from published work on methodology
- Funding applications or ethics proposal
- End of award reports
- Research meeting minutes
- Field notes/memos that form analysis
- Draftwork of initial analyses
- Codebooks or coding scheme
- Project websites...

User Guide


2 / 101
Bookmarks
SN 6616 User Guide
Interview Topic Guide
Consent Form
Sample Profile
Final Report: Exploring Attitudes to ...
Appendix A Interview topic guide
EXPLORING ATTITUDES TO GENETICALLY MODIFIED (GM) FOODS
INTERVIEW TOPIC GUIDE
Research Objectives
The interviews have three key objectives:
1. To explore respondents' views about and attitudes towards GM food
2. To explore why respondents have those views and attitudes
3. To understand what respondents consider to be the risks and benefits of GM foods
As this is an exploratory study, we wish to encourage participants to discuss their views and experiences in an open way without excluding issues which may be of importance to individual participants and the study as a whole. Therefore, unlike a survey questionnaire or semi-structured interview, the questioning (and the language and terminology used) will be responsive to respondents' own experiences, attitudes and circumstances.
The following guide does not contain pre-set questions but rather lists the key themes and sub-themes to be explored with each participant. It does not include follow-up questions like `why', `when', `how', etc. as participants' contributions will be fully explored throughout using prompts and probes in order to understand how and why views, behaviours and experiences have arisen. The order in which issues are addressed, and the amount of time spent exploring different themes, will vary between participants according to individual demographics and circumstances.
Interview materials (sheets A, B and C) are to be used in conjunction with this discussion guide. The purpose of these materials is to facilitate participants' reflection on the subject, and they will be used flexibly to achieve this.

Data List



Study Number 6616
Attitudes to Genetically Modified Foods, 2009
Sheldon, R.

Interview ID	Age	Gender	Occupation	Marital Status	Number of Children	Interview Summary	No of Pages	Text File Name
p6240_interview_a	41	Female	Healing - alternative therapies	~	5 year old boy	Shopping and cooking habits; food safety; GM organic; affects on our body, digesting chemicals; messing with nature linked with cancer & illnesses; genetically modified food and its effects; animal testing; knowledge and awareness; food preparation-microwave, tinned versus GM.	32	6616int01
p6240_interview_aa	~	Female	Business development manager for a chemical supply company	Co-habiting	none	Shopping and cooking habits; grows own vegetables for better flavour; source of food; food preservatives; agrees with modifying crops but not additives and preservatives; cross pollination; GM education; risks and benefits; provision of public information.	37	6616int02
p6240_interview_b	~	~	Zoology student	Single	none	Shopping and cooking habits; food safety; use of pesticides and impact on environment; provision of public information; risks and benefits.	31	6616int03
p6240_interview_bb	61	Female	Council employee - housing	Married	two	Shopping and cooking habits; food safety; preservatives, additives and allergies; attitudes towards GM food, mistrust; risks and benefits; provision of public information.	29	6616int04
p6240_interview_c	~	Male	Retired head of faculty, science and technology	Married	two	Shopping and cooking habits; food safety; organic food; additives, animal testing; preservatives and allergies; provision of public information; risks and benefits.	29	6616int05

Some excellent documentation

Anti-Politics: Characterising and Accounting for Political Disaffection, 2011-2012

https://discover.ukdataservice.ac.uk/catalogue/?sn=7855&type=Data%20catalogue

Availability: UK Data Service

Contact: Get in touch

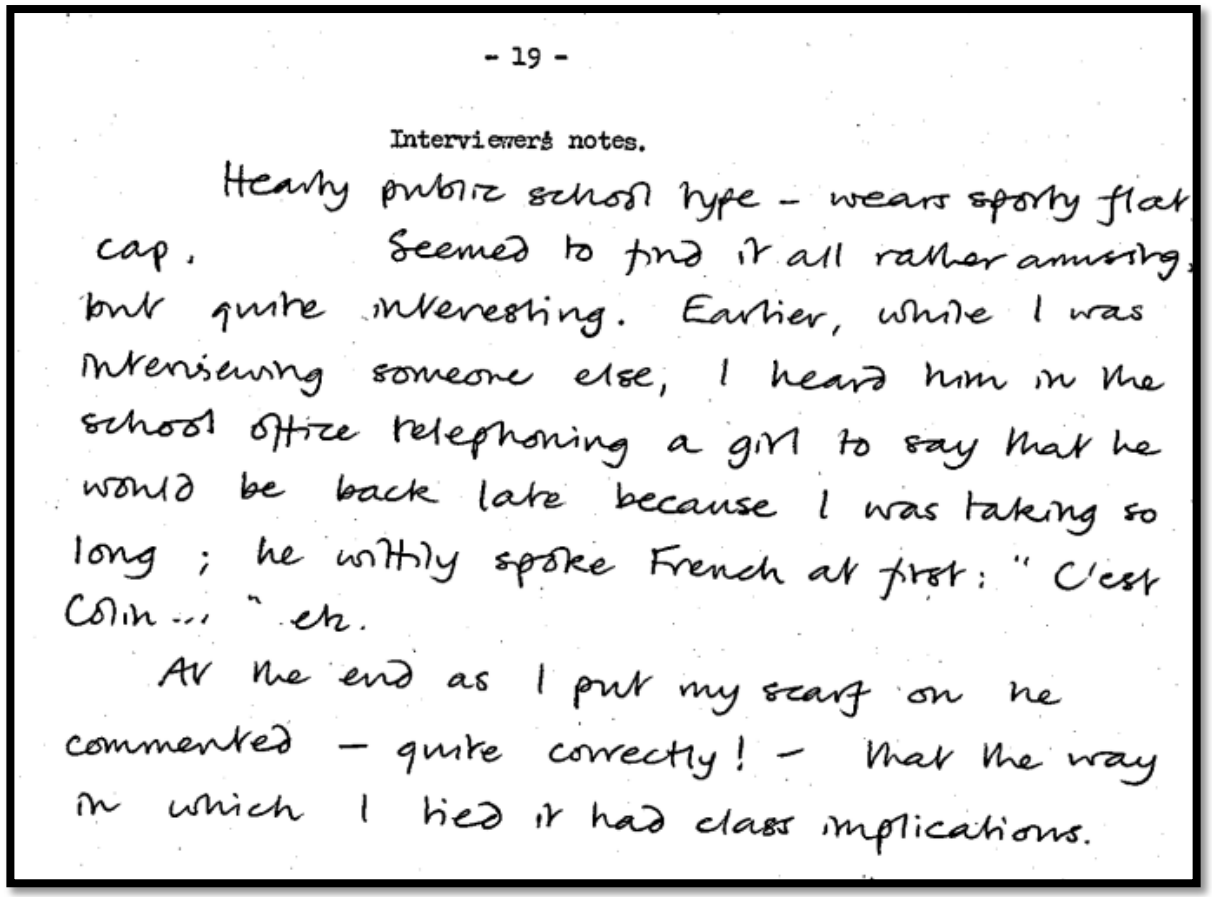
DOCUMENTATION

Title	File Name	Size (KB)
Data List (Excel)	7855_ulist.xls	34
Demographic Survey Questionnaire	7855_demographic_survey.pdf	248
ESRC End of Award Report	7855_eoa_report_res-000-22-4441.pdf	307
Focus Group Resources	7855_focus_group_resources.pdf	270
Consent Form	7855_research_consent_form.pdf	240
Participant Information Sheet	7855_research_information_sheet.pdf	248
Data List (PDF)	7855_ulist.pdf	84
Study information and citation	UKDA_Study_7855_Information.htm	6
READ File	read7855.htm	10

RELATED STUDIES AND GUIDES

+ View related studies and guides...

Affluent Worker: Interviewer comments

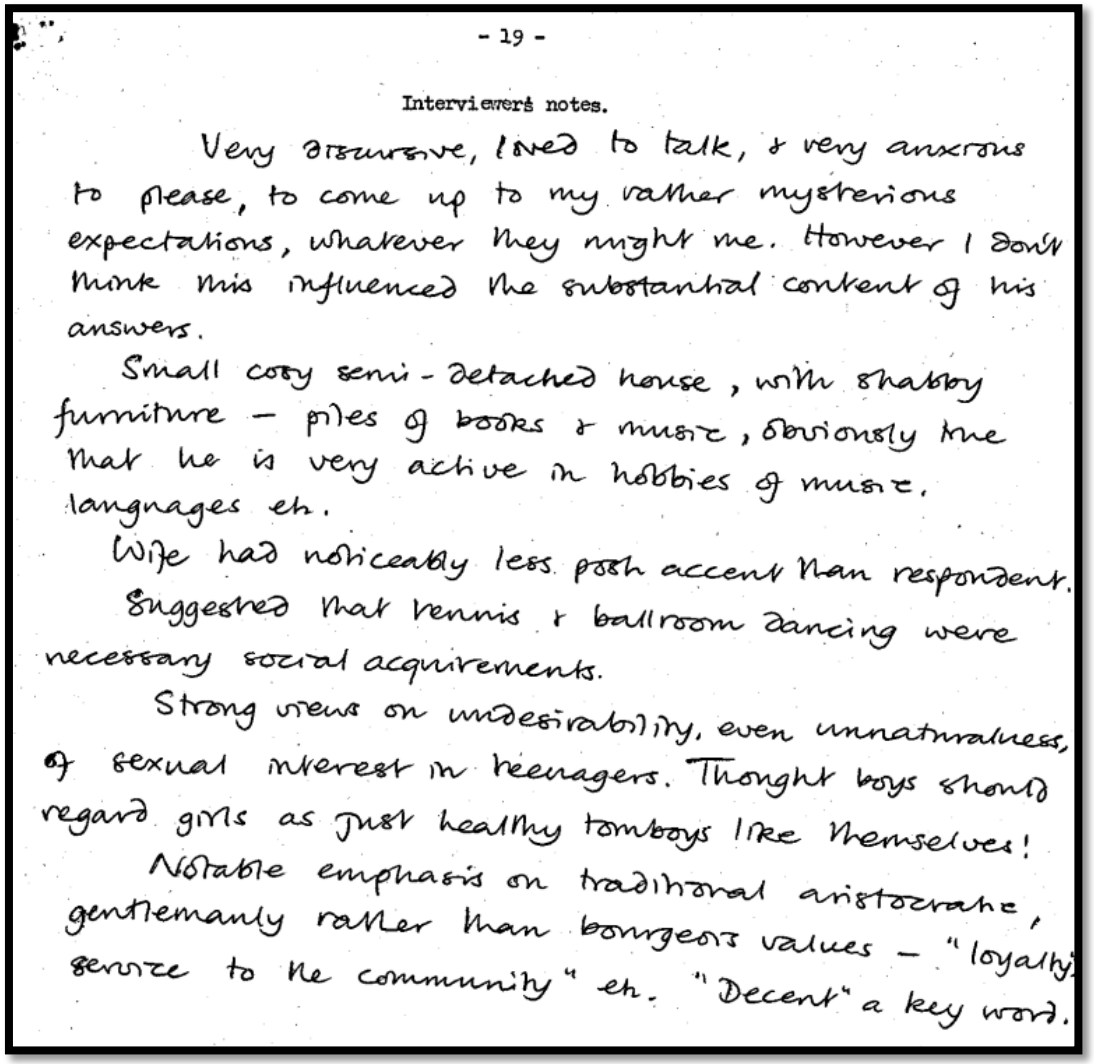


- 19 -

Interviewer's notes.

Hearty public school type - wears sporty flat cap. Seemed to find it all rather amusing, but quite interesting. Earlier, while I was interviewing someone else, I heard him in the school office telephoning a girl to say that he would be back late because I was taking so long ; he wittily spoke French at first: " C'est Colin ... " eh.

At the end as I put my scarf on he commented - quite correctly! - that the way in which I tied it had class implications.



- 19 -

Interviewer's notes.

Very discursive, loved to talk, & very anxious to please, to come up to my rather mysterious expectations, whatever they might be. However I don't think this influenced the substantial content of his answers.

Small cosy semi-detached house, with shabby furniture - piles of books & music, obviously true that he is very active in hobbies of music, languages eh.

Wife had noticeably less posh accent than respondent.

Suggested that tennis & ballroom dancing were necessary social acquirements.

Strong views on undesirability, even unnaturalness, of sexual interest in teenagers. Thought boys should regard girls as just healthy tomboys like themselves!

Notable emphasis on traditional aristocratic, gentlemanly rather than bourgeois values - "loyalty, service to the community" eh. "Decent" a key word.

https://discover.ukdataservice.ac.uk/catalogue/?sn=6512&type=Data%20catalogue

Marsden's reflection on "felt poverty"

8 'Felt' poverty

I'm not sure whether we need a separate section on how mothers feel poverty. Perhaps the feeling may best come through indirectly, and all this section will be subsumed in the others. I notice that I've already begun to use bits of material in the section on expenditure. My reason for including this title in the section headings is that we can describe objective conditions of life and quantities of this and that, but how do the mothers feel about it? Do they actually feel themselves to be worse off?

It seems obvious, but it's worth saying, that nobody felt the scales were adequate. On the other hand, most of them wouldn't say how much they needed, and the reason for this reluctance appeared twofold. They were already self-conscious about receiving 'charity', though the feeling wore off after a time, and also the long time of pinching and scraping had become a habit of mind. They just couldn't think about wants and needs, they daren't let their appetites be titillated by thoughts of more money and nice food and possessions. As one said, when going shopping her constant thoughts were 'can I do without this, can I do without that'. People were sensitive about poverty in different degrees which was quite unrelated to their needs. Mrs Dufay was one of the

https://discover.ukdataservice.ac.uk/catalogue/?sn=5072&type=Data%20catalogue

Sampling

APPENDIX B

THE SAMPLE

WEDNESDAY, JANUARY 10TH, 1983 - 7 P.M.

BRUNEL UNIVERSITY

A comfortable common room. Low tables with plates of sandwiches on them and cups of coffee. A microphone looking like a minature radar transmitter sits on one of them. Nine people in upholstered low chairs,arranged in an oval shape, sit round the tables. Six are visitors. Three men and three women - people who have responded to the appeal in the Sunday Times to take part in the study. Two are research assistants - a man working on the study with me and a woman, who is a graduate research student in the department conducting her own research on a completely different project. She has come to take charge of the recording and will later transcribe it. The ninth person is me.

I welcome everyone, thank them for their time and reassure them no-one should feel under any compulsion to speak nor to say such things that at the end of the evening they leave feeling ashamed and wishing they had not come. I tell them I hope they will enjoy the discussion. I spend several minutes on this introduction, waiting to sense that everyone is comfortable and understands the research purposes of the meeting by talking about the way they have been approached - via a newspaper article, by talking about the problems of samples selected in this way. One person takes up this point and asks whether I will not have too many people from similar economic backgrounds. I reply that I am not particularly concerned about the lack of a representative sample of the general population because the kinds of statements I wish to make do not require such a sample:

> "It depends what kind of statement you wish to make at the end of the day whether you need a representative sample of the population. I am not trying to draw any conclusions from my work about incidence or prevalence. I will not be saying 'In Britain today x% of the population has committed adultery' or 'Of men in Britain today, half will have committed adultery by the time they have been married ten years'. For that kind of statement, a representative

Cont../

> sample of the general population would indeed be needed. Actually for this kind of study there are respectable arguments to be made for choosing a fairly homogeneous group. But there is another problem which is more difficult to deal with - you are all self-selected. That is, you have chosen to take part in the study. I need to understand why. People may say you are a peculiar group simply because you are prepared to talk about your lives and about this very personal topic. What I'd like to do, therefore, is to begin by asking you all to try and cast your minds back to the moment when you decided to pick up a pen and pencil, or when you went and sat down at your typewriter and decided to write to me. What were you feeling or thinking? I know it is difficult and some time ago, but if you could cast your minds back and try to remember what it was that decided you ...; what was it you wanted to say to me ...?"

Dennis:

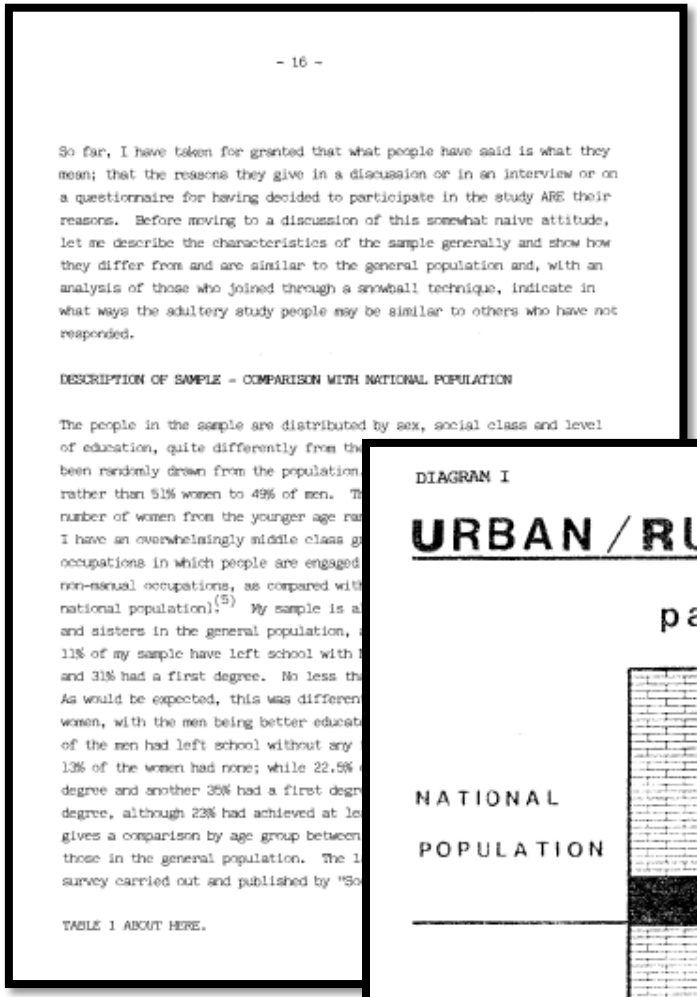
> "Well, to be honest I just wanted confirmation that there are other people like me that indulged in it - how can I put it - without necessary feelings of guilt ..."

Dennis spoke for some time, going on to describe the death of his young first wife. She was killed by a hit-and-run driver. This wife had been "so carnal" he said, "there was nothing she did not enjoy about the sex act, nothing she wouldn't do, that for me to have had an adulterous relationship and to be married to her would have been a physical impossibility". As part of his effort to cope with her loss he had spent some time in California. There, he said, even his best fantasies were met where, he told us, he found the women were "promiscuous". A further prompt from me:

> "So what actually spurred you to write to me?"

https://discover.ukdataservice.ac.uk/catalogue/?sn=4858&type=Data%20catalogue

More on sampling…



- 16 -

So far, I have taken for granted that what people have said is what they mean; that the reasons they give in a discussion or in an interview or on a questionnaire for having decided to participate in the study ARE their reasons. Before moving to a discussion of this somewhat naive attitude, let me describe the characteristics of the sample generally and show how they differ from and are similar to the general population and, with an analysis of those who joined through a snowball technique, indicate in what ways the adultery study people may be similar to others who have not responded.

DESCRIPTION OF SAMPLE - COMPARISON WITH NATIONAL POPULATION

The people in the sample are distributed by sex, social class and level of education, quite differently from th
been randomly drawn from the population.
rather than 51% women to 49% of men. T
number of women from the younger age ra
I have an overwhelmingly middle class g
occupations in which people are engaged
non-manual occupations, as compared wit
national population).[5] My sample is a
and sisters in the general population,
11% of my sample have left school with
and 31% had a first degree. No less th
As would be expected, this was differen
women, with the men being better educat
of the men had left school without any
13% of the women had none; while 22.5%
degree and another 35% had a first degr
degree, although 23% had achieved at le
gives a comparison by age group between
those in the general population. The 1
survey carried out and published by "So

TABLE 1 ABOUT HERE.

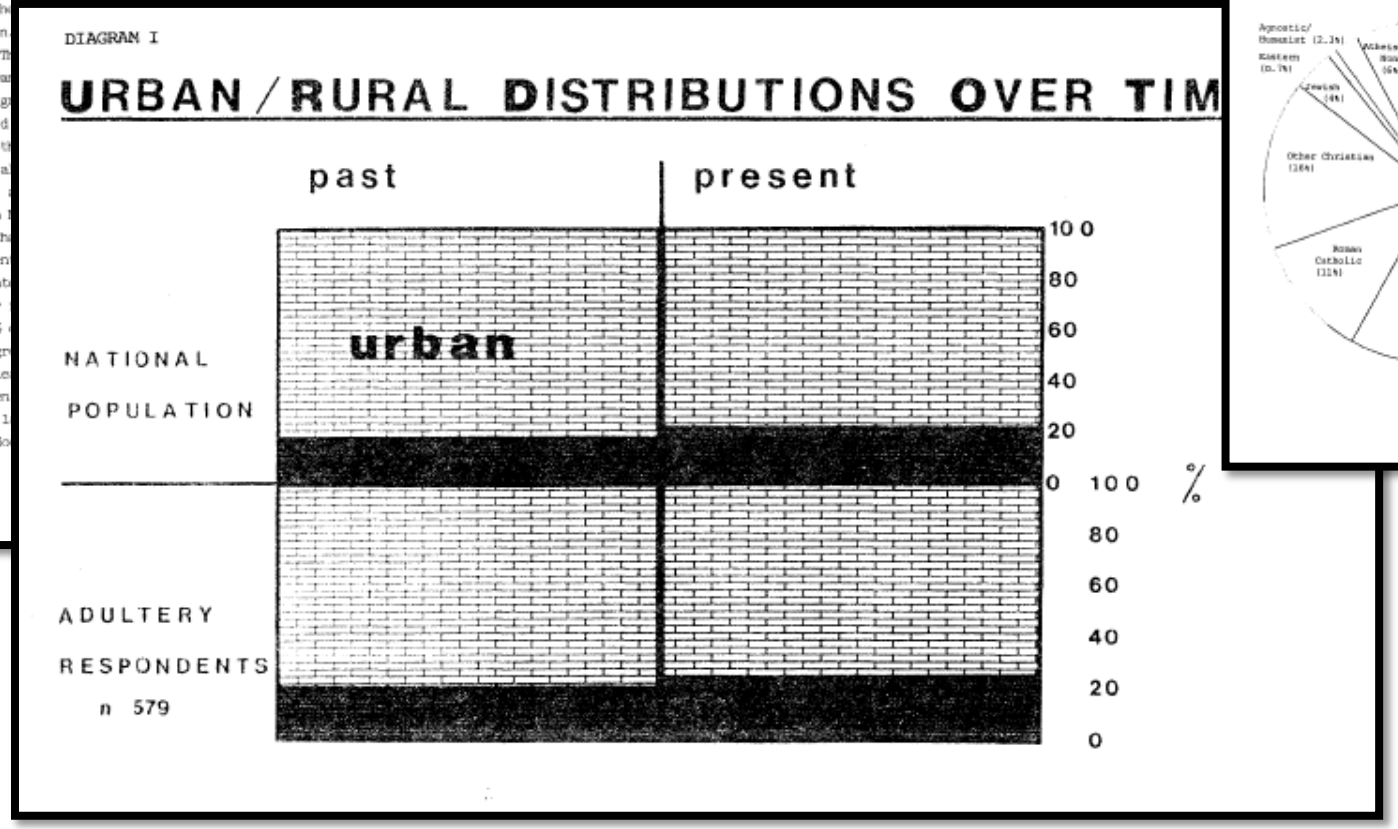
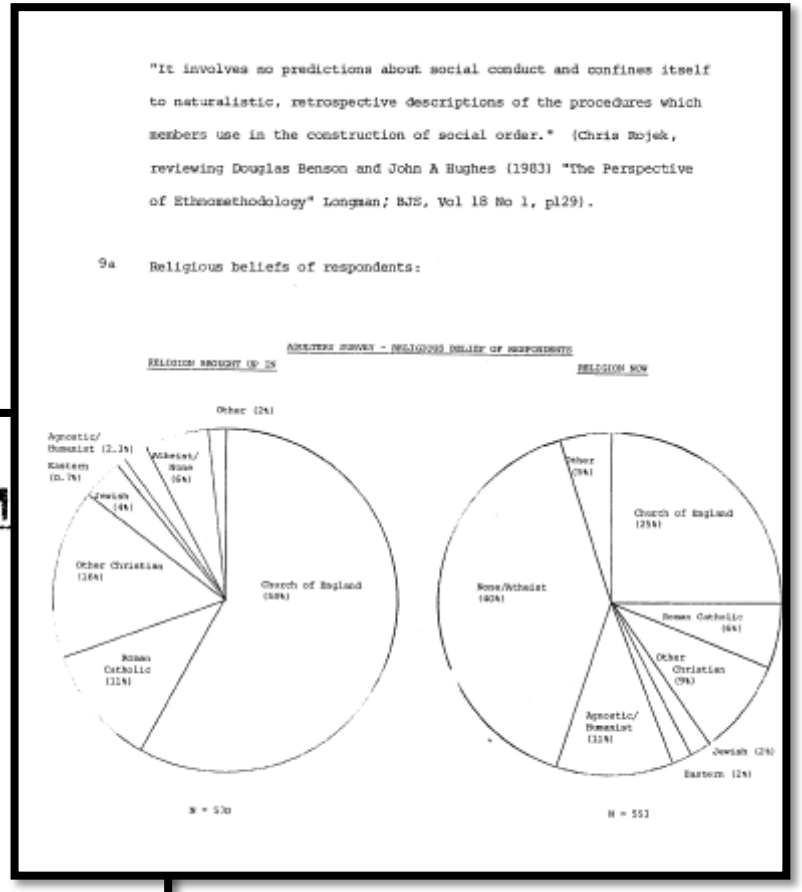


DIAGRAM I

URBAN/RURAL DISTRIBUTIONS OVER TIM



"It involves no predictions about social conduct and confines itself to naturalistic, retrospective descriptions of the procedures which members use in the construction of social order." (Chris Rojek, reviewing Douglas Benson and John A Hughes (1983) "The Perspective of Ethnomethodology" Longman; BJS, Vol 18 No 1, p129).

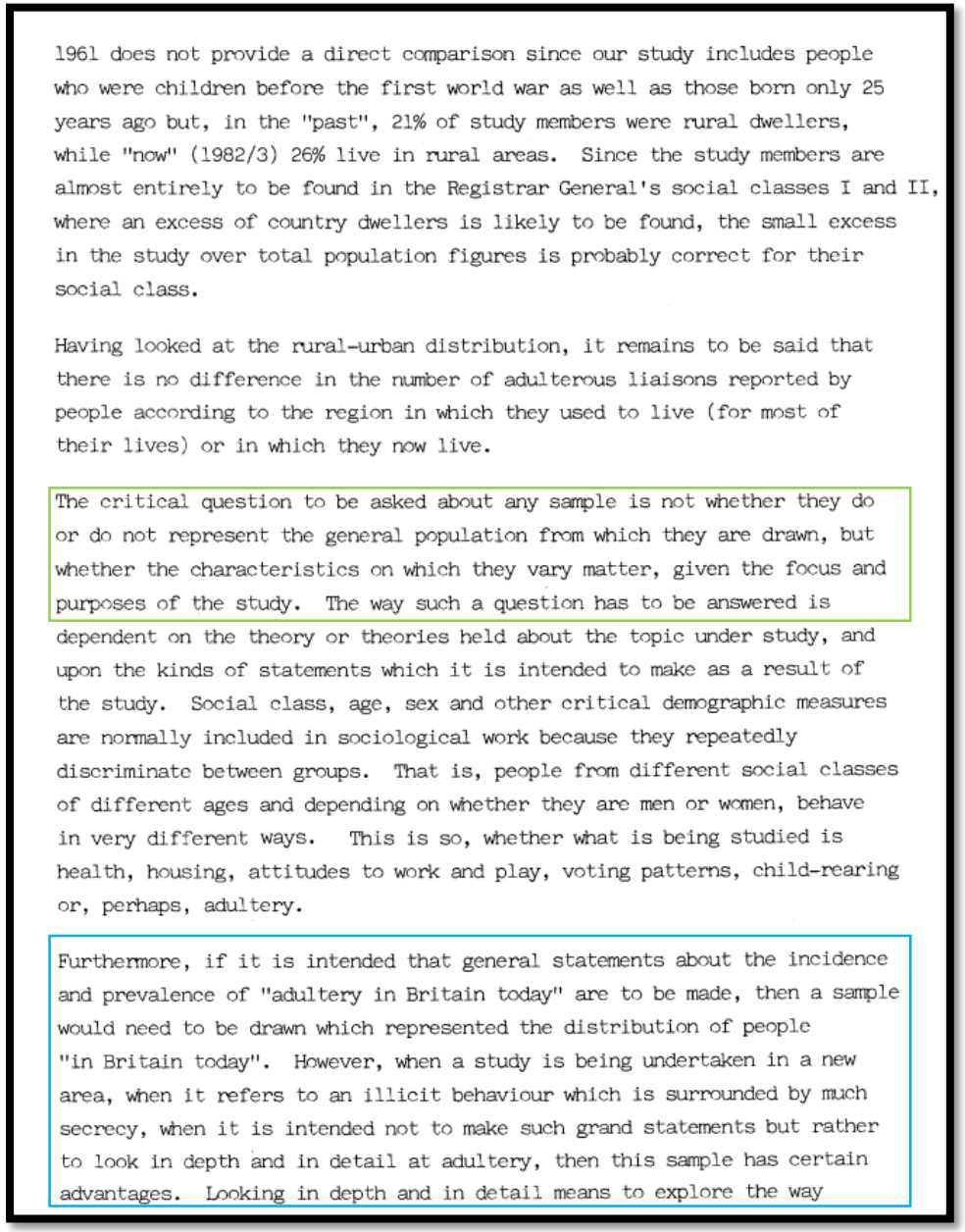
9a Religious beliefs of respondents:

https://discover.ukdataservice.ac.uk/catalogue/?sn=4858&type=Data%20catalogue

…and even more on sampling

The critical question to be asked about any sample is not whether they do or do not represent the general population from which they are drawn, but whether the characteristics on which they vary matter, given the focus and purposes of the study…

Furthermore, if it is intended that general statements about the incidence and prevalence of “adultery in Britain today” are to be made, then a sample would need to be drawn which represented the distribution of people “in Britain today”. However, when a study is being undertaken in a new area, when it refers to an illicit behaviour which is surrounded by much secrecy, when it is intended not to make such grad statements but rather to look in depth and in detail at adultery, then this sample has certain advantages…



1961 does not provide a direct comparison since our study includes people who were children before the first world war as well as those born only 25 years ago but, in the "past", 21% of study members were rural dwellers, while "now" (1982/3) 26% live in rural areas. Since the study members are almost entirely to be found in the Registrar General's social classes I and II, where an excess of country dwellers is likely to be found, the small excess in the study over total population figures is probably correct for their social class.

Having looked at the rural-urban distribution, it remains to be said that there is no difference in the number of adulterous liaisons reported by people according to the region in which they used to live (for most of their lives) or in which they now live.

The critical question to be asked about any sample is not whether they do or do not represent the general population from which they are drawn, but whether the characteristics on which they vary matter, given the focus and purposes of the study. The way such a question has to be answered is dependent on the theory or theories held about the topic under study, and upon the kinds of statements which it is intended to make as a result of the study. Social class, age, sex and other critical demographic measures are normally included in sociological work because they repeatedly discriminate between groups. That is, people from different social classes of different ages and depending on whether they are men or women, behave in very different ways. This is so, whether what is being studied is health, housing, attitudes to work and play, voting patterns, child-rearing or, perhaps, adultery.

Furthermore, if it is intended that general statements about the incidence and prevalence of "adultery in Britain today" are to be made, then a sample would need to be drawn which represented the distribution of people "in Britain today". However, when a study is being undertaken in a new area, when it refers to an illicit behaviour which is surrounded by much secrecy, when it is intended not to make such grand statements but rather to look in depth and in detail at adultery, then this sample has certain advantages. Looking in depth and in detail means to explore the way

https://discover.ukdataservice.ac.uk/catalogue/?sn=4858&type=Data%20catalogue

Field notes

- There are a lot of processes that the average villager needs to go through in order to simply get building materials/ Need to pay a lot of money/ It didn't used to be like that/ When asked when this change began, the participant stated that it was when Kikwete (current TZ president) took power [This is the CUF political party representative – whilst I do not doubt that there are some real hassles in 'simply cutting building poles,' I wonder if it really has any connection to Kikwete's inauguration, 2002-2005].
- Fire which is lit by people to clear land for agriculture gets out of control and burns the forest causing some of the trees to die
- Problem animals that live in the forest (ie baboon, vervet monkey, lion, elephant) and eat crops. Apparently two people were killed in Ruhatwe in 2014 by elephants.
- Legal but unsustainable harvesting of timber leading to the loss of forest and severe soil erosion.

The elder interjects at this point and claims that this area in unlike other areas in Kilwa in that it is highly fertile. He states that you can cut a tree and in 5 years it will have regenerated [Whilst this is clearly an exaggeration, part of the village lands appear to be highly fertile (floodplain) where they grow rice during the rainy season when the lands become inundated with water. At present they are hard, cracked heavy clay soils (see image below)].

5 October 2014

The main focus of this day was the development of the storyboard, which was an arduous process. The group initially decided to focus on water and developed a storyboard over a period of approximately 2.5 hours. Upon its completion, the group was visibly distressed by the fact that they had not included trees – we suggested that to try and cover more in the allotted time would be difficult. Noticeably the three women in the group that had been largely silent throughout the development of the water storyboard, stated that had they had known they had limited time, they would've chosen to focus on the trees (forest). We told them they could certainly develop a new storyboard or try to incorporate their concerns in the present storyboard. After about 30 minutes into the development of a new tree-focussed storyboard, the group gave up and decided to continue with the creation of their first, water-focussed storyboard. [My impression is that this was a decision primarily driven by fatigue].

6 October 2014

A smaller group of us (4 participants) drove out to film the impact of the cattle and flooding (~ 5 km from the main village); this is also where the Sukuma are. On the way back I asked a young Sukuma girl if she had been born in the village. She replied that she had in fact been born in Morogoro. In asking further I learned that a few years ago the TZ government relocated a large number of Sukuma from Morogoro due to conflicts between their cattle and famers. Apparently Lindi and Mtwara Regions were designated for the re-homing of the Sukuma, but no infrastructure (ie water) was put in place, so naturally the Sukuma began to move elsewhere. At this time, village leaders in these regions were told [not sure by whom – perhaps the District?] that if the Sukuma came and asked for land, the village should not refuse them. The Ruhatwe participatory video group claims that although Kikole offered them land, Ruhatwe never agree [as a village] to do so.

I also asked how/if the history of the Sukuma, having been removed from their home, gave them any pause – whether they [Ruhatwe] felt any sympathy for them. The PV folks responded that being removed from your home is 'serious' and that they do in fact feel for them because of this. BUT, when the Sukuma came to Ruhatwe they moved to an area where there were already people living (from Ruhatwe) and 'disturbed their way of life.' So the way they see it is, although they feel for them, they also feel for their people and that this is an injustice. [I think this may be a potential recognition issue for us to explore].

Apparently the Sukuma began to come to Ruhatwe in 2012 [according to the PV group].

Responses of community members who attended rush screening

Prepared by Peter Paul (MCDI)

1. During the meeting with expert from the district (in this scene) people should not have interrupted each other, they should have given each other sometime to talk rather than interrupting.

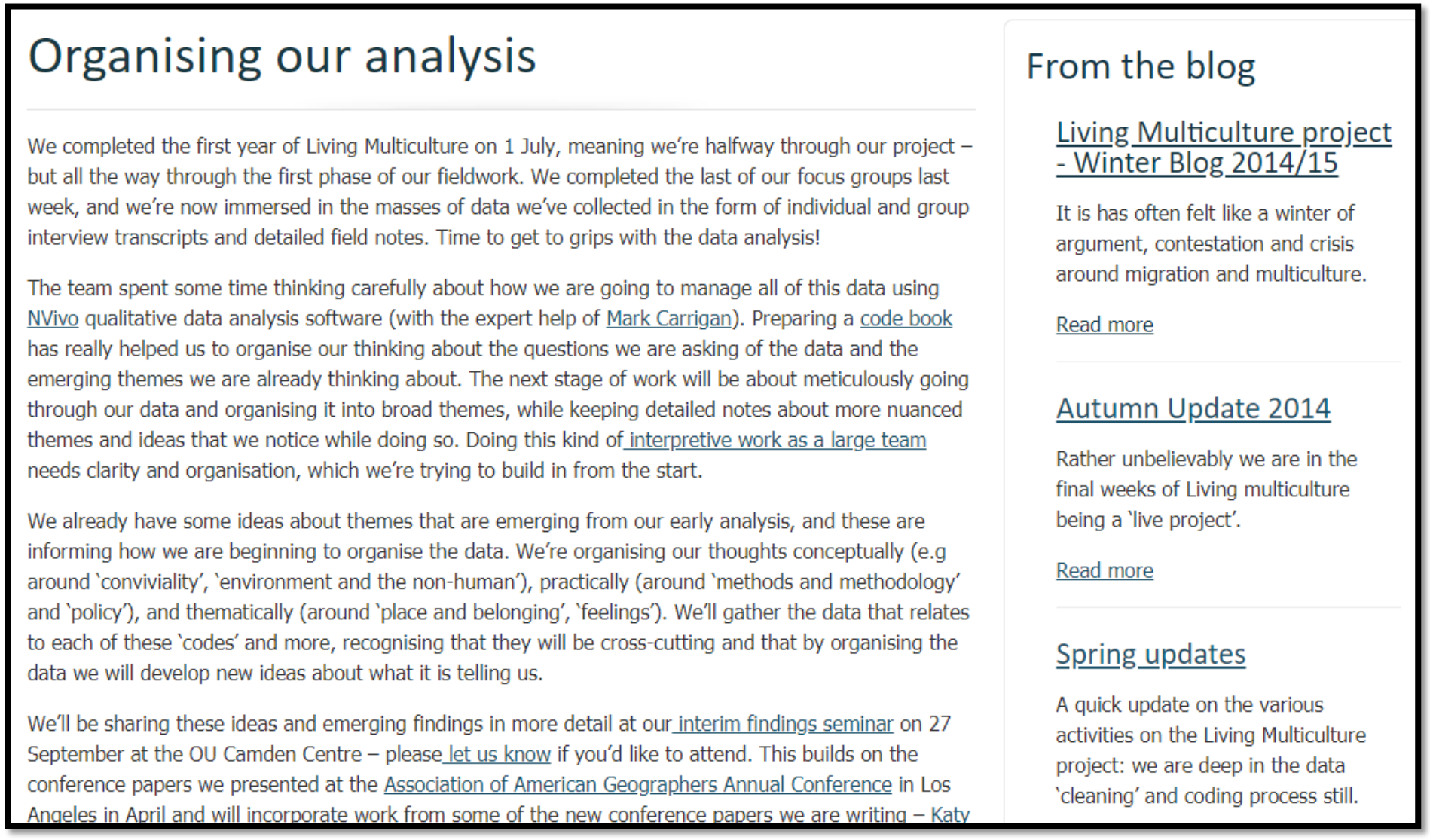
https://discover.ukdataservice.ac.uk/catalogue/?sn=852476&type=Data%20catalogue

New possibilities with changing technology

List of nodes and their description

Hierarchical Name	Description
Nodes\\Attitude	A person's general feeling of favorableness or unfavorableness. Synonym: affect (Ajzen and Fishbein) Evaluation of people, objects, and ideas (Aronson). A set of beliefs.
Nodes\\Attitude\Mixed	A person's statement that they are indecisive as to whether their attitude is positive or negative. If a person states two different views, such as that they are positive about one aspect and negative about another, this is not mixed, but rather multiple, feelings.
Nodes\\Attitude\Negative	A negative opinion or a disagreement where an opinion is not shared with someone or with a group.
Nodes\\Attitude\Positive	specifying positive opinion oro agreement with other stakeholders or with an idea. A shared opinion
Nodes\\Balance	
Nodes\\Community	Text coded to topics around the concept of community (not around specific named Down East communities)
Nodes\\Community\Community change	Change to the people, culture, or built structures of the communities Down East

Living Multicultural: blog



Organising our analysis

We completed the first year of Living Multiculture on 1 July, meaning we're halfway through our project – but all the way through the first phase of our fieldwork. We completed the last of our focus groups last week, and we're now immersed in the masses of data we've collected in the form of individual and group interview transcripts and detailed field notes. Time to get to grips with the data analysis!

The team spent some time thinking carefully about how we are going to manage all of this data using NVivo qualitative data analysis software (with the expert help of Mark Carrigan). Preparing a code book has really helped us to organise our thinking about the questions we are asking of the data and the emerging themes we are already thinking about. The next stage of work will be about meticulously going through our data and organising it into broad themes, while keeping detailed notes about more nuanced themes and ideas that we notice while doing so. Doing this kind of interpretive work as a large team needs clarity and organisation, which we're trying to build in from the start.

We already have some ideas about themes that are emerging from our early analysis, and these are informing how we are beginning to organise the data. We're organising our thoughts conceptually (e.g around 'conviviality', 'environment and the non-human'), practically (around 'methods and methodology' and 'policy'), and thematically (around 'place and belonging', 'feelings'). We'll gather the data that relates to each of these 'codes' and more, recognising that they will be cross-cutting and that by organising the data we will develop new ideas about what it is telling us.

We'll be sharing these ideas and emerging findings in more detail at our interim findings seminar on 27 September at the OU Camden Centre – please let us know if you'd like to attend. This builds on the conference papers we presented at the Association of American Geographers Annual Conference in Los Angeles in April and will incorporate work from some of the new conference papers we are writing – Katy

From the blog

Living Multiculture project - Winter Blog 2014/15

It is has often felt like a winter of argument, contestation and crisis around migration and multiculture.

Read more

Autumn Update 2014

Rather unbelievably we are in the final weeks of Living multiculture being a 'live project'.

Read more

Spring updates

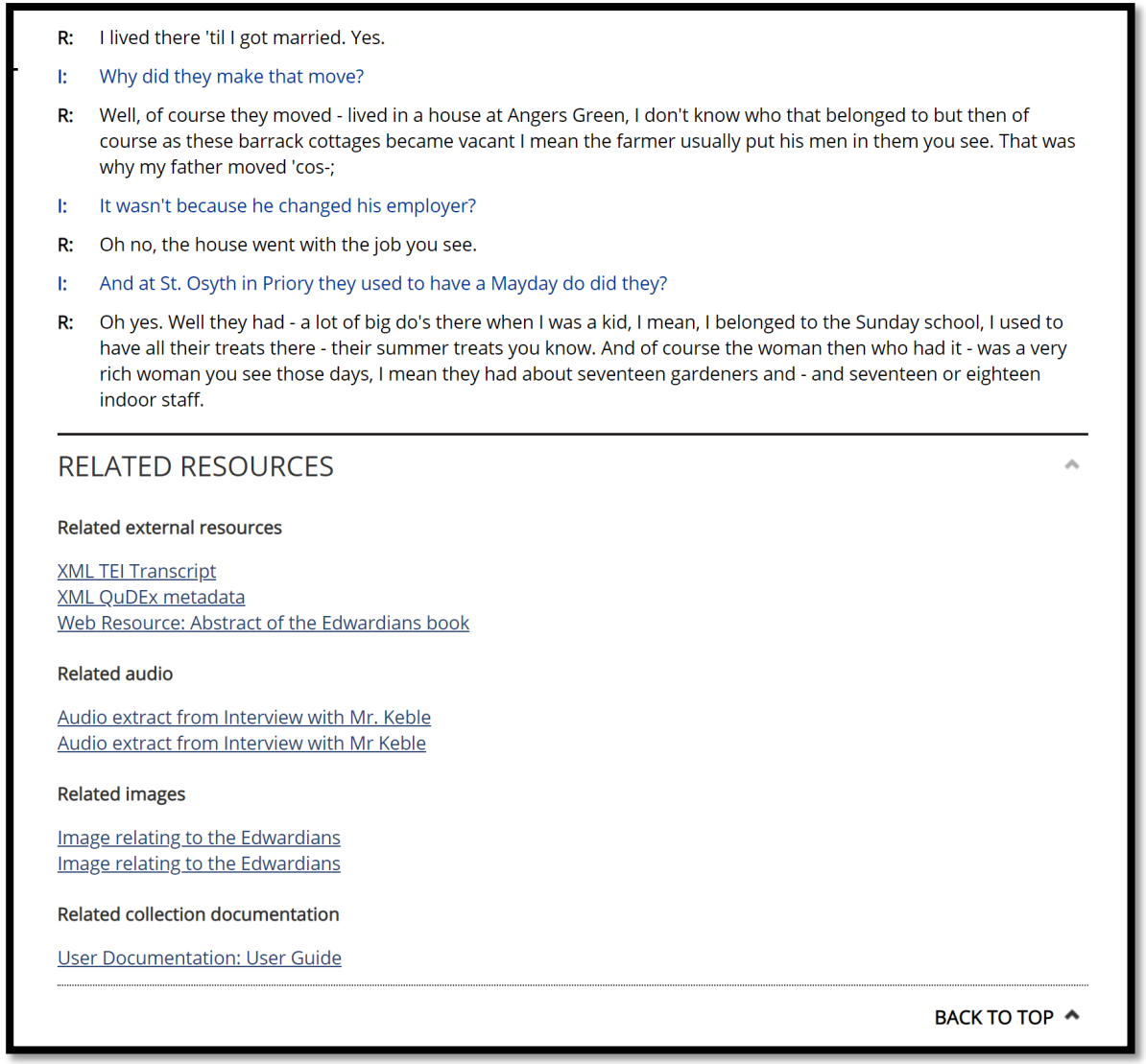
A quick update on the various activities on the Living Multiculture project: we are deep in the data 'cleaning' and coding process still.

https://discover.ukdataservice.ac.uk/catalogue/?sn=851852&type=Data%20catalogue

Creative documentation

https://discover.ukdataservice.ac.uk/catalogue/?sn=852476&type=Data%20catalogue

Linking specific documentation to specific data



R: I lived there 'til I got married. Yes.

I: Why did they make that move?

R: Well, of course they moved - lived in a house at Angers Green, I don't know who that belonged to but then of course as these barrack cottages became vacant I mean the farmer usually put his men in them you see. That was why my father moved 'cos-;

I: It wasn't because he changed his employer?

R: Oh no, the house went with the job you see.

I: And at St. Osyth in Priory they used to have a Mayday do did they?

R: Oh yes. Well they had - a lot of big do's there when I was a kid, I mean, I belonged to the Sunday school, I used to have all their treats there - their summer treats you know. And of course the woman then who had it - was a very rich woman you see those days, I mean they had about seventeen gardeners and - and seventeen or eighteen indoor staff.

RELATED RESOURCES

Related external resources

XML TEI Transcript
XML QuDEx metadata
Web Resource: Abstract of the Edwardians book

Related audio

Audio extract from Interview with Mr. Keble
Audio extract from Interview with Mr Keble

Related images

Image relating to the Edwardians
Image relating to the Edwardians

Related collection documentation

User Documentation: User Guide

BACK TO TOP

Re-using documentation

- 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.

How will the data be used?

I'm asking for permission to use anonymised quotations and narrative themes, along with any photographs and video you provide in the interviews or diaries for research purposes. All diaries and interviews from all participants will be analysed together for common themes about what everyday life is like when pregnant. As I work through my analysis, I will transcribe any audio recordings or handwritten diary entries. As I transcribe, I'll anonymise any identifying details, such as your name and address. All digital files will be saved on a password-protected computer at University of Essex and all paper documents will be stored in a locked drawer at my office at the University of Essex, to which only I have access.

Throughout the project, I will be the only one with access to un-anonymised data, and my supervisors will have access to anonymised data. Since this project has gone through ethical approval from the Health Research Authority, NHS Trust staff may also be audit this project to ensure I am protecting your information appropriately, and may ask to see relevant sections of data.

Researching with children and vulnerable groups

My story

(draw or stick a picture of yourself here)

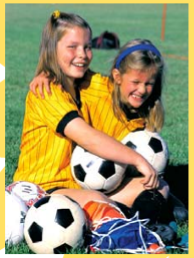
My name is.....

.....

I'm .....years old

© Ruth Evans 2006

Who makes you laugh?



Our project

We are working on a project about who is important in children's lives.

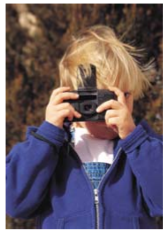
We'd like to know about *who matters to you.* It might be your family, carers, grandparents, friends, teacher, cousins or your pets! Why are they important? What sort of things do you do together?

We're asking lots of children what they think so we can write a report about who is important from children's point of view.

What do YOU think?

We'd like to talk to you and hear what YOU have to say. If you want to take part, we will arrange an **interview** with you- it's a chat where we ask you some questions and you can tell us what you think.

Before the interview we'll give you a camera. You can take photos of anything that makes you think about important people.



We'll develop the photos for you. Then in the interview we'll talk about the photos you took.

You can also do some drawing or writing if you want. It's not all talking!

Who will be there?

One of us will come and do the interview at your house, or somewhere else if you prefer. You can have the interview on your own or ask a friend or relative to be there with you.

Is what I say private?

Yes. When we write about what children tell us, we'll change everyone's name so no-one will know what you said.

How do you keep in touch with important people?

Who's always there for you?

Developing an interview guide

UK Data Archive SN 6226 - Pioneers of Social Research, 1996-2018

Pioneers of Social Research: Interview Guide

Pre-interview
Look for an outline biography of cv of the person you are recording (Wikipedia, Who's Who, or they may send you a cv)

Try to give a hard look at what seem to be the researcher's key books

For every interview it is essential to establish in the early part
a) the date and place of birth of the interviewee, their family background and occupations;
b) what led them towards research and university work;
c) what were they key influences in leading them towards their pioneering research themes.
d)

Opening questions

Date of birth, family and own occupations.

Can we begin by talking about your family background and early life, and any early influences which may have led you towards university work?

FAMILY AND EARLY LIFE

- Memories of ***family background*** and ***grandparents:*** were they an influence?
- ***Parents*** - where they came from, their jobs, their characters, could you talk to them, were you close to them;
- ***Education -*** school and ***University***: important friendships and influences
- ***Youth*** - important friends and their influences; what you did in ***National Service***

UK Data Service

Documentation for Quantitative Data

- What documentation accompany quantitative data?
- Embedded documentation
- What good documentation looks like, with examples

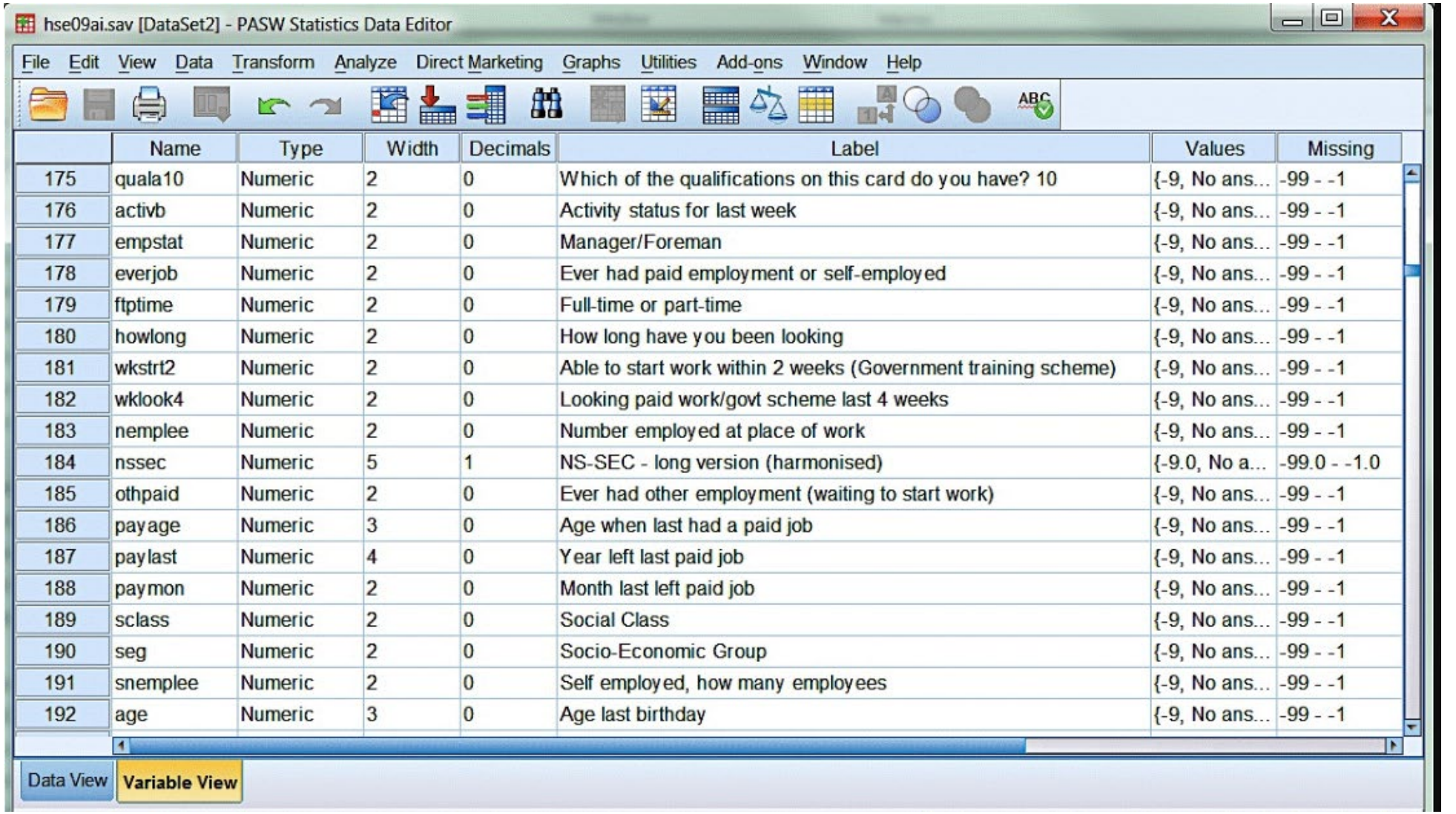
Documentation for quantitative data - Summary

- Questionnaire
- Codebook (embedded or separate)
- Data dictionary
- User Guide
- Experiment protocol
- ReadMe file

Documentation for quantitative data

- For quantitative data (structured, tabular data), documentation can be embedded within data files, such as variable and code descriptions in databases.
- Most data analysis software packages have facilities for data annotation and description, as variable attributes, data type definitions, table relationships and so on.
- Alternatively, information about data items can be recorded in a structured document such as a codebook or a data dictionary.
- *What would someone with no prior knowledge of the project/data need to be able to understand or use the data correctly in their own research*?

Embedded documentation in SPSS file



	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

User Guides

A user guide should contain a variety of documents that provide context information, such as: data collection methods, fieldwork, consent procedures, interview schedule, experiment protocol, transcription notes, study findings, weights applied to the data etc.

University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: COVID-19 Study, 2020-2021. [data collection]. 11th Edition. UK Data Service. SN: 8644, DOI: http://doi.org/10.5255/UKDA-SN-8644-11

Durrheim, K., Dixon, J. (2009). Intergroup Contact and the Construction of Racial Inequality and Injustice in Post-Apartheid South Africa, 2006-2007. [data collection]. UK Data Service. SN: 6315, DOI: http://doi.org/10.5255/UKDA-SN-6315-1

ScotCen Social Research. (2020). Scottish Social Attitudes Survey, 2016. [data collection]. UK Data Service. SN: 8628, DOI: http://doi.org/10.5255/UKDA-SN-8628-1

Codebooks, Data Dictionaries

- In previous slides we saw how we can have documentation embedded in data files*. Alternatively, information about data items can be recorded in a structured document such as a codebook or a data dictionary.
- Should contain detailed and sufficient information about all data items – all variables new and derived, (ideally including frequencies as well as code, algorithm or command files used to create them).
- Codebook creation tools:
  - DDIEditor - is aimed at data processing for curation purposes (accommodating the requirements for a data archive).
  - Nesstar Publisher - advanced data management program consisting of data and metadata conversion and editing tools, enabling the user to prepare these materials for publication.

*Possibly under access restrictions.

Codebooks, Data dictionaries

Examples:

- Codebook:
  - University of Essex, Institute for Social and Economic Research, University of Manchester, Cathie Marsh Centre for Census and Survey Research, UK Data Service. (2014). Understanding Society, Wave 3, 2011-2012: Teaching Dataset. [data collection]. University of Essex, Institute for Social and Economic Research, [original data producer(s)]. University of Essex, Institute for Social and Economic Research. SN: 7549, DOI: http://doi.org/10.5255/UKDA-SN-7549-1

- Data dictionary:
  - Department for Education. (2019). Further Education Learners and Apprentices Survey, 2018: Secure Access. [data collection]. UK Data Service. SN: 8492, DOI: http://doi.org/10.5255/UKDA-SN-8492-1
  - Baumberg Geiger, B., Edmiston, D., Summers, K., de Vries, R., Robertshaw, D., Young, D., Gibbons, A., Scullion, L. (2023). Welfare at a (Social) Distance, 2020-2022. [data collection]. 3rd Edition. UK Data Service. SN: 8689, DOI: http://doi.org/10.5255/UKDA-SN-8689-3

Data-level documentation

- All structured, tabular data should have adequate variable names as well and variable and value labels.
- Variable names
  - Question number system matching questions in the questionnaire used e.g Q1a, Q1b, Q2, Q3a, Q3b, Q3c etc.
  - Numerical order system e.g V1, V2, V3 etc.
  - Meaningful abbreviations or combinations of abbreviations referring to meaning of the variable e.g. 'GOR' for Government Office Region, 'moocc' for mother's occupation.
  - Consistent in naming conventions across the entire project.
  - For interoperability across platforms, variable names should not be longer than 8 characters and without spaces.

Data-level documentation

Similar principles for variable labels:

- Brief and concise, maximum 80 characters
- Include unit of measurement where appropriate
- Coding or classification schemes used, with a bibliographic reference e.g., Standard Occupational Classification 2000, ISO 3166 alpha-2 country codes
- Reference the question number of a survey or questionnaire

e.g., variable '**q11hexw**' with label Q9b:hours spent taking physical exercise in a typical week' - label includes unit of measurement (hours) and references question number in the questionnaire (Q9b).

Data-level documentation

For value labels:

- Ensure no out of bounds values for categorical variables.
- Avoid having blanks, system missing or 0.
- Ensure to code and include reasons for missing values.

  e.g '99 – not recorded', '98 – not provided' (no answer), 97 – not applicable/skipped', '96 – not known', '95 – error'.

Variable information

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format	Missing Values
Sserial	1	Archive Serial	Scale	Input	10	Right	F6	F6	Lowest through -1
WtFactor	2	Final BSA weight	Scale	Input	11	Right	F9.4	F9.4	Lowest through -1.0000
RSex	3	Sex of respondent	Nominal	Input	8	Right	F2	F2	-1, 8, 9
RAgeCat	4	Age of respondent (grouped)<7 category> dv	Ordinal	Input	8	Right	F2	F2	-1, -2, 8

Variable values

Variable Values

	Value	Label
Sserial	-9[a]	Refused
	-8[a]	Don't know
	-2[a]	Schedule not applicable
	-1[a]	Not applicable
WtFactor	-2.0000[a]	Skip, version off route
	-1.0000[a]	Item not applicable
RSex	-1[a]	Item not applicable
	1	Male
	2	Female
	8[a]	Don`t know
	9[a]	Refusal

Data-level documentation

Pos. = 3 **Variable =** NEW_CHECK2 **Variable label =**
Attended any learning or an apprenticeship in a college or Further Education setting in the last year
This variable is *numeric*, the SPSS measurement level is *NOMINAL*

Value label information for NEW_CHECK2

Value = 1.0	Label = Yes
Value = 2.0	Label = No

Pos. = 95 **Variable =** AEM_BENEFITS **Variable label =** Whether learner/apprentice was claiming benefits or tax credits before starting course/apprenticeship
This variable is *numeric*, the SPSS measurement level is *NOMINAL*
SPSS user missing values = -1.0 thru -2.0

Value label information for AEM_BENEFITS

Value = 1.0	Label = Yes
Value = 2.0	Label = No
Value = -2.0	Label = Prefer not to say
Value = -1.0	Label = Dont know

UK Data Service

Metadata

- What is it?
- What qualifies as good metadata?
- Metadata standards
- How is metadata produced?

What is metadata?

- Essential subset of core data documentation providing standardised, structured information.
- Intended for machine-reading.
- Important for the purposes of cataloguing, citing, discovering and retrieving data collections.
- Without good quality metadata, data collections become of limited value because potential users will not be able to search for these collections in data catalogues or cite the data collection.

Metadata - Examples

- Abstract
- Keywords
- Topics
- Dates of fieldwork
- Country
- Spatial units
- Observation units
- Population
- Number of units
- Method of data collection
- Kind of data (numeric, text, image etc.)
- Weighting

Good metadata- FAIR Principles

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 right)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, they need to know how they can be accessed, *possibly including authentication and authorisation*.

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

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage and processing.

- 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

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- 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

Producing good metadata

- Archives or repositories will collect information using a data deposit form, which is completed by the data depositor. They will then enhance this information to create a conformant metadata record.
- Metadata standards are used by data archives to enhancing discoverability, interoperability and reusability. When you submit your dataset to a trusted data repository, these standards are automatically applied.

- The UK Data Service uses DDI to structure catalogue records. DDI stands for Data Documentation Initiative

Producing good metadata

- At UKDS, we use the Data Documentation Initiative (DDI), which is a rich and detailed metadata standard originally designed for describing social, behavioral and economic sciences data. It is 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.
- There are other metadata standards, varies (usually) by discipline.

Metadata – Controlled Vocabularies

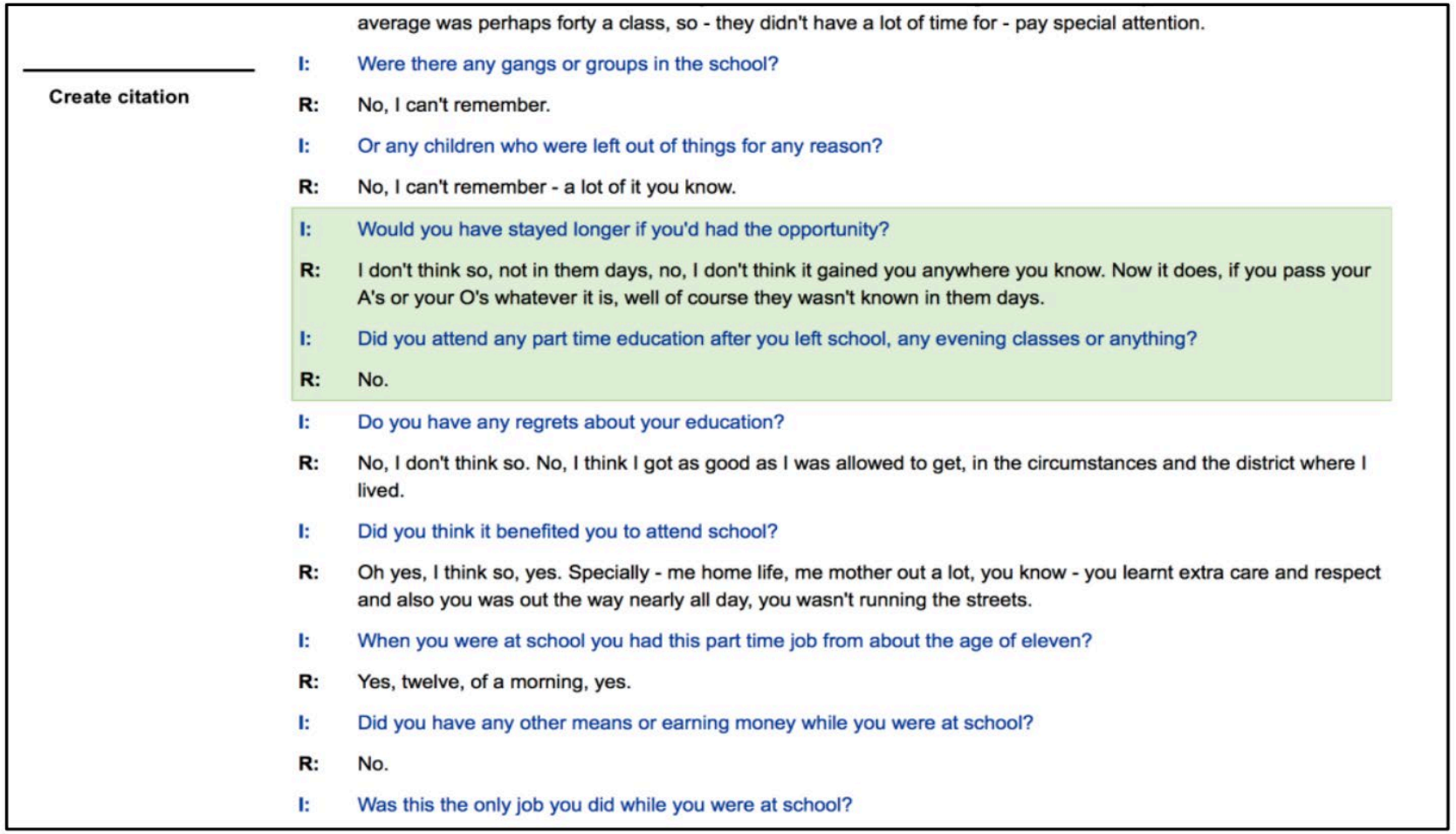
- A controlled vocabulary is a consistent and organised way of describing data and is essential to make data findable and sharable within research communities.
- e.g., subject headings, thesauri, ontologies, and taxonomies

- The UK Data Archive is guardian of the Humanities and Social Science Electronic Thesaurus (HASSET) and curator of its multilingual sister, the European Language Social Science Thesaurus (ELSST).

Linking to persistent data

QualiBank: persistent citation of sources

https://discover.ukdataservice.ac.uk/QualiBank/Document/?cid=q-1dba72b1-d148-40e7



average was perhaps forty a class, so - they didn't have a lot of time for - pay special attention.

Create citation

I: Were there any gangs or groups in the school?

R: No, I can't remember.

I: Or any children who were left out of things for any reason?

R: No, I can't remember - a lot of it you know.

I: Would you have stayed longer if you'd had the opportunity?

R: I don't think so, not in them days, no, I don't think it gained you anywhere you know. Now it does, if you pass your A's or your O's whatever it is, well of course they wasn't known in them days.

I: Did you attend any part time education after you left school, any evening classes or anything?

R: No.

I: Do you have any regrets about your education?

R: No, I don't think so. No, I think I got as good as I was allowed to get, in the circumstances and the district where I lived.

I: Did you think it benefited you to attend school?

R: Oh yes, I think so, yes. Specially - me home life, me mother out a lot, you know - you learnt extra care and respect and also you was out the way nearly all day, you wasn't running the streets.

I: When you were at school you had this part time job from about the age of eleven?

R: Yes, twelve, of a morning, yes.

I: Did you have any other means or earning money while you were at school?

R: No.

I: Was this the only job you did while you were at school?

Reviewing a dataset

1. Generic project-level checks:

- project-level metadata and consent forms are checked for legal and ethical information specified by the depositor, that influences data sharing and reuse (e.g. anonymisation carried out; consent in place for data sharing; consent for disclosing participant names, ….)
- for qualitative data (interviews, images, video) consent agreements should not preclude data sharing
- copyright status and permissions are checked, e.g. for research where third party data have been used
- for data collections resulting from a RCUK-funded project, that a link to the project record on the Gateway to Research is included in 'Related resources'

2. Generic file-level checks:

- each file opens without problems
- file formats of data files and documentation files conform to the UK Data Archive's recommended file formats: http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx
- file properties of data files contain no names of people or other disclosive information (typically the case where interviewee name has been used as file name before anonymisation)
- access and licence have been selected in accordance with any confidentiality concerns (e.g. safeguarded data, permission access data) and documentation files are open access

http://reshare.ukdataservice.ac.uk/reshare-review-procedures/

Data Sharing Checklist – Documentation needed

- ✓Check archive/repository guidelines and create necessary documentation files depending on what data you are sharing.

  There might be templates already available which you can just download and adapt to your project (data list, ReadMe).

- ✓Fill in data offer and/or data deposit form in as much detail as possible - this will allow the archive/repository to create machine readable metadata.

- ✓Ensure your data files also contain data-level documentation.

Accessing Data

Access Options

OPEN

Available for download/online access under open licence without any registration

•SAFEGUARDED

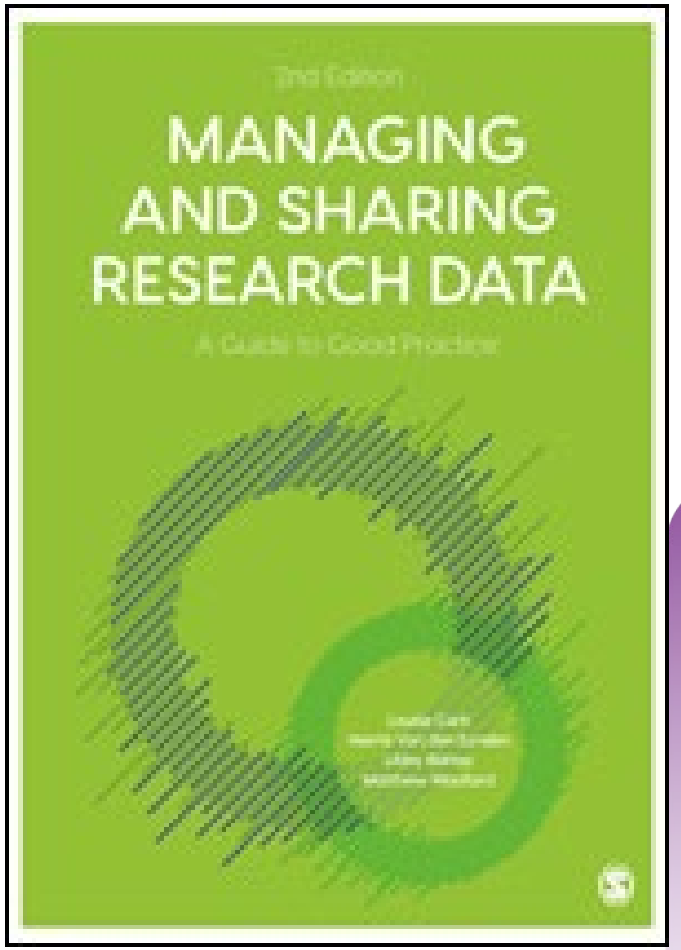
•Available for download / online access to logged-in users who have registered and agreed to an End User Licence; special agreements (e.g. depositor permission or approved researcher); embargo for fixed time period

•CONTROLLED

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

UKDS data management guidance

- Best practice guidance: www.ukdataservice.ac.uk/manage-data.aspx
- *Managing and Sharing Research Data: A Guide to Good Practice* (Sage Publications Ltd)
- Training: https://ukdataservice.ac.uk/training-events/
- Twitter: @UKDSRDM



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
- 3 interviews and 5 diaries were embargoed until 2015
- audio files archived and only available by permission from researchers

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

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

Tools 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

Further resources

- Anonymising Research Data - ESRC National Centre for Research Methods, Working Paper 7/06
- Guide to Social Science Preparation and Archiving from the Inter-University 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
- Advice from med.data.edu on anonymisation

Get connected

http://ukdataservice.ac.uk/about-us/contact.aspx

https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKdataservice

https://twitter.com/UKDataService
@UKDataService

https://www.facebook.com/UKDataService

https://www.youtube.com/user/UKDATASERVICE

***Powerpoint slides will be available on our website in due course and you can catch up on the recording on our Youtube channel.
Check out our Twitter for more updates.***

UK Data Service

Thank you.

Maureen Haaker
mahaak@essex.ac.uk

Anca Vlad
advlad@essex.ac.uk