MANAGING QUALITATIVE DATA

LOUISE CORTI
ASSOCIATE DIRECTOR
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

DATUM WORSHOP, NEWCASTLE
26 May 2011
PRACTICAL ADVICE FOR QUALI RESEARCHERS

• too much DM advice out there is generic and confusing

• data management is not a word quali researchers understand/like much – we use ‘manage and share’

• guidance needs to be meaningful and relevant

• qualitative data has its own peculiarities – I hope to offer you useful take home messages and tips
WHERE WE ARE COMING FROM

• UK Data Archive experience of providing support for data creators of social science data

• We operate the ESRC Data Policy (since 1995) – you must share your data!

• our best practice approaches to making data shareable based on:
  • challenges faced by data creators over the years
  • over 40 years of handling others’ social science data - 17 years with qualitative data
• highly skilled staff comprising researchers, technical and information specialists
<table>
<thead>
<tr>
<th>Doc. 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPORTANT</td>
<td>Thesis Final final</td>
</tr>
<tr>
<td>My study</td>
<td>My study</td>
</tr>
<tr>
<td>Interview with Jane</td>
<td>Interview with Janet</td>
</tr>
<tr>
<td>Int. 1</td>
<td>Int. 1</td>
</tr>
<tr>
<td>My data</td>
<td>My data</td>
</tr>
<tr>
<td>My Passwords</td>
<td>My Passwords</td>
</tr>
<tr>
<td>Thesis version 12</td>
<td>Thesis version 12</td>
</tr>
<tr>
<td>Data chart for interviews</td>
<td>Data chart for interviews</td>
</tr>
<tr>
<td>Int 1 (2)</td>
<td>Int 1 (2)</td>
</tr>
<tr>
<td>My thesis (copy)</td>
<td>My thesis (copy)</td>
</tr>
</tbody>
</table>

UK DATA ARCHIVE

JOURNEY FROM YOUR DATA ....
...TO OUR DATA

Study Number: 4410
Interview Reference No. 931

Female, White British, D.O.B - 14.03.64, Married, Employed, CBS Score Time 1 - 22

NOTE: Machine noise throughout interview.

Q: So I know Clare’s been asking you about things about the area, particularly things like changes that have happened over the last twelve months. Could you just sort of give us what you think’s changed in the area over the last twelve months.

A: Erm … well, (Willow Park’s) took over a big project here, you know, they bought a lot of property off the council and they seem to be trying to build the area back up. You know, trying to make it look nicer for … doing a lot with houses, you know, that they took over. Like double glazing, shower cubicles, new bathrooms, fitted kitchen and …

(Brief interruption - laughter)

A: … generally trying to make the properties look better as well.

Q: And is your house anything to do with Willow Park?

A: Yes. Yes.

Q: And have you done anything to your house?

A: They’ve recently done the bathroom. I’ve had a new shower cubicle put in instead of a bath. I’ve been rewarded, I’ve not had a fitted kitchen because I’ve got my own, and generally I find the work, if you’ve got a problem they seem to address it quite quickly.

Q: So has it made a big difference to you personally?

A: Yeah. Yeah.
BENEFITS OF GOOD DATA MANAGEMENT

• efficiency – makes research easier
• quality – better research data – for you and others
• safety – protect valuable data
• reputation – enhances research visibility
• compliance – with ethical codes, data protection laws, journal requirements, funder policies
WHAT IS DATA MANAGEMENT?
KEY DATA MANAGEMENT AREAS FOR QUALIS

- data confidentiality and conditions of use
- describing and documenting data for re-use
- practicalities of looking after data
  - formats, version controlling, encryption, storage, back-up, file-sharing
ETHICS AND DATA SHARING

Ethical duties in research

- confidentiality towards informants and participants
- protect participants from harm
- treat participants as intelligent beings, able to make their own judgements and decisions on how the information they provide can be used, shared and made public (through informed consent)
- duty to wider society to make available resources produced by researchers with public funds

Consider data management and sharing during ethical review
LEGISLATION AND DATA SHARING

Data Protection Act (1998)

- ‘personal data’
  - relate to living individual
  - individual can be identified from those data or from those data and other information
  - includes any expression of opinion about the individual

- only disclose personal data if consent given to do so (exc. legal reasons)
- DPA does not apply to anonymised data

processed fairly and lawfully obtained and processed for specified purpose adequate, relevant and not excessive for purpose accurate not kept longer than necessary processed in accordance with the rights of data subjects, e.g. right to be informed about how data will be used, stored, processed, transferred, destroyed; right to access info and data held kept secure not transferred abroad without adequate protection
**PRINCIPLES FOR ETHICAL / LEGAL DATA SHARING**

Researchers to consider

- obtaining *informed consent*, also for data sharing and preservation / curation
- protecting identities
  - e.g. anonymisation, not collecting personal data
- restricting / regulating access where needed (all or part of data)
  - e.g. by group, use, time period
- securely storing personal or sensitive data

Consider jointly and in dialogue with participants
Plan early in research
INFORMED CONSENT

Information sheet and consent form must include consent for

- engaging in the research process, and right to withdraw
- use of data in outputs, publications
- data sharing and possible future uses

Process or one-off consent? - repeat interactions?

Written or verbal consent? - how realistic?

Consent needs to be suitable for the research purposes

UK Data Archive sample consent form
http://staging.data-archive.ac.uk/create-manage/consent-ethics/consent?index=3
ANONYMISING DATA

Identity disclosure
• direct identifiers – often not essential research info
• indirect identifiers

Anonymise data
• remove direct identifiers
• reduce precision/detail through aggregation / generalisation
• restrict upper lower ranges variables to hide outliers
• replace rather than remove
• pseudonyms
• maintain maximum meaningful info
• log edits
DATA ACCESS CONTROLS

at the UK Data Archive

- archived research data NOT in public domain
- use of data for specific purposes only after user registration
- data users sign legally binding End User Licence
  
  *e.g. not identify any potentially identifiable individuals*
- stricter access regulations for sensitive data (case to case basis):
  - access to approved researchers only (approved researcher/special license)
  - data access permission from data owner prior to data release
  - data under embargo for given period of time
If someone was using your data for the first time, what would they need to know?

• context information about research and data
  • final report, publications, fieldnotes, thumbnail about interview setting

• data collection methodology and processes: sampling, data collection process, instruments used, tools used, temporal/geographic coverage, data validation

• documentation: descriptions of codes or classifications used

• data listings for qualitative data

• any conditions of use and access?
CAN YOU UNDERSTAND/USE THESE DATA?

SrvMthdDraft.doc
SrvMthdFinal.doc
SrvMthdLastOne.doc
SrvMthdRealVersion.doc
DATA QUALITY CONTROL IN THE RESEARCH PROCESS

• data collection – data must reflect facts, responses, observations, events
  • examples: standardised protocols, computer assisted interviews, interview recording and transcription

• data entry, digitisation, transcription and coding – avoid errors - use standardised and consistent procedures
  • examples: validation rules for data entry, controlled vocabularies or choice lists, transcription template

• data checking and verifying - automated and/or manual
  • typos, check for data completeness, peer review of data
DATA FORMATS

- choice of software format for digital data
  - planned data analyses/discipline-specific customs
  - software availability
  - hardware used – e.g. audio recorders
  - discipline-specific standards and customs

- best formats for long-term preservation
  - standard formats
  - interchangeable formats
  - open formats
    - tab-delimited, comma-delimited (CSV), ASCII, RTF, PDF/A, OpenDocument format, XML

- beware of errors in data conversion! Always check
FIELDWORK FROM HELL

“I’m sorry but we had to blow up your laptop”

“What….all my client case notes and testimony, writing, pictures, music and applications. Years of work. NO!!!! What?? Are you insane?? What were you thinking? THAT’S ALL MY WORK!?”

Source: Lilysussman’s Blog
DATA STORAGE

• ALL digital storage media are fallible

• file formats and physical storage media ultimately become obsolete
  • optical (CD, DVD) and magnetic media (hard drive, tapes) degrade

• best practice:
  • use data formats with long-term readability
  • storage strategy - at least two different forms of storage and locations; maintain original copy and external copies
  • check data integrity of stored data files regularly (checksum)
  • know your personal / institutional back-up strategy: network server/PC/laptop; data retention policies
  • what to protect? Not only data, and not only digital
DATA SECURITY

- protect data from unauthorised access, use, change, disclosure and destruction
- personal data need more protection – always keep separate
- control access to computers
  - passwords
  - anti-virus and firewall protection, power surge protection
  - networked vs non-networked PCs
  - all devices: desktops, laptops, memory sticks, mobile devices
  - all locations: work, home, travel
  - restrict access to sensitive materials e.g. consent forms, patient records
- proper disposal of equipment (and data)
  - even reformatting the hard drive is not sufficient
- control physical access to buildings, rooms, cabinets
- but beware of “requirements” to destroy data
ENCRYPTION

- always encrypt personal or sensitive data

- easy to use
  - Safehouse
  - Truecrypt
  - Axcrypt

- basic principles
  - use an algorithm to transform information (A=1)
  - need a “key” to decrypt

- encrypt anything you would not send on a postcard
  - for moving files e.g. transcripts
  - for storing files e.g. shared areas, mobile devices
FILE TRANSMISSION

Sharing data between researchers and teams

• virtual research environments
  • MS Sharepoint
  • Sakai

• file transfer protocol (ftp)

• Yousendit, Dropbox

• via physical media

• too often email attachments
KEY DATA MANAGEMENT INTERVENTION POINTS

The Social Science Research Life Cycle

I. Initial Data Collection
   - Shared data sharing protocols

II. Data Collection
   - Licensing, terms and conditions for sharing, formal documentation

III. Final Data Preparation & Analysis

IV. Publication & Sharing
   - Licensing, terms and conditions for sharing, formal documentation

V. Long-term Management
   - Data formats, data migration

I. Discovery & Planning
   - Sign off consent form
   - Agree data & metadata templates

Green and Gutmann, 2007
BASICS OF WHAT TO PUT IN A DATA MANAGEMENT PLAN

• need for access to existing data sources
• data planned to be produced
• planned quality assurance and back-up procedures for data
• plans for management and archiving of collected data
• expected difficulties in making data available for re-use and measures to overcome such difficulties
• who holds copyright and intellectual property rights of data
• data management roles and responsibilities
WHAT RESEARCH DATA ARE WORTH KEEPING?

- rich data, breadth, unique, topical, time series
- format, usability and condition of material
- data that have further analytic potential than the original investigation (depth; large-scale; longitudinal)
- relative importance or impact of the study
- confidentiality issues unproblematic (consent)
- copyright is not prohibitive
HOW DO PEOPLE RE-USE DATA?

- descriptive material
- comparative research, restudy or follow-up study
- re-analysis/secondary analysis
- research design and methodological advancement
- replication of published statistics
- teaching and learning
IS IT REALLY DIFFICULT

• no, it’s just common sense.

• it’s all about good research practice

• be organised and accountable, and anticipating disaster at any time!

SIMPLES!
RESEARCH DATA MANAGEMENT SUPPORT SERVICES
UK DATA ARCHIVE
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX CO4 3SQ

T  +44 (0)1206 872001
E  datasharing@data-archive.ac.uk
www.data-archive.ac.uk/sharing