ARCHIVING YOUR DATA: PLANNING AND MANAGING THE PROCESS

LIBBY
BISHOP

RESEARCHER LIAISON
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

TCRU/NOVELLA SPECIAL SEMINAR - LONDON

29 MAY 2012

UK•DATA ARCHIVE

UK · DATA ARCHIVE

THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES

HOME   ABOUT US   CREATE & MANAGE DATA   DEPOSIT DATA   HOW WE CURATE DATA   FIND DATA   NEWS & EVENTS

SEARCH OUR SITE   GO

AUTHENTICITY AND PROVENANCE

What are they? How do social scientists understand them? How are they used in the digital preservation community?

READ MORE

What's new | Data lifecycle | Find data | Who are we?

FIRST TIME HERE?
HELPFUL INFORMATION

A QUICK GUIDE TO THE ARCHIVE
6 of 8: We provide best practice guidance and training in managing and sharing research data

WHO GIVES US DATA?
Find out what kind of data are available to you

DEPOSITING YOUR DATA
Depositing your data with the Archive ensures that they will be professionally curated and accessible

DEPOSIT DATA

FINDING DATA TO USE
We can help you find data for research and teaching with our catalogue of over 5,000 data collections

OUR CATALOGUE

OUR DATA IN USE
Labour Force Survey: How does the health service distribute resources equitably...

OUR DATA IN USE

OUR SERVICES
Census.ac.uk
providing a gateway to 1971-2001 UK censuses

CENSUS.AC.UK

VIEW SITE

CONSENT & ETHICS
Find out how to gain informed consent for sharing, protect identities and securely store research data

READ ON

PREPARING YOUR DATA
Prepare your data from the start of research so that they can be shared and re-used in the future

READ ON

LATEST NEWS & EVENTS
MORE

Nesstar 4.0 released
ESDS has recently upgraded Nesstar, its online ...

DwB: call for research proposals
The Secure Data Service is a lead partner in the ...

Survey on public sector data
Are you a (potential) user of public sector data? ...

Archive moves forward with DDI
The world of digital data is changing rapidly, and ...

WHY RE-USE DATA?

http://www.livingandworkingonsheppey.co.uk/

Living and Working on Sheppey

**Welcome to the Living and Working on Sheppey: Past, Present and Future project website.**

Living and Working on Sheppey explores the recent history and changes in working lives on Sheppey in the last decades of the 20th century and into the 21st. It concentrates on the period since the closure of the dockyard at Sheerness in 1960.

The site contains **materials from older people and their memories of work** in or around the dockyard and everyday life in Blue Town, the area of Sheerness closed to the dockyard. You can listen to clips of these interviews on different themes including the closure of the Dockyard, various trades, and changes in family life across different generations, and you can read the interviews in full (you need to register to do that).

You can read and listen to **short essays written by young people** in 2009-10 on what they imagine their futures have in store for them. They talk about their hopes and aspirations for their working lives, families, and travel as well as some of the difficulties they foresee.

You can watch the **two videos made by the artists, Tea,** as part of the Living and Working on Sheppey project. In 'Back and Forth on High Street Blue Town' you can see images of Blue Town High Street past, present – and future! And you can hear older people's memories of everyday life on the High Street through the twentieth century. In 'Sheerness Port', you can go on a journey around the site of the former Dockyard today.

The Living and Working on Sheppey project was funded by the South East Coastal Communities Programme from 2009 to 2011. See Peter Hatton and Jenny Hurkett and discuss the project, the Blue Town Heritage Centre and the Isle of Sheppey in the video clip below.
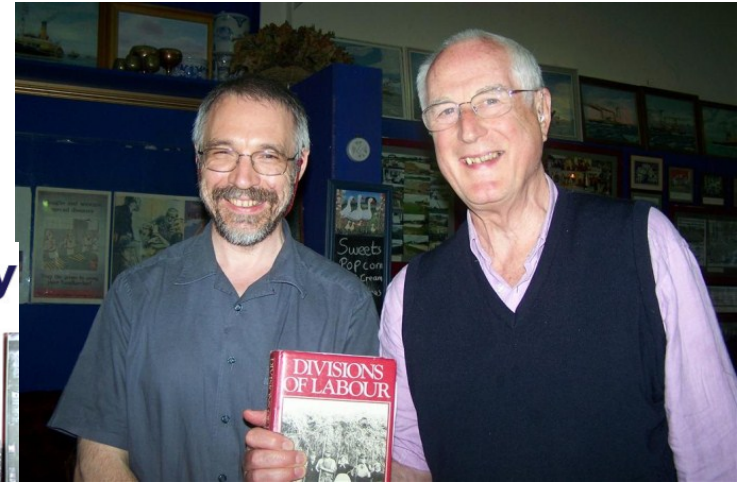
Home
Project overview
Project team
The Isle of Sheppey: context
The Isle of Sheppey: gallery
Older people's memories
Young people and the future
Blue Town High Street: a video by Tea
Sheerness Port: a video by Tea
Lessons learned
Get involved
1978 Essay Writers Facebook Page
Access to data
Presentations and events
Publications and publicity
Ray Pahl's Sheppey Studies
Links
Contact us

Scholarly collaboration across generations

BENEFITS OF GOOD DATA MANAGEMENT

- efficiency – makes research easier

- safety – protect valuable data

- quality – better research data

- reputation – enhances research visibility

- compliance – with ethical codes, data protection laws, journal requirements, funder policies

# DATA MANAGEMENT CHALLENGES

- data confidentiality and how to archive and share such data:
    - consent for data sharing
    - anonymising data
    - access regulation to data

- describing and documenting data for re-use

- data copyright

- practicalities of looking after data
    - formats, version controlling, encryption, storage, back-up, file-sharing

AND THE PRACTICAL CHALLENGES ARE

- having technical knowledge, support and tools
  - software, formats, storage

- understanding legal implications from a researcher's point of view
  - data protection, duty of confidentiality, copyright

- never enough staff or time!
  - sharing or archiving data will be low on the priority list
  - contextualising raw data can be very time consuming

KEY DATA MANAGEMENT INTERVENTION POINTS

Green and Gutmann, 2007

UK **DATA ARCHIVE**
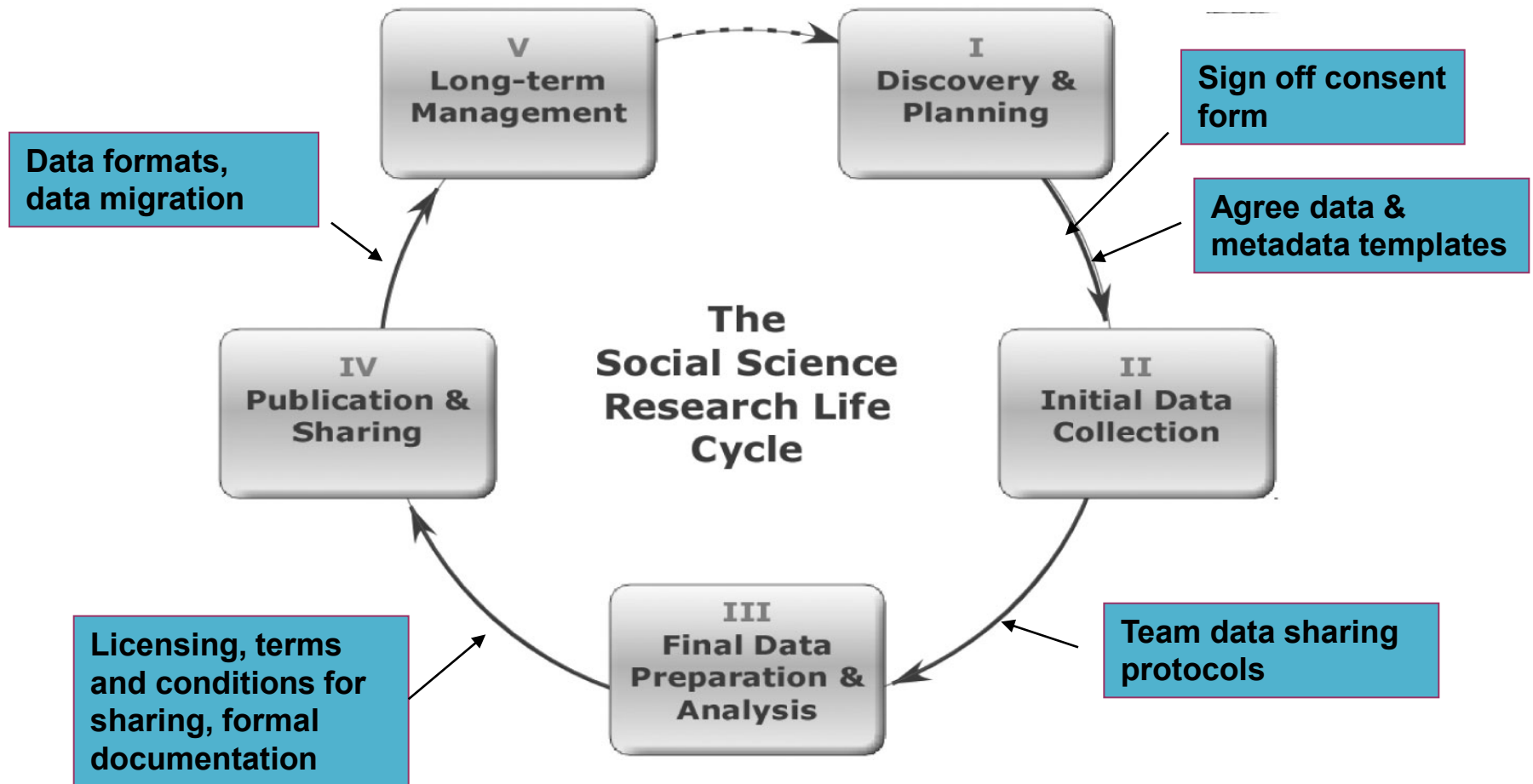
ETHICS AND DATA SHARING

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
- research data (even personal) exempt if: no indiv ID or effect; research only; no harm

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

UK **DATA ARCHIVE**

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 (future uses)

Process or one-off consent? - repeat interactions?

Written or verbal consent? - how realistic?

Consent needs to be suitable for the research purposes

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

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

UK **DATA ARCHIVE**

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 (special license)
  - data access permission from data owner prior to data release
  - data under embargo for given period of time
  - secure access to data (data analysis without actual access to or download of data)

UK **DATA ARCHIVE**

CAN YOU UNDERSTAND/USE THESE DATA?
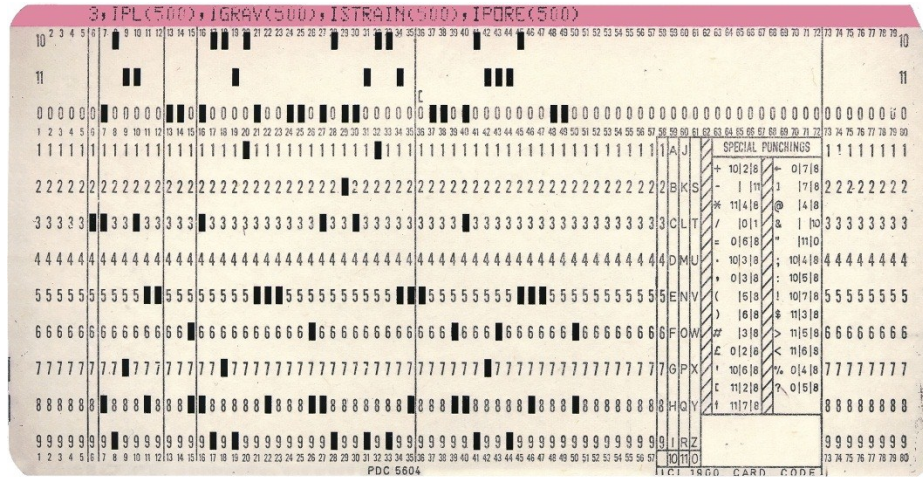
SrvMthdDraft.doc
SrvMthdFinal.doc
SrvMthdLastOne.doc
SrvMthdRealVersion.doc

UK **DATA ARCHIVE**

DOCUMENTING DATA

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 or lab books

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

- variable documentation: labels, codes, classifications, missing values, derivations, aggregations

- data listings for qualitative data

- any conditions of use and access?

DATA FORMATS

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

- best formats for long-term preservation
    - standard formats
    - interchangeable formats
    - open formats
    - .doc/.rtf, .pdf/a, .jpeg/.tiff, .mp3, .wav

- beware of errors in data conversion! Always check

DATA STORAGE

- ALL digital storage media are unreliable

- 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, external local copy and external remote copy
  - copy data files to new media two to five years after first created
  - check data integrity of stored data files regularly (checksum)
  - know your personal / institutional back-up strategy: network server/PC/laptop
  - test file recovery
  - know data retention policies that apply: funder, publisher, home institution
  - what to protect? Not only data, and not only digital

NEWS TECHNOLOGY

8 December 2010 Last updated at 11:43

Nasa sells shuttle PCs without wiping secret data

US space agency Nasa has been left red-faced after selling off computers without ensuring that highly sensitive data had been removed.

An internal investigation found 10 cases where PCs were sold despite failing data removal procedures.

Another four PCs - which were about to be sold - were found to contain data restricted under arms control rules.

Nasa is selling of hundreds of PCs used Space Shuttle programme

DATA SECURITY

- protect data from unauthorised access, use, change, disclosure and destruction

- personal data need more protection – always keep separate personal data

- 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
  - store most sensitive materials separately 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

DATA TRANSMISSION & ENCYRPTION

Sharing data between researchers / teams

- content management systems / virtual research environments
  - e.g. MS Sharepoint, Sakai (open source)
- file transfer protocol (ftp)
- Yousendit
- via physical media
- too often email attachments

- consider security needed / encryption
  - use an algorithm to transform information (A=1)
  - need a "key" to decrypt
  - should be easy to use, or won't be used (*.zip)
  - examples
    - Pretty Good Privacy (PGP) http://www.pgpi.org/
    - TrueCrypt: http://www.truecrypt.org/

RECENT DM RECOMMENDATIONS FOR CENTRES

- generate a data management resources **library**
- provide a data management **contact** for each project
- create a centre-wide data log using an agreed template
- use standard ethical review forms (append additional items to standard institutional forms where necessary)
- use agreed consent forms and information sheets
- collate an anonymisation log using a proforma
- use transcription proformas and rules/confidentiality agreements for transcribers
- set up a security **policy** for storing and sending data
- set up a policy for retention and destruction of data
- create statement for **copyright** and ownership of data
- provide recommendations on using standard data formats
- set up file sharing and storage procedures
- set up version control and file naming guidelines
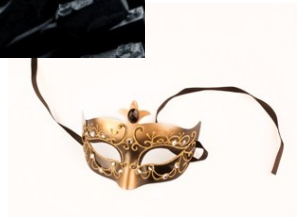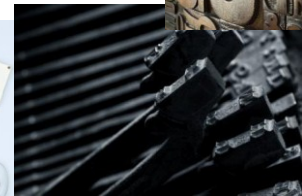
UK **DATA ARCHIVE**

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
- http://www.esds.ac.uk/aandp/create/esrcfaq.asp

OUR DATA MANAGEMENT GUIDANCE

- arguments for sharing data
- ethical and legal aspects of data sharing and re-use
- suitable data formats and software for long-term preservation
- documentation and metadata to understand and use data
- adequate security and controlled access to data
- data copyright
- quality control of data
- ensuring authenticity and version control of data
- backing-up data and files
- appropriate data storage

www.data-archive.ac.uk/create-manage

CONTACT

RESEARCH DATA MANAGEMENT SUPPORT SERVICES
UK DATA ARCHIVE
UNIVERSIY 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

UK **DATA ARCHIVE**