Sheffield Hallam University

Doctoral Data Think Tank

17 February 2017

Sheffield Hallam University

Doctoral Data Think Tank

Outline for today

10:00	**Arrival/Welcome**	
10:15	What is Data?	**Dr Murray Clark**
11:00	Data Management Planning (DMP) I -requirements and good practice	**Paul Ashwell**
12:00	Personal experience of DMP for the doctorate	**Natalie Haynes**
12:30	Lunch (provided)	
13:30	Data Management Planning (DMP) II -writing your own data plan	**Paul Ashwell**
14:30	Data analysis and tools	**Discussion with invited researchers**
15:30	**Round-up and Close**	

What is Data?

the concept of data
the methods of collecting
and using it

Sheffield Hallam University | Sheffield Business School

Ingredients of Research

- Construction of theory
- Collection of data
- Design of methods for gathering the data

- Research aims to add knowledge by applying various methods and strategies
- Assumptions about what is being researched and how it can be known
- Should different kinds of reality be approached in different ways?
- Can any approach guarantee certain knowledge or absolute truth?

What do you understand the following words to mean?

- Analysis
- concept
- explanation
- theory
- understanding

Sheffield Hallam University | Sheffield Business School

Data

- One of the most important words in a scientist's vocabulary
- Science – from French and Latin first appeared in English in 14th century – meaning knowledge
- Now associated with specific body of knowledge or skill or a branch of learning and in particular with theoretical and methodical study of nature
- Science is typically conceptualized as 'an empirical enterprise in which data are obtained and analysed in order to establish facts about the world and to test explanations' *(Thomas, 2006)."*
- *An understanding of data is crucial for the researcher because of its central role in the research process*

Sheffield Hallam University | Sheffield Business School

Language of Data

- Word data is derived from the Latin verb *dare* – to give and its past participle *datum.* A 'piece of information', literally 'something given'.
- **Origin:** Mid 17th century: Originally recorded as a term in philosophy referring to 'things assumed to be facts or givens.
- This identity with facts is retained in the modern core meaning. In the 'standard model' of science data are equated with observations.
- Through observations there is also a link with experiment in its original meaning of experience.
- Empirical data = experience data
- "By implication the objects of our knowledge the things we seek to acquire knowledge of, are independent of ourselves (not imagined or supposed) and posses qualities that we may or may not be able to discover depending on the methods we use to seek knowledge of them" *(Thomas, 2006:88)*
- Things and their properties are real and have an objective existence.

Language of Data

- Fact commonly seen to refer to that which is true or real
- Modern usage of the term fact or data often takes on a strong quantitative connotation
  - Data analysis until relatively recently implied statistical manipulation of numeric information – from experiments, surveys and observations.
  - Tendency to associate data with numbers of statistics, of measurements
- Now it is common to discuss both quantitative and qualitative data and their different forms of analysis.

Quantitative Research

Numerical data

- advantages: it can be easy and relatively cheap to collect);
- disadvantages: may not tell the whole story; may be misleading;

‘Quantitative Research’

- data are usual distinguished in terms of their level of measurement as nominal, ordinal, interval and ratio data.
- References usually made to processes involving data
  - E.g. data analysis, data reduction, data mining, data set, etc.
- Primary distinction for data itself is that between quantitative and qualitative data.
- This distinction is problematic – notion of qualitative data is ambiguous.

Main preoccupations of Quantitative Researchers
(Bryman, 2001)

- Measurement
- Causality
- Generalization
- Replication
- Reliability & Validity

Qualitative data

- Less likely to be conceptualized as variables– aim is to understand the meanings of human actions
- Typically data takes the form of texts or meanings derived from textual sources
  - Field notes or Interview transcripts
  - Research diaries, visual materials, etc.
- Qualitative data are not represented numerically and are not analysed using statistical analytical processes
- Data are categorized and examined for patterns and themes that can be integrated into narrative accounts

Main Preoccupations of Qualitative Researchers
(Bryman, 2001)

- Seeing through the eyes of the people being studied
- description & the emphasis on context
- Emphasis on process
- Flexibility & lack of structure
- Concepts & theory grounded in data

Common Threads of Data Analysis *(Hardy & Bryman, 2004)*

- **Quantitative & Qualitative Techniques** (Schism between camps)
  - Reinforcing this division is a mistake
  - Limits intellectual exchanges – learn more by viewing subject through a variety of lenses
  - Both approaches attempt to 'tell a story' from the data
- **Method of data analysis** – quantitative or qualitative?
  - Hypothesis testing or modeling – "to what extent empirical observations are represented by motivating theoretical model"
  - Qualitative researchers may or may not invoke models
  - Choice of method of analysis can involve assumptions about:
  - the nature of social reality, how it should be studied, the kind of research questions of interest, how errors of observation, measurement & estimation should be addressed.

Sheffield Hallam University | Sheffield Business School

Similarities between Quantitative & Qualitative Approaches *(Hardy & Bryman, 2004)*

Both are concerned with:

- Data reduction condensing vast amounts of data collected in course of fieldwork
- Answering research questions – even though precise nature of relationship between RQs and data analysis may be different
- Relating data analysis to research literature – relate issues that drive and emerge to existing literature
- Treating frequency as a starting point for analysis
- Ensuring deliberate distortion doesn't occur
- Importance of transparency

Suicide Rates in Selected Countries

Bahamas	2.2
Greece	3.1
Italy	5.5
Spain	5.7
UK	6.3
United States	10.2
Sweden	14.4
France	15.9
USSR	17.4
Austria	18.9
Denmark	20.1
Finland	23.7
Hungary	32.0
Sri Lanka	35.8

What types of data will you be working with?

Methodology

- Experimental research
- Survey Research
- Ethnography
- Phenomenological research
- Grounded theory
- Heuristic inquiry
- Action research
- Discourse analysis

Sheffield Hallam University | Sheffield Business School

Methods - Data Collection

- **Measurement & scaling**
- **Structured Observation**
- **Questionnaire**
- **Observation -** participant, non-participant
- **Interview – s**tructured, semi-structured, conversation (with a purpose)
- **Focus group**
- **Case study**
- **Document analysis**
- **Interpretative methods**
- **Content analysis**

Sheffield Hallam University | Sheffield Business School

Types of Data

- According to **Meaning**
  - Facts - characteristics or situations that exist or have existed;

    descriptions of individuals' behaviour
  - Awareness or knowledge of some object or phenomenon
  - Intentions - acts that people have in mind to do
  - Attitudes and Opinions - views, preferences, inclinations or feelings of people towards some object or phenomenon
  - Motives of individuals - internal forces that channel behaviour in a particular way

Diamantopoulos & Schlegelmilch (1997)

Types of Data

According to **Source**

- Primary data - collected with specific purpose in mind
  - usually gathered by researcher via
    surveys, experiments, observation methods
- Secondary data –data which have not been expressly gathered for the immediate study but for some other purpose
  - Published statistics, annual report, database services, etc.

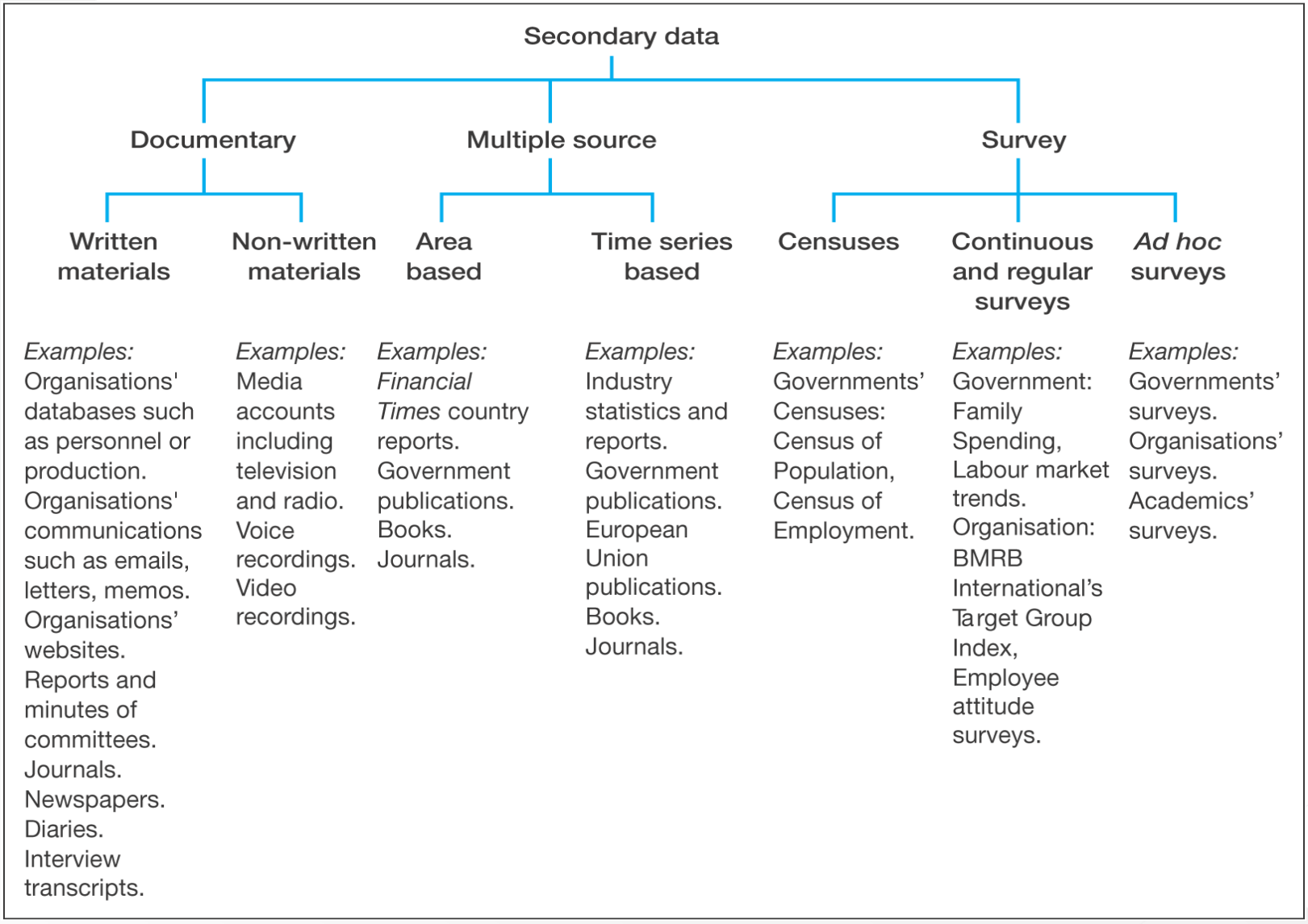
According to Time Dimension

- Cross-sectional Data – single point in time
- Longitudinal data –relating to a number of time periods
  Trend data

Diamantopoulos & Schlegelmilch (1997)

Sheffield Hallam University | Sheffield Business School

Types of Secondary Data



Source: © Mark Saunders, Philip Lewis and Adrian Thornhill, 2006

21

Conclusion - Data

- The concept of data - ambiguous and elusive
- uncertainty (especially in social sciences) as to where data (is) are located - difficult philosophical questions- considerations of the 'real' and our relation to it
- So, 'What are data?'
  - Conventional usage - distinguishes between the world (the source of data) and textual records of features of the world (the data)
- Much **qualitative** 'research' involves 'ambiguity' when **text** (an aspect of that world) is used itself as a source of data.
  - Texts are represented and described by other texts - relativism?
- **quantitative** - world is assumed to be directly accessible to the senses - empiricism -facts speak for themselves
  - problem arises when we encounter non-observables
- **Social constructionism** -data are never 'innocent' of presuppositions - they are constructed both by the researcher and those whom the researcher studies.

Sheffield Hallam University | Sheffield Business School

The words - together with synonyms and related terms, are at the heart of the process of analysis

- *concepts* - abstract or general ideas, which are important to how we think about particular subjects or issues
- *theories* - suppositions which explain, or seek to explain something
- *explanations* - statements which make something intelligible about why things are the way they are
- *understanding* - perception of the meaning of something, in this case the subject area, the issues and/or the research questions under consideration
- ANALYSIS is about the search for explanation and understanding, in the course of which concepts and theories are likely to be advanced, considered and developed

Sheffield Hallam University Library Research Support

Data Think Tank

Data Management Planning

The benefits of Research Data Management and how do it

Paul Ashwell

Research Support Librarian

p.g.ashwell@shu.ac.uk

Session outline

- Research Data Management in a nutshell
- Why, when and what to plan
- Managing live data
- Selecting data for preservation
- Preserving and sharing data
- Write your own data management plan

RESEARCH DATA MANAGEMENT IN A NUTSHELL

CC0 Public Domain

What, why and how?

- Video: The what, why and how of data management planning (Research Data Netherlands)

Research Data Management

Plan

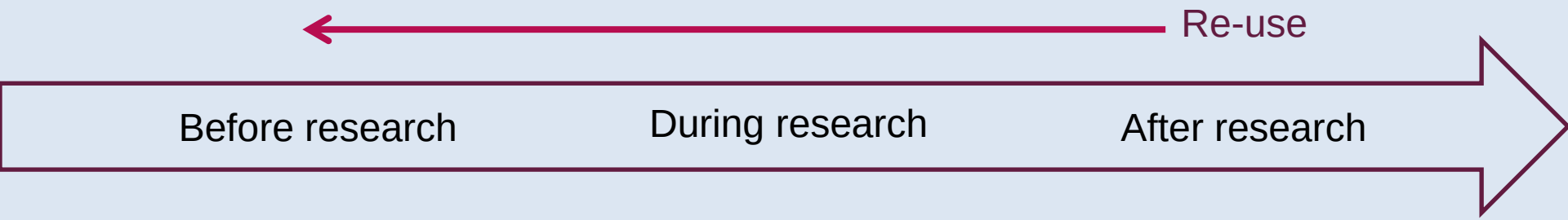
- Including ethics and copyright

Manage

- Documenting and organising data
- Storing and backing up data

Archive & Share

- Selecting which data to keep
- Preserving data
- Giving access to data



Research Data Management

Plan

- Data Management Planning Tool

  online tool for planning research data, also as pdf

  http://dmponline.dcc.ac.uk
  http://bit.ly/shu-dmp-doctoral

Manage

- Research Store (Q:\Research)

  safe and secure storage of 'live' research data

  http://bit.ly/shu-research-store

Archive & Share

- SHU Research Data Archive (SHURDA)

  archive for digital and non-digital research data

  http://shurda.shu.ac.uk

Before research | During research | After research

Benefits for the researcher

- Direct benefits of managing live data
  - **Risk management -** avoid data loss and unauthorised access - *storing and back-up*
  - **Usability and sustainability** of research data and projects through *documentation*
  - **Efficient organising** through logical *folder structures, file naming and versioning, and accessible formats*
- Benefits of data archiving and sharing
  - Research **integrity**: openness and transparency
  - Personal and institutional **reputation**: increase in citation rate of associated research output of up to 69%, opportunities to collaborate
  - Altruistic benefit of re-using primary research data in new ways may create **new insights and advance academic progress**
  - Increased **success in acquiring research funding** and optimal choice of journals for publication through compliance

WHY, WHEN AND WHAT TO PLAN?

Why plan?

- **Decisions made at the beginning** determine what you can do with your data later on:
  - **informed consent and appropriate anonymisation** should allow for data preservation and data sharing at the end of your project
  - when re-using **secondary data** there may be restrictions to what you can do (IP issues)
- **Effective storage of live data** - You can name, structure, document and choose appropriate file formats for your data as you collect or create them from the very beginning - doing this later on can be time consuming and sometimes impossible
- **Help** you and your team to understand your own data
- **Prepare** for data preservation and sharing

When to plan?

- At RF1 stage you need awareness of possible issues - e.g.
  - using third party materials (secondary data - IP)
  - informed consent of human participants for preserving and sharing
  - storing your live data
  - available support in the University
  - any requirements from your funder or external partners (if you have any)
- At RF2 stage you write a core DMP using a template
- A plan is a living document: not all details of your project may yet have emerged, and research projects can develop in unpredictable ways, so keep your plan up-to-date as your research develops
- Discuss your plan regularly with your supervisor, e.g., once a year

RF2

- You are strongly encouraged to write a data management plan as part of your RF2 or equivalent confirmation
- Discuss your plan with your supervisor
- Revisit your plan regularly and keep discussing it with your supervisor
- Use our online tool - DMPOnline - it contains useful SHU specific guidance (or for webphobes - PDF template)

What to plan? A template

1. What data will you collect or create?
2. How will you deal with any ethical and copyright issues?
3. How will your data be documented and described?
4. How will your data be structured, stored, and backed up?
5. What are your plans for the long-term preservation of data supporting your research? (selecting and preserving)
6. What are your plans for data sharing after submission of your thesis or article?

http://bit.ly/shu-dmp-doctoral

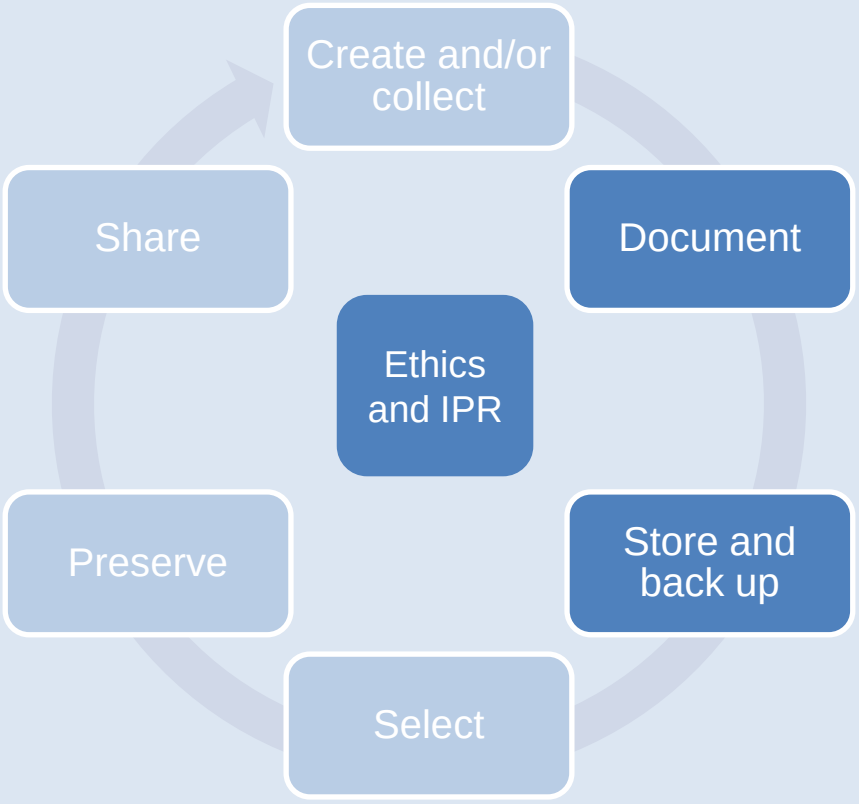
MANAGING LIVE DATA

CC BY-NC-SA 2.0

What to plan? - Live data

1. What data will you collect or create?
2. How will you deal with any ethical and copyright issues?
3. How will your data be documented and described?
4. How will your data be structured, stored, and backed up?
5. What are your plans for the long-term preservation of data supporting your research? (selecting and preserving)
6. What are your plans for data sharing after submission of your thesis?

http://bit.ly/shu-dmp-doctoral



Documentation

- All the information that a future user may need to replicate your research, and to re-use and interpret your data
- Data-level documentation
  - describes the data contained in a file
  - often embedded in the file (spreadsheets, SPSS, NVIVO), sometimes as a separate .txt file
  - explains variables, codes, missing values, etc.
- Supporting documentation that describes the research project, the data creation and processing process, and the general context
  - often already available as laboratory notebooks, questionnaires, interview guides, protocols, reports, PhD chapters, and publications
  - could be added as separate files

File formats for archiving

Type of data	Recommended formats	Acceptable
Tabular data	SPSS (extensive metadata), CSV (minimal metadata)	MS Access, MS Excel
Text	plain text (.txt), rich text format (.rtf), XML	HTML, MS Word, NVivo
Images	TIFF	JPEG, PDF, RAW, PSD
Audio	FLAC	mp3, AIFF (Apple), WAV (Microsoft)
Video	MPEG-4 (.mp4), motion JPEG 2000 (.mj2)	MKV / FFV1 (momentum to adopt this as a standard)

- It is best to use formats that are:
  - lossless (tiff, flac) rather than lossy (jpeg, mp3)
  - standard formats (MS Office, rtf, SPSS, NVivo)
  - non-proprietary / open formats that are well documented (pdf/a, csv, tiff, flac, odf, xml)

http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx

File naming

Tips

- Make sure your file names are unique, and keep them independent of their location ('interview_2015_05_01' is better than '2015_05_01' even if the file is located in a folder called 'interview')
- Use file names that are concise but informative, so that you can tell the contents of the file without having to open it
- Be consistent
- Think about what comes first in the filename, because operating systems usually sort files alphabetically
- Do not use special characters in a filename such as : & * % $ £] { ! @

FG1_CONS_2016-02-12 *is the file that contains the transcript of the first focus group with consumers, that took place on 12 February 2016*

File naming conventions could include

- which terms you use in your file names (vocabulary)
- which abbreviations you use
- punctuation and spelling, e.g. will you use CamelCase or not, and will you use dashes (-) or underlines (_) instead of spaces
- format of dates, e.g. YYYY-MM-DD is easier to sort than DD-MM-YYYY
- versioning
- the order of the elements in the filename

Int024_AP_2015-06-05 *is an interview with participant 024, interviewed by Anne Parsons on 5 June 2015*

Activity

What are the advantages and dangers of the following storage options?

- Networked drives
- Local drives on PCs and Laptops
- Cloud-based storage
- External portable storage

What would be the best way to back up?

Research store (Q:\Research)

Places to store data

- Local drives on PCs
  and laptops (risk of data corruption, data loss, unauthorised access if unencrypted)
- Cloud-based storage (host has access to all of your data, they may have the right to use/publish your information)
- External portable storage (risk of data corruption, data loss, unauthorised access if unencrypted)

The Q:\Research - network storage which is

- **secure** (firewalls, passwords)
- **safe** (daily automatic backup to two remote locations)
- **flexible** (enough space, flexible access)
- Ask your supervisor for your own space on the Research Store (Q:\Research)
- More info at
  research.shu.ac.uk/rdm/research-store

SELECTING DATA FOR PRESERVATION

CC0 Public Domain

Activity

- It may not be feasible or desirable to keep all the data you collect or create, so what data would you select for long-term preservation?
- **Discuss the following questions**
  - What gives research data long-term value?
  - Why would ***your*** data be interesting to others?

When are data?

- What is the data? Sound files of interviews, transcripts, notes on interviews, NVIVO files with coded transcripts?

- When are data?
- Temporary reifications of on-going processes
- "Moments of organisation" in a continuing flow of research activity, rather than stable entities (visual arts)

The purpose of data

- To produce original research results

  "Research data refers to any type of data created, collected or generated in a digital or non-digital form that is analysed *to produce original research results*."
  SHU's research data management policy
- To allow validation of research findings

  "Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as *necessary to validate research findings*; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created."
  Engineering and Physical Sciences Research Council (EPSRC)

Selecting data

1. *Why* are you making the data available?
   - To allow further analysis
   - to support validation/scrutiny of research findings
2. What data ***must*** be kept because of policies and regulations?
   - e.g. the University, your funder, your journal, relating to informed consent
3. What data ***should*** be kept because it is of long term value?

Long-term value

- Is the data of good enough quality in terms of completeness, sample size, accuracy, validity, reliability or any other criterion relevant in your subject domain?
- Is the data sufficiently documented to allow re-use by your peers?
- Is there likely to be a demand for your data?
- Is it difficult to replicate your data?

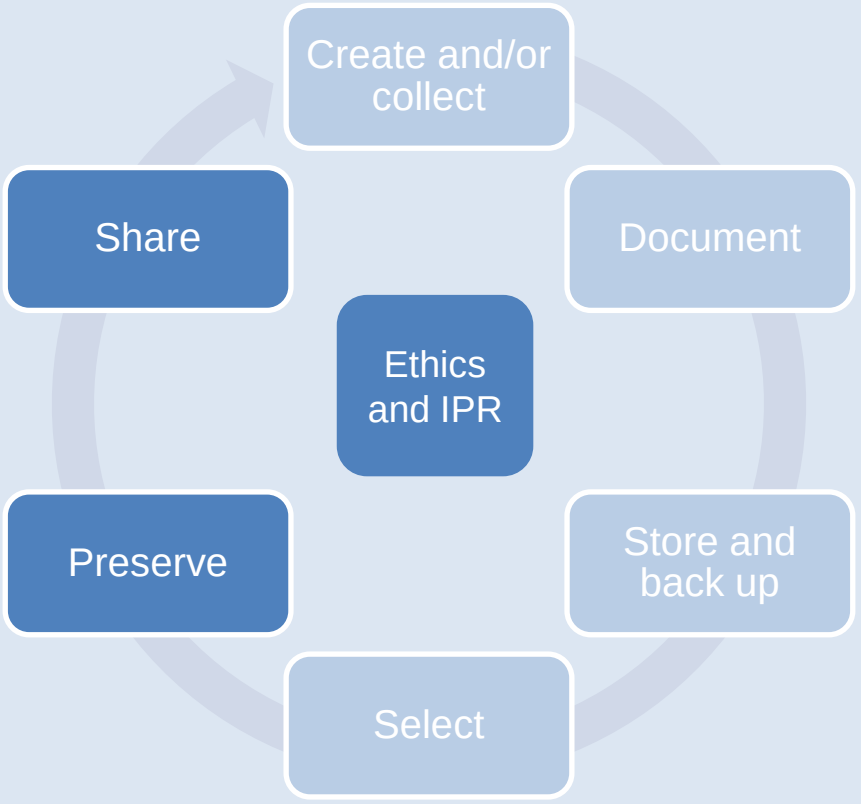
PRESERVING AND SHARING DATA

CC0 Public Domain

What to plan? Preserve and share

1. What data will you collect or create?
2. How will you deal with any ethical and copyright issues?
3. How will your data be documented and described?
4. How will your data be structured, stored, and backed up?
5. What are your plans for the long-term preservation of data supporting your research? (selecting and preserving)
6. What are your plans for data sharing after submission of your thesis?

http://bit.ly/shu-dmp-doctoral



Preserve your data - SHURDA

- Don't do it yourself - it is difficult to curate your data and who is going to find it?
- SHU Research Data Archive, SHURDA (also for registration of all your preserved datasets): http://shurda.shu.ac.uk/
- SHURDA gives you a DOI for your dataset to be used in your thesis and published papers
  - "All data created during this research are available from the Sheffield Hallam University Research Data Archive at http://dx.doi.org/10.15125/12345."
- (Inter)national subject-specific data repositories: http://re3data.org/
- University policy is to preserve primary research data for 10 years since the last time any third party has requested access to the data, or as long as any external funder or contractual partner requires

Share your data

What is data sharing?

- "Data sharing is the practice of making data used for scholarly research available to other investigators." (Wikipedia)
- But also with your future self

How can you share?

- By request
- Linked to a publication
- Open data in a repository

Who is involved?

- The data sharer
- The data repository and/or journal
- The secondary data user
- Support staff (e.g., Library)
- Research participants
- Research collaborators and external partners (e.g., English Institute of Sport)
- Research funders and sponsors

Activity

- List some of the reasons why you, or other researchers you know, may feel your ability to share data is restricted
- What actions could be taken to reduce or overcome these restrictions?

Constraints on sharing	Solutions and/or approaches

Managing restrictions on sharing

Ethical and legal (Data Protection Act)

- Balance data protection with data sharing
  - informed consent: cover current *and* future use
  - confidentiality: anonymisation
  - controlled access: giving access to whom, and under what conditions?

Intellectual Property Rights (IPR)

- Consider licensing options -
  such as Creative Commons, and controlled access
  - Creative Commons: attribution required (CC BY), share-alike required (CC SA), non-commercial use only (CC NC), no derivative works (CC ND)

Sheffield Hallam University

Personal experience of data management planning for the doctorate

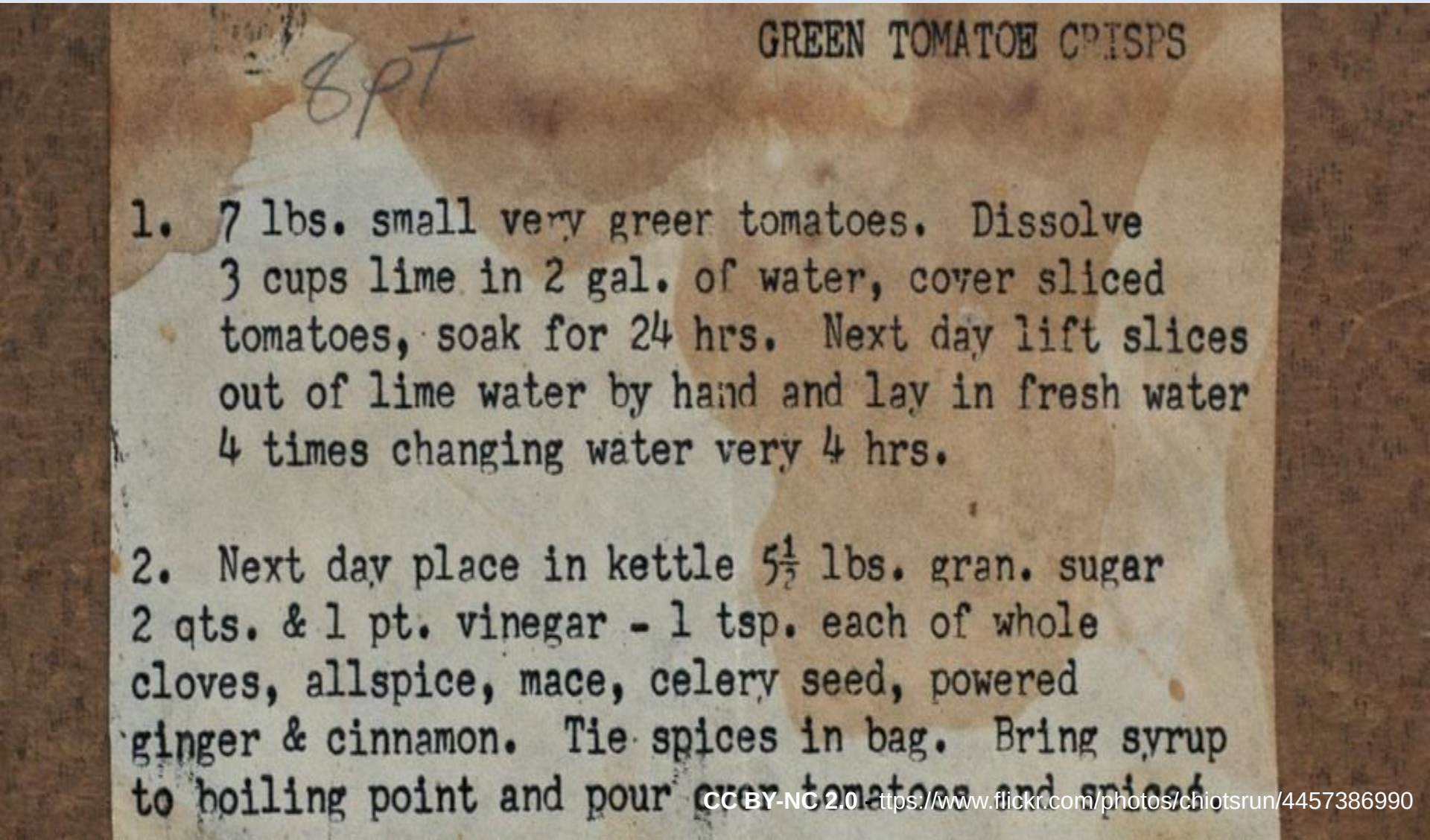
Natalie Haynes

Sheffield Hallam University

Discussion points

- Think of your data management plan as a live document and go back and review it as you start your data collection
- It is useful to read the SHU research data management policy and get the terminology right
- Think about your DMP from your own point of view and link it to your methodology
- Natalie's plan is available online as an example

WRITING YOUR DATA MANAGEMENT PLAN

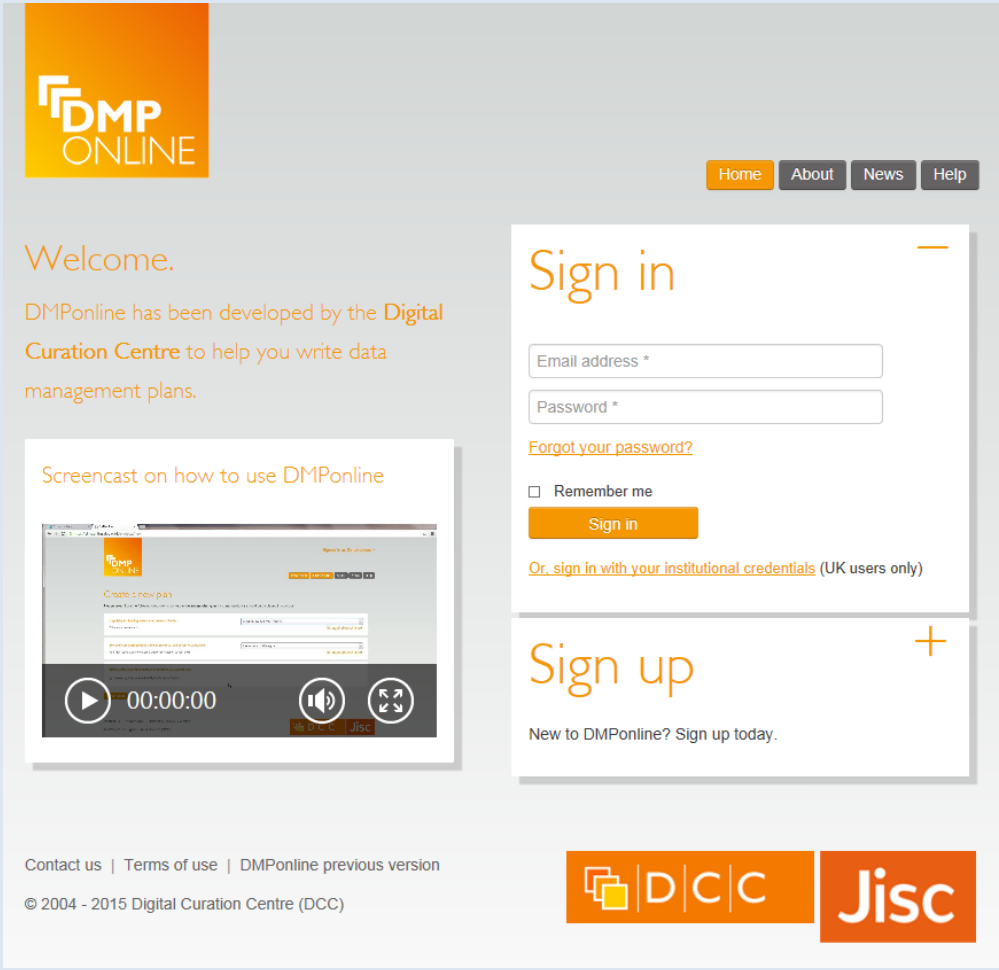


CC BY-NC 2.0 - ttps://www.flickr.com/photos/chiotsrun/4457386990

DMPOnline

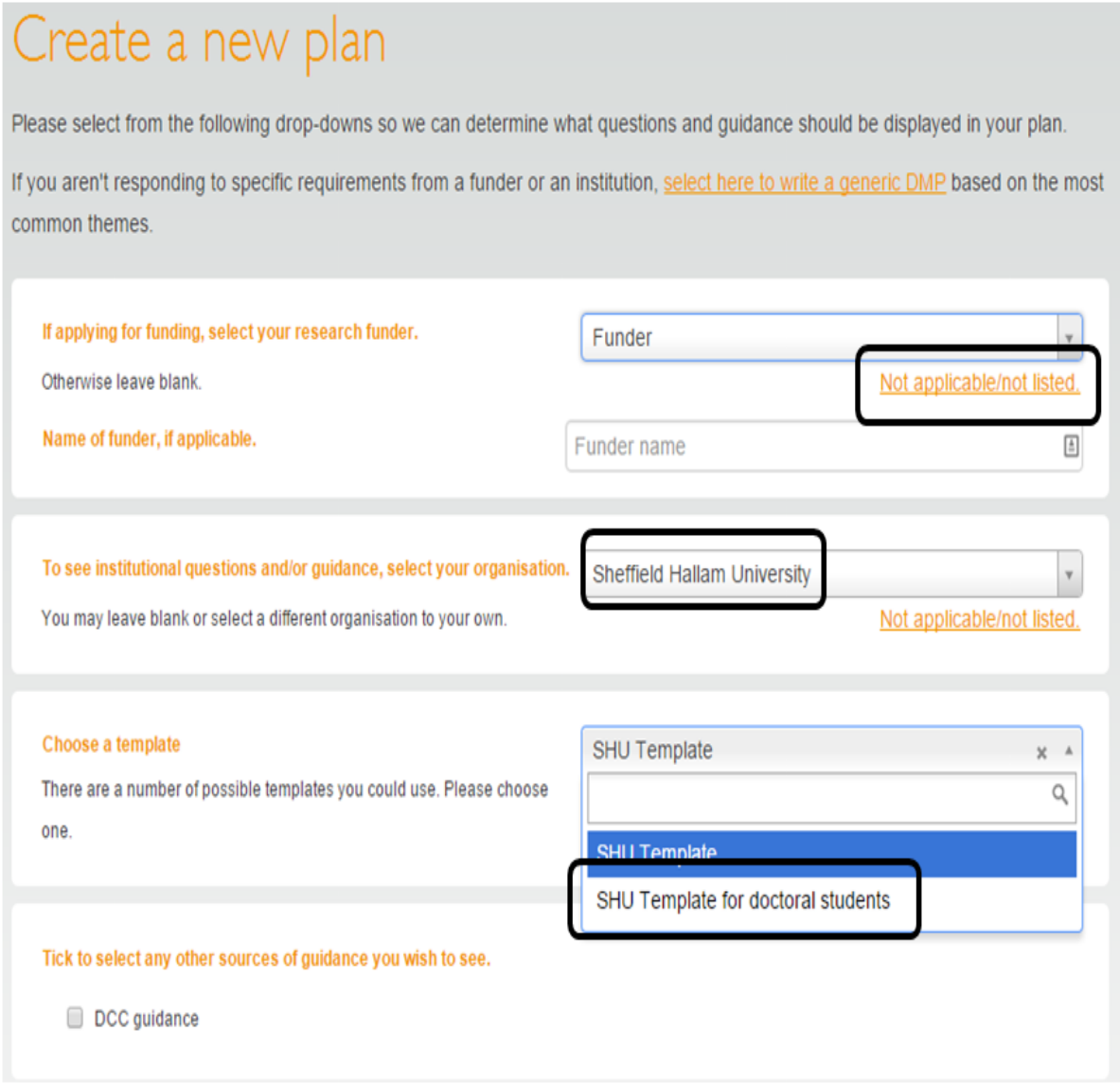
- DMPOnline
  - Templates for funders, a generic template for other SHU-based research, and a generic template for doctoral research
  - SHU-specific guidance
  - Allows your plan to grow
  - *Share* your plan with others that have a DMPOnline account, or *export* it as Word, pdf or html
  - http://dmponline.dcc.ac.uk

- A template is also available as a pdf from http://bit.ly/shu-dmp-doctoral



Create your plan in DMPOnline

1. Go to http://dmponline.dcc.ac.uk and create a new plan
2. Click on "Not applicable /not listed" for research funder
3. Select "Sheffield Hallam University" to see institutional guidance
4. Select "SHU Template for doctoral students"



Tips

- **Don't write your plan in isolation** -
  talk to colleagues, your supervisor(s), ethics committees, or the Library Research Support team
- **Be pragmatic** -
  plan what is useful, and keep it simple, short and specific
  - "A good plan implemented today is better than a perfect plan implemented tomorrow" (George Patton)
- Give the plan space to change over the life of the project

RDM and SHURDA Support

Online

- RDM guidance website - http://research.shu.ac.uk/library/rdm/
- Online learning module on Blackboard as part of the **Academic CPD Online Courses**
  - 30-60 minutes, added to Epigeum courses

Personal advice

- library-research-support@shu.ac.uk
  - DMP advice and DMP read-through
  - depositing data

Drop-in sessions

- Monthly drop-in sessions on both campuses - http://bit.ly/shu-oa-rdm-dropin

Where can I get more help?

Website - http://research.shu.ac.uk/library/

Workshops - https://blogs.shu.ac.uk/doctoralschool/events/

RIO (for REF related enquires)

Contact us - library-research-support@shu.ac.uk

@SHUResearch

0114 225 3852

Sheffield Hallam University

Discussion of data analysis and tools

Sheffield Hallam University

Discussion points

- How would you respond to a request to share your data?
  - think about who **owns** it, informed consent (where relevant) and **when** is the best time to share it.
- What contingency plans can you put in place for if you are unable to collect your data as planned?
  - pros/cons of 'snowballing' technique to generate data
- Is your data 'useful' without your analysis?

Sheffield Hallam University

Tools - questions to consider
(with thanks to Karen Kilner)

- Does it do all the necessary analysis?
- Will I have to justify my choice of tool?
- Do I need to learn a new package / tool / language?
- Is it available to me?
- Is there expertise / support / training?
- Are there specific requirements for reporting or publication in my discipline which make specific software vital?
- Is the output of sufficient quality for publication?

Sheffield Hallam University

Useful links

Library research support team pages

- http://research.shu.ac.uk/library/rdm/begin.html

Online courses on shuspace (Academic CPD Online Courses)

- Research Data Management
- Planning your data analysis - Research Methods in the Sciences
- Data analysis - Research Methods in the Social Sciences

Vitae resources

- Focus on open research and open researchers

Available software

- https://students.shu.ac.uk/shuspacecontent/it/software-list

Training via Lynda.com (Nvivo, SPSS, R, Matlab etc)

- http://extra.shu.ac.uk/lynda/

Sheffield Hallam Doctoral School

https://blogs.shu.ac.uk/doctoralschool/

doctoralschool@shu.ac.uk

@SHDocSchool