

**I've collected my data, so
what do I do with it now?**

Research data management

Session 3

Problems and Practical
Strategies and Solutions

Tutor Notes

DATUM for Health

www.northumbria.ac.uk/datum

Project funded by JISC

Copyright holder: Northumbria University, School of Computing, Engineering & Information Sciences, 2011

Materials made freely available under a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 UK: England & Wales license

Session 3 Problems and Practical Strategies and Solutions Notes for Tutors

SESSION DETAILS

Aims and Objectives / Learning Outcomes:

By the end of this session participants will have:

- an awareness of research data management problems
- knowledge of practical solutions to research data management problems
- the ability to develop a research data management strategy for the remainder of their PhD and in their future research career

Session Content:

- Problems with managing research data
- Problem triggers and practical solutions for managing data
- Data confidentiality and data protection
- Data storage and data security
- Data classification, file structures, file names and metadata
- Practical exercise on strategies for data anonymisation and data sharing
- Practical exercise on file management

Structure:

10.00 - 10.05	Introduction
10.05 - 10.40	Data problems and conflicts
10.40 - 10.55	Exercise 1 – Anonymisation Coffee / tea
10.55 - 11.15	Discussion on the issues raised by the exercise
11.15 - 11.35	Information security
11.35 - 11.45	File management and metadata
11.45 - 12.20	Exercise 2 - File Management
12.20 - 12.30	Final comments and questions

NB The timings are very tight. Coffee/tea can be taken during the anonymisation: this can help student engagement with the exercise.

Directed Learning Tasks:

- Reflect on your strategies and procedures for research data management, e.g.
 - Do you need to take any further actions to ensure your research data is appropriately anonymised and capable of being shared in the future, if appropriate?
 - Do you need to revise your file / document naming system, and / or the file structure for storing your research on your university drive
 - Are the ways in which you physically store your data sufficiently safe and secure?
- Embed research data management into your PhD work activities, e.g.
 - Make an entry in your research diary / portfolio about RDM strategy and actions going forward
 - Make an entry in your training plan about RDM

- Meet with your supervisory team to report on RDM learning and plans going forward

Handouts:

- PPT slides 3 per page handout (provide the students with a hardcopy for use during the session)
- Model consent form, provided by the UK Data Archive ⁱ (provide the students with a hard copy for use in the session)
- Exercise 1 - Anonymisation – used with the consent of the UK Data Archive. (Contact Louise Corti (corti@essex.ac.uk) for further information)
- Exercise 2 – File management exercise (provide the students with a hardcopy - as 3 separate pages - for use in the exercise)
- University of Edinburgh Records Management Section, Standard Naming Conventions For Electronic Records: The Rules ⁱⁱ (provide hard copy for students to use in file management exercise)
- Evaluation form for session (provide the students with a hardcopy for use at the end of the session)

NOTES TO ACCOMPANY POWERPOINT SLIDES

INTRODUCTION (Slides 0 - 3)

This part includes a general explanation about who the presenter is and that the session builds on the two earlier DATUM sessions.

Slide 1: Overview

This session will look at some of the issues that confront you in managing your research data, and suggest some strategies & tactics for dealing with these problems. I will look in particular at confidentiality and data sharing, data storage and security and file structures, file names and metadata. There will be two worked exercises, on anonymisation and on file management. Sessions 1 and 2 have outlined the issues and what you need to do; this session is looking at how to do it.

Slide 2: Learning outcomes

We've established in the previous sessions that by planning it is possible to pre-empt and manage many potential research / data management problems.

In considering the problems that confront the researcher in managing data you need to remember the benefits that underpin good data management. The skills from these research data management sessions are essential for your future research and regardless of your future career choices these are skills that are transferable into the wider workplace beyond the academic context.

Slide 3: Why is Research Data Management Important?

This slide is taken from the first session and is just a reminder of why RDM is important.

DATA PROBLEMS AND CONFLICTS (Slides 4 -12)

This part dealt with the key problems and solutions for managing research data.

Slide 4: What Problems Do You Have Managing Your Research Data?

There are a number of reasons why you might encounter problems: conflicts between different processes, e.g. trying to protect a participant's privacy but also wanting to share data, lack of resources and so on.

After Session 2 you were asked to think of any problems you'd encountered in managing your data. Do you want to share any of those now? They'll be opportunities to talk about particular aspects during the session, and we'll have a round up at the end for final comments and questions.

Tutor: Allow 2 minutes for open discussion.

Examples of some research management issues

Universities do change their research and ethics guidance and rules over time. This can impact on a 3 year PhD programme – it is important to keep up-to-date with any changes. For example, in 2009 Northumbria University changed the rules on the intellectual / commercial property rights over PhD students' developments to ensure that the University's contribution to PhDs was established. Within research any ownership rights should be understood and documented at the start of the project to avoid any later conflict.ⁱⁱⁱ Northumbria University does have a business advisor who can discuss any potential commercial aspects underpinning a piece of research.

Where research is being conducted across different legislative regimes then the research must be developed to take account of the legislative requirements wherever data is being gathered, held and disseminated. In addition to legislation different cultures have different expectations about consent.

Within the context of a PhD novelty is critical. It is possible to publish articles in order to protect one's originality and document key findings before others reach the same conclusions. These publications can and should be quoted / cited.

Slide 5: Conflicting Data Management Considerations

There is a balance to be struck in terms of ensuring the maintenance of research confidentiality and handling appropriate data sharing. This process is easier if it is managed from the outset so that data ownership and consent for sharing is obtained.

Slide 6: Data Ownership and Consent

Consent is a critical part of the process for good research data management.

Examples of consent forms are available from the UK Data Archive 'create and manage' webpages^{iv}. In your handouts there is an example of their model consent form.

Although you may hold the physical data sets (e.g. Interviews, questionnaires etc) this does not mean you own all the copyright within that data. A participant still owns their own words unless copyright is assigned to the researcher. The UKDA's model consent form, e.g., contains the statement "I agree to assign the copyright I hold in any materials related to this project to [name of researcher]". So you must obtain clear consents that their data can be used for a range of research purposes.

You must not forget that if you ask someone else to transcribe your data then you should also require they complete an undertaking of confidentiality. In addition you need to consider your rules for transcription. The UKDA has some useful practical resources on this topic.^v

Slide 7: Confidentiality and Data Sharing

There are ways of managing data access and maintaining confidentiality. For example, data can be anonymised. When data is anonymised you will need a key that relates the code / pseudonym of the participant to their actual personal details. You will also need to maintain an audit trail between the various forms of the data, e.g. for an interview this could be the audio record, the transcript, the anonymised transcript. You can use 'member checking', i.e. asking the participant to agree that the anonymised transcript is acceptable to them. You would also need a record of this agreement, such as a signed form. The UKDA provides a

number of resources about anonymisation.^{vi} We will carry out an anonymisation exercise in this session.

For small data sets it is possible to use additional techniques to maintain confidentiality e.g. Barnardisation. In small statistical datasets, even when anonymised, individuals might still be identifiable, e.g. a village containing one person with a rare medical condition. Barnardisation involves randomly altering the figures by a small amount.

The slide gives the URL of a case relating to the Barnardisation of data that has caused controversy. The data related to a small sample of patients with leukaemia. This data was requested as a Freedom of Information request and after a legal battle released. The issues surrounding the release focused on whether the data was barnardised, whether the patients could be identified and what the public interest in the data was. Full details are available at the link but you should realise that data can be requested and made available as the result of Freedom of Information requests or Data Protection subject access requests. The Freedom of Information Act gives people the right to request official information (which can include research data) held by public authorities (which includes universities), unless there are good reasons to keep information confidential. The Data Protection Act gives people the right to know what information an organisation (or you as a researcher) holds about them and the right to correct information that is wrong.

If you receive a request for data under FoI or DP, then the process of making data available should be handled through the university's compliance officer. For Northumbria University refer to the Freedom of Information Officer. If assurances, regarding confidentiality, have been agreed within the consent form then it is very unlikely (although not inconceivable) that data will be released. The circumstances under which the data had been gathered would then be taken into account. An example of where data might be released is if it had damaged the rights of another person. Where data is very sensitive and potentially damaging to an individual it may be possible to keep the data closed (embargoed) for a period of time.^{vii}

Slide 8: Problem Triggers

The requirements for managing personal data within the UK are set out in the Data Protection Act 1998.

Slide 9: Data Protection Act – Managing Personal Data

There are eight key Data Protection Principles. Data should be:

1. Obtained and processed fairly and lawfully
 - effectively this means following these DP principles
2. Obtained and processed only for specified lawful purposes
 - organisations should be open about what they are going to use the data for, and these uses should not be unreasonable or for an unlawful purpose
3. Adequate, relevant and not excessive
 - the data should be related to the use(s) the organisation has specified, and be the minimum required
4. Accurate and, where necessary, up to date

5. Kept no longer than necessary
 - data no longer needed should be deleted
6. Processed in accordance with the rights of the data subjects
 - there are a number of these rights, e.g. to see a copy of the data held about them, and to have incorrect data corrected / deleted
7. Kept securely and safely
8. Only transferred outside EEA with specific protections
 - not all countries have the same level of data protection

You can see how these principles are directly relevant to your research data. However, with appropriate consents and good research data management procedures you should be able to meet all these requirements. JISC provides guidance on data protection in the HE sector

Slide 10: Data Protection Act – Sensitive Personal Data

In addition, under the terms of data protection legislation, where **sensitive** data is held then there are additional legal requirements. The slide lists what data is deemed to be sensitive. Health research is by definition collecting sensitive data.

The subjects of that data should normally explicitly consent to each purpose for which it is used. There are some exceptions. For example, it would be possible to use and anonymise data from patient hospital records without the specific consent of an individual – although hospitals have ethics committees that would require clarification about the research purpose for which the data is being used and therefore this is a different type of check.

A number of the sensitive data categories will have a bearing in a health context. For example religious beliefs may influence health choices. But if you are gathering data about religious and health opinions/beliefs then it is important to inform the research participants that both categories are being collected. Individuals can be content to have health information captured but be sensitive about other details, such as their sexual life choices or opinions, being recorded.

Slide 11: Problem Triggers

This slide considers some other problem triggers.

Some examples:

When you are obtaining third party funding for research it is important to be clear that the research will still be conducted and delivered impartially. This can be achieved by signing a research agreement with the funder. This establishes and sets out the expectations from the outset that the research must be conducted ethically and rigorously in accordance with research principles. The research data management plan (DMP) also acts as a record demonstrating the rigour with which research has been conducted. Consider the case of a PhD student researching the nutritional value of sausages. The research was funded by one sausage manufacturer and this was a problem when that manufacturer's sausage did not compare well to other sausages. Funding was withdrawn from the student and it made it harder for them to complete their PhD.

We've already mentioned that organisations, such as universities, researcher funders / sponsors, could alter their requirements during the timespan of your PhD.

In addition to individuals consenting to take part in research, it is also important to consider any research impact on private or public sector participants and to gain the appropriate organisational consents for anonymisation. In small sectors it can be much harder to anonymise and protect an organisation whilst delivering the full value of the research and therefore this must be discussed and considered at the outset.

It is normal to provide individuals with the right to withdraw from research. It is important to ensure that where an individual withdraws then either their data can be easily extracted or alternatively any data already gathered whilst that person was participating can still be used in an anonymised context. The latter of these options is preferable. However, this may become contentious if there is a high profile outcome from a piece of research. One such example, was the case of Andrew Wakefield's research determining whether there was any link between the MMR vaccine and autism. This case received huge media attention over a number of years and therefore had a potential impact for the small number of participants. In this instance many of the participants in the research continued to support the research. The research has now been discredited.^{viii}

It is important to understand the requirements of your sector. Discussing research can seem daunting and negative. Many researchers are afraid that it will put individuals and organisations off participation. However a free and frank discussion can improve the quality of the research. It is also better to be honest about any possible impacts rather than having participants withdraw part way through a project.

Slide 12: Some Health Related Guidance

This slide provides links to some of the guidance the health sector has published within different contexts e.g. on ethics and confidentiality. NB Caldicott Guardians are responsible for the management of personal data within hospitals.

ANONYMISATION EXERCISE (Slide 13)

Note: In the timing of the session, students can have their tea / coffee whilst doing this exercise.

This exercise was prepared by the UK Data Archive who gave permission for its use and inclusion in these materials (Contact: Louise Corti - corti@essex.ac.uk)

The handout pack from the UK Data Archive explains the full context of the exercise, sets out the exercise and provides an answer sheet to be handed out to the students.

The exercise is based on a real piece of research - the foot and mouth disease epidemic in Cumbria, 2001-3 - where the research data has been deposited in the UKDA.^{ix}

In brief, students are given a piece of text - an interview - and asked to read it through and anonymise it. [Allow 15 minutes for this stage]. This is followed by open discussion about what anonymisation decisions the students make and why. Model answers are provided at the end of the discussion. [Allow 20 minutes for this stage].

As noted before, the UKDA provides rules on model transcription.^x If data is transcribed by a third party, then you need to give that person a clear set of transcription guidance (e.g. are

all the ums and ers to be included?) and they should also be asked to sign an undertaking of confidentiality so that they are sensitive to the confidentiality of research data.

INFORMATION SECURITY (Slides 14 - 23)

This section discusses how research data can and should be managed to ensure its security across a range of domains.

Slide 14

This slide is a quick ice breaker to the topic of information security. TASK: In open discussion, ask the students to spot any problems within the office set up!

Examples

1. Open shelving and open filing cabinet
2. Computer password on post-it note
3. Confidential waste by bin
4. Possible suspicious colleague – which one the one in the mask or the one smiling (not all risks are dressed as such!)
5. Online system issue with identity theft
6. Unattended computer – but it does have a screen saver
7. Unattended papers

[NB It is not a very sophisticated picture and it would be possible to work on it to improve it]

Slide 15: Data Security and Storage

Securing and protecting data is a key part of the research management process.

The ICO gives some tips.

Slide 16: Data Security and Storage

Data must be managed in different locations/contexts, e.g. at home, in the office (and this can be at your university, and / or health organisation), in the field (in the UK or in other countries). The level of security in these setting will be variable. Can you lock things away at home? Are you the only person using your computer at home? Is it id/password controlled? Is your laptop id / password controlled? Do you carry personal data on your laptop / data USB sticks? Is the data encrypted? By definition, in the field you have personal data. How can you protect this? And the data you've collected is precious - and could be very difficult if not impossible to recollect. Can you get it safely home?

Slide 17: Data Security and Storage

Data can also be gathered / held in virtual spaces. When storing information in the 'cloud' or through social media networks then it is important to understand that you cannot necessarily ensure the confidentiality of that data and need to be careful about the ownership of the data - service providers might claim rights - and the security of that data. Data held in the Web may be managed in accordance with another country's legislation e.g. the US does not have data protection legislation and has different legal requirements in terms of ownership and confidentiality.

Slide 18: Collecting Data

Information may be collected and held on a range of devices. When in the field you need only take with you data you will refer to whilst in the field and not all the research data: and you can take anonymised versions of this data. Though, as I noted earlier, as you will be collecting data you will have to have with you personal details of participants and your collected data will contain personal details. Many of these devices are potential risks to your data. So think about can using id / passwords, encryption, creating backups, putting data into the cloud, accessing the university system via 'desktop anywhere' facilities to download your data, checking that the computer facilities you use are secure, etc.

Slide 19: Storing Data Within the University

This slide shows some of the facilities / systems offered by Northumbria University. Research data should be stored in accordance with university guidance, e.g. a key Northumbria University recommendation is that data is stored on your network drive (provided by the University), which is protected and backed up.

Slide 20: Home Access

Many universities enable you to log in and continue to use the appropriate university systems from home. This means that some protections are in place, even when the information is being accessed remotely.

In some instances there will be a functional benefit to using a host (e.g. Web 2.0) with a lower level of security assurance. However you must be clear about the impact and ensure that anyone participating in the research through these mechanisms is also aware of the data risks.

Slide 21: Northumbria University Guidance

This links to the existing guidance at Northumbria University. NB This is updated regularly and therefore any changes should be checked.

Slide 22: Sending Emails

Too often research data is sent as attachments within an email. Emails are not a secure medium. Encryption can provide some protection for emails.

Slide 23: Physical Security

Good procedures and simple techniques like locking things up are in your control. Find out what the 'encryption' possibilities are in the software you use, or provided by your IT department. Universities provide confidential shredding, or you can buy your own shredder quite cheaply. Software can be purchased to overwrite data on memory sticks.

Summary: Managing your data will allow you to access and use it effectively over time.

FILE MANAGEMENT AND METADATA (Slides 24 - 30)

These slides detail some of the key conventions for naming and managing files ensuring their context is maintained.

Slide 24: Organise your Documents and Files

Organise your files / documents into folders. If you think about how you might organise paper documents, then the same approach can be used to create folders on your computer. So think about your paper consent forms: individual forms are put into a cardboard folder named consent forms, this is placed into a hanging pocket, labelled consent forms, in a drawer in a filing cabinet, the hanging pockets in the drawer are in A/Z order of their labels, you might have one drawer of the filing cabinet for admin documents, another for data, and so on. This hierarchical structure can be created in your computer. You are in effect creating a fileplan: "A logical and systematic arrangement of files into subject groups or other categories based on some definite scheme of natural relationships"^{xi}

Paper and electronic filing systems should be compatible and use the same naming conventions. Keep a record of your 'fileplan', e.g. in a Word document or Excel spreadsheet.

Slide 25: Folders and Documents

In the continuum from raw to published data, your data will undergo many processes. You need to keep copies of the previous form of the data, e.g. in separate folders, and have an audit trail from one form of the data to the next. Keep a record of the audit trail of the different types of documents and types of data along the continuum. This record could be a Word document, an Excel spreadsheet, an Access database, depending on its complexity.

Slide 26: File Names

As well as organising your documents and data into folders, you also need to give your files names: names that are helpful to you in organising your material, and later in finding what you need efficiently! I'll give you the links to a number of resources which provide tips for creating file names. Here are some examples of the points to consider:

Keep file names short but meaningful

- Where file path names are lengthy problems can be caused with automated programmes run to back up data.

Avoid special characters, e.g. \/:[]

- They have special meanings to computer programs

Underscore spaces _ or %20 may replace it

- If you don't like underscores, then you can demarcate the separate parts of the name by capital letters

Reverse dates so they sort usefully YYYYMMDD e.g. 20111805

Ensure document version control, e.g. different versions V01, V02; different drafts of a given version V0.1, V0.2

Slide 27: File Names

A filename can be a code, which gives you some of the audit trail of the data, and could help in finding specific bits of data.^{xii} Explain the system as on the slide. However you would need to keep a key to this code for it to have meaning.

Compare it with a simple name: this tells you it was an interview with the participant with the ID 02. You would need a further record somewhere that listed all the interviews undertaken: the participant IDs, dates undertaken and so on. But you would need to keep such a record with the previous naming version too.

Choose a naming convention that is helpful to you, keep a record of this convention, and be consistent.

Slides 28 - 29: Resources

These are some links to filing guidelines.

Slide 30: Metadata

Metadata is "Data about data"

So for an interview:

- metadata about the interview
 - date
 - time
 - duration
 - location
 - participant ID
 - interview schedule used
 - check that consent forms signed
- metadata about the file containing the data
 - author
 - date
 - format
 - version

and so on

Microsoft allows you to add some basic properties to a document.

Use customisation options e.g. Microsoft document templates

You can create your own document templates, e.g. for an interview transcript you could set up a template with a section at the top that asked you to input all the required metadata.

Utilise tools to add metadata to files.

A lot of metadata will need to be kept in various form of project documentation.

FILE MANAGEMENT EXERCISE (Slide 31)

The purpose of this exercise is to enforce the rules on file management in the context of a research scenario. The research scenario was developed with the assistance of a social work team. The research vocabulary is based on this one team and will differ in other authorities.

The exercise is broken into three questions:

1. What research data and documentation would you expect to exist, including raw data and analysed data? Create a list.

2. Taking this list of data can you develop a list of potential file names?
3. Can you structure the file names into a file plan with folder names?

Question 1 (15 minutes). Using page 1 of the handout, give the students ~7 minutes to identify / detail what research data and documentation one would expect to exist, including raw data and analysed data. Then in open discussion pull out possible answers, and then look at page 2 of the handout.

Questions 2 and 3 (20 minutes) - in open discussion, write up answers on a whiteboard or flipchart. Spend ~5 minutes on file names, the rest on the file plan. Then look at page 3 of the handout.

The case under discussion raises the opportunity to further discuss consent. Any project with potentially vulnerable children / young people and adults would require additional consent and ethical considerations. This case demonstrates the complexities of processes of consent. For example, depending upon the age of the child you would require parental consent forms plus children's assent forms. This particular case is set in England. If the research were in another country this would alter the consent paperwork, e.g. the law in Scotland relating to children's and adult's capacity to consent differs. The key to consent is the legal status and capacity of each individual in the research context. In England, family representatives would only be interviewed for adult clients if the client consented or the representative had Lasting Power of Attorney. In addition foster parents can only comment on their perspective as a foster parent and not necessarily on individual children they are fostering.

ROUNDUP / WHAT NEXT? (SLIDE 32)

For ~7 minutes ask the students to share any tips / good practice they had for RDM, or to raise any queries / make any comments.

Then outline their directed learning tasks:

- Reflect on your strategies and procedures for research data management, e.g.
 - Do you need to take any further actions to ensure your research data is appropriately anonymised and capable of being shared in the future, if appropriate?
 - Do you need to revise your file / document naming system, and / or the file structure for storing your research on your university drive
 - Are the ways in which you physically store your data sufficiently safe and secure?
- Embed research data management into your PhD work activities, e.g.
 - Make an entry in your research diary / portfolio about RDM strategy and actions going forward
 - Make an entry in your training plan about RDM
 - Meet with your supervisory team to report on RDM learning and plans going forward

ⁱ UK Data Archive Model Consent Form <http://www.data-archive.ac.uk/create-manage/consent-ethics/consent?index=3>

- ⁱⁱ University of Edinburgh Records Management Section, Standard Naming Conventions For Electronic Records: The Rules,
<http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/RMprojects/PP/FileNameRules/Rules.htm>
- ⁱⁱⁱ Northumbria University, Regulations for the Degrees of Master of Philosophy (MPhil) & Doctor of Philosophy (PhD), 2011, http://www.northumbria.ac.uk/static/5007/graduateschool/regs_mphilphd11.pdf
“The copyright of the submission rests with the student.” (Section 11.2)
“As a condition of enrolment, the student agrees that the University shall own any Intellectual Property (IP) that may result from his/her research activity and in return the student is eligible for a share of the revenue generated under the same procedures as members of academic staff ... In the case where a student is funded wholly or partly by a third party external to the University, the supervisor(s) and the student should clarify intellectual property ownership issues with the third party at the outset of the research project.” (Section 11.3)
- ^{iv} UKDA, Consent / Consent forms, <http://www.data-archive.ac.uk/create-manage/consent-ethics/consent?index=3>
- ^v UKDA, Transcription, <http://www.data-archive.ac.uk/create-manage/format/transcription>
- ^{vi} UKDA, Anonymisation, <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation>
- ^{vii} Jisc, Freedom of Information and research data: Questions and answers,
<http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx>
- ^{viii} RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith *The Lancet*, Vol 351 No 9103 28 Feb 1998 p 637-41; see <http://www.guardian.co.uk/science/blog/2011/jan/12/andrew-wakefield-fraud-mmr-autism>, Brian Deer investigative journalist.
- ^{ix} Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003, http://www.esds.ac.uk/doc/5407%5Cmrdoc%5CUKDA%5CUKDA%5CStudy_5407_Information.htm
- ^x UKDA, Transcription, <http://www.data-archive.ac.uk/create-manage/format/transcription>
- ^{xi} The Cura Consortium, Glossary of Information Management Terms,
<http://www.cura.org.uk/curaconsortium/glossary.html>
- ^{xii} Laditka, S. B., Corwin, S. J., Laditka, J. N., Liu, R., Friedman, D. B., Mathews, A. E. & Wilcox, S. (2009) 'Methods and management of the healthy brain study: a large multisite qualitative research project. ', *Gerontologist*, 49 (S1), pp. S18-S22.
Johnson, B. D., Dunlap, E. & Benoit, E. (2010) 'Organizing "mountains of words" for data analysis, both qualitative and quantitative. ', *Substance Use & Misuse*, 45 (5), pp. 648-670.