
Ethics and data management

TRANSCRIPT

Research Ethics Online Course

1. Ethics of data management.
2. Most research collects data of some kind. It is collected, analysed and stored in some shape or form in most research projects. This presentation focuses on the ethical aspects of data management, the technical issues are well explained for example in the Data Management Guidelines , see link below. This lecture is built around hypothetical PhD student Alex.
3. As part of his PhD research he collects data from people in multiple different ways including interviews, questionnaires. They both present somewhat different challenges for data management as he discovered.
4. As he was starting to write his informed consent procedure, he realised he needed to consider his whole data management plan in advance. His first thought were about his duty of confidentiality and his thinking broadened from there on.
5. In the end he came to realise there were five different aspects on data management he needed to figure out before he could even write his informed consent process! I shall talk you through all of them in turn and they are: firstly about what should he ask in his interviews and questionnaires in other words, what kind of data should he collect. Secondly there are issues related to accessing the data he collects, who and when and why needed to be considered, thirdly he had to consider process of anonymisation of the data, fourthly he needed to make plans for storing the data during the research process and lastly he had to think through plans for the data life-span past his PhD.
6. In his thinking phase he benefitted from the guidance in the Finnish Advisory Board on Research Integrity 2009 Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review. You can find a link to this document below.
7. The core ethical challenge can be viewed as seeking a balance between the need for confidentiality and for the benefits of scientific openness and completeness of data.
8. One of the key considerations relating to privacy, is to decide what data is to be collected. The less identifiable data he collect, the less risk of a privacy breach. The Privacy Act states quite clearly that personal data cannot be collected unnecessarily. Alex needs a good reason for any personal identifier he collects in his data. If his

research can be completed without a particular identifier like name or address, he cannot collect it.

9. Then he continued considering access to data as an aspect closely linked with his duty of confidentiality. To protect the privacy of his subjects, he needed to make arrangements regarding access, who could access particularly the raw data. Restricting who can see original data could be of significant interest to research subjects and his thinking is that this group should be relatively small. He will need access to it and will have access as the person conducting the interviews. It might be a good idea to grant access to his supervisor as well as he may need assistance in the anonymisation of the data. he am also working closely with another PhD student who may use some of this data in a different part of the process and she may benefit from access to raw data as well. In his informed consent procedure he will share this with the informants and make a commitment that the raw data is not available to anyone else.
10. Anonymising the data means that others looking at the data should not be able to trace back to the actual subject, whose data is being included. There are multiple different techniques to anonymise data including removing, recoding or categorising variables, or using pseudonyms. When anonymisation is done well, the risk of subjects being identified based on the data they have provided becomes very small. You will find a link below for detailed descriptions on anonymisation practises.
11. However, even the best techniques applied in the most diligent manner may not always deliver anonymisation. In very small samples, where subjects come from defined cohorts, it may be possible to trace information and data back to the subject. If his interviews include Anna, Tom and Sam and his data will indicate that these three all work for the same company doing a particular task. It is then unlikely he could guarantee that their identity remains anonymous if the data they are sharing is rich and full of direct and indirect identifiers. Because his interview data is very rich and his population is quite small, he decided to discuss these risks with his subjects in advance as part of the informed consent procedure and share with them the process he will adopt to maximise confidentiality to allow them to make a decision if they are ready to risk any possible association made from the data they share.
12. Similarly in very large data sets where diverse data is extensively collected from subjects, it may be possible to piece together accurate identification despite of the best efforts to anonymise the dataset. So while we commit protecting privacy, we need to be aware of the potentials of not being able to maintain confidentiality in all situations and be transparent about these risks
13. As of storing his data, he decided firstly to keep his paper copies in a locked filing cabinet with two keys, he having one and his supervisor the other. He will store the original soft copies in a password secured electronic location. The password would be something only he will know. Lastly, he will need to make sure to consider confidentiality in all his discussions about the data with others and be conscientious not to share aspects that would risk maintaining confidentiality.

-
14. Alex is aware that the European Code of Conduct for Research Integrity notes in its guidelines (1) that data needs to be stored in a secure and accessible form. Similarly the Finnish Guidelines on Responsible Conduct of Research identify 'inadequate record-keeping and storage of results and research data' as disregard for the responsible conduct of research.
 15. The last aspect of his planning is to consider the fate of the data once the project is finished. At the beginning of the project we are not sure what the path ahead will look like and we do our best to plan ahead. As the project progresses we find out new things and opportunities for our data. At the end we may be able to see great potential for using the data further, or it may appear quite exhausted at that point. We never know where research will take us, so having old data to go back to has significant research advantages. This is also a recommendation in the European Code of Conduct for Research Integrity first guideline, which recommends archiving for 5-10 years. In some countries that balance is interpreted differently and data is typically required to be destroyed at the end of each project to avoid any risks to privacy of the informants.
 16. So finding the balance between confidentiality and scientific openness guides you in your data management decision-making. It invites you to apply both types of ethical thinking in your decision-making. A consequentialist approach would focus you to look at the potential harms and benefits of each option - how great would the potential research benefit be in comparison to the potential harm of future confidentiality breach would be an example of a harm benefit evaluation. From a rule based approach the questions would focus you to think on the inherent importance of allowing subjects to make a truly informed decisions about the information they are giving to research. A good plan for managing data is reached when it does not breach any rights or duties and maximises research potential and minimises harm to subjects.