

## **ACP021 Data Governance and Ethics Report**

### **Project objectives**

The overall objective of the proposed project is to provide greater understanding into how adult University Hospitals Leicester (UHL) patients access and use health information, and how this affects their decision making when it comes to treatment. The first stage of the project will involve small samples of participants from five UHL outpatient clinics (urology, gynaecology, cardiology, clinical immunology, and rheumatology) taking part in digital storytelling workshops. These workshops will provide researchers with detailed information about the sources used by patients to access health information, the reasons why they trust these sources over others, and how this impacts their health-related decisions. These workshops also serve a second purpose by informing the researchers on the items to be included in the second part of the project – the online questionnaire. The questionnaire will be distributed to around 6,000 patients from these five outpatient clinics, to observe more accurately how patients access and use health information, and how this differs between ages and genders.

It is intended that participants will also find some value in their participation, particularly in the storytelling workshops. Allowing participants to reflect on their health information seeking behaviours in groups will enable them to improve their ability to find and appraise trusted sources of health information and be better informed in the future. More broadly, the results of the project will inform the National Health Service (NHS) of where patients get their health information, areas in which patients could be better informed, and ways to improve access to trustworthy sources of health information.

### **General principles**

Both parts of this project involve participants disclosing potentially private and sensitive information. Due to the number of participants involved, it is also likely that some participants will be elderly, disabled, or otherwise vulnerable. It is therefore of upmost importance that the data collected from this project is handled safely, ethically, and in a manner compliant with the Data Protection Act (DPA; 2018) principle of integrity and confidentiality.

To achieve this, the project will follow some general guiding principles. First, the researchers will intend to be as transparent as possible with participants about the aims of the project, how the data collected will be stored and processed, who will have access to the data, and so on – complying with the DPA principle of lawfulness, fairness, and transparency. Participants will be encouraged to ask questions about anything they are unsure of and will be informed of their DPA right to withdraw consent – both to taking part in the experiment and to using their data. This will be communicated through participant information sheets, though it is important that researchers communicate this verbally as well as some patients may struggle with the volume and complexity of information contained in participant information sheets (Ennis and Wykes, 2016).

As recommended by a recent Information Commissioner's Office report (Denham, 2017, p.40), the project intends to keep the amount of personal data collected to a minimum throughout. The procedures for achieving this will be discussed in following subsections.

To comply with the DPA requirement of privacy by design, it is important that data protection is considered from the outset. Thus, a Data Management Plan has been appended to this report (Appendix A). This document will be updated when required as the project is ongoing. Loughborough University will take responsibility as the data controller for the project, and will be identified as such to all participants, researchers, and doctors involved. All processing will be performed by Loughborough University researchers, to minimize the number of people who have access to sensitive data obtained. The project is subject to Loughborough University's approved ethical clearance process. The legal basis of the project will be provided by the consent of the participants.

### **Data collection: digital storytelling**

The digital storytelling workshops, and the data collected from them, pose the greatest risks to the privacy of the participants. Participants will be video and audio recorded whilst they discuss potentially sensitive topics, including their behaviours, attitudes, and motivations relating to their health conditions. They will also create short videos which will be shared amongst the project team to be analysed.

The workshops will be delivered to groups of 6-8 participants. This means that participants in each group will be privy to sensitive information about other participants. Therefore, when consenting to take part in the experiment, participants will also be required to sign an agreement not to share anything disclosed by other members outside the session. Participants will also be asked to use pseudonyms from the outset to reduce the risk of reidentification during and after the project (Graham, 2012, p.36). To maintain consistency, participants will be referred to only by these pseudonyms from this point onwards.

Participants will be made aware that the workshops will be recorded and that researchers at Loughborough University will analyse these recordings, as well as the digital storyboards produced, after the session. Participants will be made aware that, though they will use pseudonyms, reidentification could still be possible if a participant's acquaintance saw or heard the voice of a participant in the recordings. This risk should be minimal, however, as only a select number of researchers at Loughborough University will have access to the recordings. Protective measures will also be taken when storing the recordings after project completion, though this will be discussed in more detail later in the report.

The information discussed in the previous three paragraphs will be summarised on a participant information sheet. Participants will be asked sign consent to take part in the workshop, to have their data processed and stored, and to transfer copyright to Loughborough University. Due to the participants consenting to sharing their data before they take part, some participants may disclose information in the session that they do not wish to be used in the project. Therefore, participants will be asked to review the recordings and digital storyboards before they are analysed and remove any information they do not wish to be processed.

#### **Data collection: questionnaire**

Participants' responses in the digital storytelling workshops will form the basis for the items used in the online questionnaire. The Data Management Plan will also be updated before the questionnaire is distributed, to consider any issues which may have arisen by this point in the project. Participants in this part of the project will be asked to rate the extent to which they agree with a set of statements relating to health information-seeking behaviours

on a Likert scale. This mitigates the possibility of participants being reidentified through their responses, as in a sample of around 6,000 participants there is unlikely to be any unique responses to any items.

No direct identifiers, such as name or address, will be collected from the surveys. To achieve this, the clinic manager for each of the outpatient clinics will be asked to disseminate the surveys. Two indirect identifiers, age range and gender, will be collected in the questionnaire, as age and gender differences have been found in health information behaviours (Ybarra and Suman, 2008). Age range, rather than age, will be collected to reduce granularity and thus reduce risk of reidentification (Graham, 2012, p.51) – whilst still providing sufficient information to be analysed. Again, with the number of expected participants in mind, this should mitigate the possibility of reidentification here.

Similarly to the storytelling workshops, participants will be given a participant information sheet before starting the questionnaire. Participants will be informed of the purpose of the study, the steps that have been taken to protect their privacy, how their data will be stored, and so on. Participants will ask to confirm their consent to the questionnaire and to the intended use of their data. Participants will also be informed of their right to withdraw consent from the questionnaire and the use of their data.

### **Data storage**

After the project has been completed, it will be stored in a way which complies with FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles (Wilkinson et al., 2016). This is to meet the DPA requirement of data portability. The data from both parts of the project will be stored in the UK Data Service (UKDS), making it findable and accessible. Participants will be made aware of this. The data will be made interoperable and reusable by using common generic file formats, such as *.csv*, *.mp4*, *.txt*, and *.tif* (UK Data Service, 2020a).

The questionnaire data will be stored in the open level of access (UK Data Service, 2020b), to make it as accessible as possible for future researchers. Reidentification should be practically impossible with this data, so there is no reason for it not to be made freely available.

The storytelling workshop data, however, needs to be stored much more securely, so will be stored under the controlled access level (UK Data Service, 2020b, 2020). Though Pseudonyms were used, other direct identifiers, such as participants' voices will remain in the data. The controlled level of the UKDS will keep this data completely safe, allowing access only to approved researchers, for approved purposes, and with approved methods.

To maximise accordance with FAIR principles, the data from both parts of the project will include metadata, using the Data Document Initiative (DDI; UK Data Service, 2020d). This will contain descriptions of the aims of the project, variables, data files and formats, methodology, and more. The data will also be assigned a unique Digital Object Identifier (DOI; Economic and Social Research Council, 2020) to make it as findable and accessible as possible.

## **Conclusion**

By using a mixed methods design, with storytelling workshops forming the basis of a questionnaire disseminated to many patients, this project should provide some great insight into NHS patients' health information-seeking behaviours. The results of this project could improve public access to and use of health information. Though the project involves the collection, use, and storage of personal and sensitive information, the risks associated with this information have been minimized and mitigated wherever possible. The resulting project is safe, legal, and ethical, and certainly in the interest of the general public.

**Report Word Count: 1663**

**Total Word Count: 3262**

## References

- Brown, S. (2015). *Patient and surgeon decision-making in risk-reducing mastectomy: an ethical and empirical analysis*, [data collection], UK Data Service, Accessed 10 Nov 2020. SN: 851946, <http://doi.org/10.5255/UKDA-SN-851946>
- Copyright, Designs and Patents Act 1988*, c. 48 Available at <https://www.legislation.gov.uk/ukpga/1988/48/contents> (Accessed: 9 November 2020).
- Data Protection Act 2018*, c. 12 Available at <http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted> (Accessed: 9 November 2020).
- Denham, E., 2017. *Big Data, Artificial Intelligence, Machine Learning and Data Protection*. [online] Ico.org.uk. Available at: <<https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>> [Accessed 9 November 2020].
- Economic and Social Research Council, 2020. *Data Citation*. [online] Esrc.ukri.org. Available at: <<https://esrc.ukri.org/funding/guidance-for-grant-holders/data-citation/>> [Accessed 9 November 2020].
- Ennis, L. and Wykes, T., 2016. Sense and readability: participant information sheets for research studies. *The British Journal of Psychiatry*, 208(2), pp.189-194.
- Graham, C., 2012. *Anonymisation: Managing Data Protection Risk Code of Practice*. [online] Ico.org.uk. Available at: <<https://ico.org.uk/media/1061/anonymisation-code.pdf>> [Accessed 9 November 2020].
- UK Data Service, 2020a. *UK Data Service » Recommended Formats*. [online] Ukdataservice.ac.uk. Available at: <<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>> [Accessed 9 November 2020].

UK Data Service, 2020b. *UK Data Service » Access Control*. [online] Ukdataservice.ac.uk. Available at: <<https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control.aspx>> [Accessed 9 November 2020].

UK Data Service, 2020c. *UK Data Service » Five Safes*. [online] Ukdataservice.ac.uk. Available at: <<https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control/five-safes.aspx>> [Accessed 9 November 2020].

UK Data Service, 2020d. *UK Data Service » Catalogue Metadata. Catalogue, Documenting, Data, Subset, Documentation*. [online] Ukdataservice.ac.uk. Available at: <<https://www.ukdataservice.ac.uk/manage-data/document/metadata.aspx>> [Accessed 9 November 2020].

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), pp.1-9.

Ybarra, M. and Suman, M., 2008. Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. *Health education research*, 23(3), pp.512-521.

## Appendices

### Appendix A: Data Management Plan

#### Assessment of existing data

There are no existing available data on health information-seeking behaviours and resultant decision making. The closest related data found was data on decision making relating to mastectomies (Brown, 2015), and the proposed project will be much wider-reaching and will involve a much larger sample of participants. The mastectomy data is also safeguarded, whereas the bulk of the data from the proposed project will be open access.

#### Information on new data

The proposed project will address a gap in the knowledge of how NHS patients access and use health information when making health-related decisions. It will do this in two parts.

First, digital storytelling workshops with groups of 6-8 participants, all of whom are patients at University Hospitals Leicester, will produce short videos where patients talk about their experiences finding health information and how this affects their decision-making. The data produced from these workshops will be:

- The video for each participant's digital story (.mp4)
- The script for each participant's digital story (.txt)
- The storyboard for each participant's digital story (.tif)
- Video recordings of the workshops (.mp4)
- Researchers' field notes (.txt)

These digital storytelling workshops will form the basis for a questionnaire, which will be distributed to around 6,000 participants, all of whom are patients at UHL patients. Because the digital storytelling workshops have not yet taken place, the exact items included is unknown. Still, an example of a potential item could be *"I generally consult more than one source of information when I am making decisions about my health."* Participants will indicate the extent to which they agree with each item on a Likert scale. Participants will also be asked to state their age range and gender. The responses to the questionnaires will be saved in a .csv file. Incomplete responses will be removed from the final dataset.



No unnecessary data, such as names or addresses, will be collected at any point in the project, and so the data collected will comply with the DPA principle of data minimisation.

### **Quality assurance of data**

By using a mixed-methods design, where the qualitative data informs the quantitative data, it is hoped that the questionnaire will be well-equipped to accurately measure patients' attitudes to seeking and using health information. For the storytelling workshops, pseudonyms will be used consistently from the start of the workshops. Audio recordings will be used to type up transcripts so that no information is lost in digitisation. The questionnaire will be conducted online using Qualtrics, which allows for responses to be downloaded in widely compatible .csv format. Questionnaire responses will be checked, and responses will be removed if incomplete or if likely to be erroneous – for example providing the same response to every item.

### **Backup and security of data**

Whilst the project is ongoing, all data and files relating to the project will be encrypted and stored on password protected computers, accessible only to researchers working on the project. Backups will be stored in encrypted, password protected hard drives which are prohibited from leaving the room they are stored in.

After completion of the project, the data will be stored in the UK Data Service (UKDS). Once transferred to the UKDS, all local copies of the data will be deleted from the Loughborough University computers and hard drives. These measures are compliant with the DPA principle of storage limitation. The storytelling workshop data will be stored under the UKDS controlled access level, as it will contain potentially identifying information. The questionnaire contains no such identifying information, and so will be stored at the UKDS open access level, for maximal future use and accessibility.

### **Ethical issues**

The research participants will all be patients at University Hospitals Leicester. Due to this, and the number of participants, it is likely that this will include people belonging to vulnerable groups. Patients will also be talking about their health, which is a private and

sensitive topic. Responses to the questionnaire will be fully anonymised, however data from the storytelling workshops will necessarily contain identifying information.

Therefore, researchers will be instructed to be as transparent as possible with participants and make sure they are fully informed before they take part in either part of the project. Participant information sheets will be given to participants before they engage in the workshops and questionnaires. These information sheets will inform participants of:

- The aims of the project overall
- The aims of the current part of the project
- The experimental procedure
- How their data will be used and stored
- Whom the data controller is and their contact details
- Copyright information
- The legal basis for the project
- The fact that the data will be stored indefinitely with the UKDS
- Participants' right to withdraw consent from the experiment
- Participants' right to withdraw consent to their data being used and stored

It is important that this information is also communicated verbally to participants, to ensure they can give informed consent. Participants should be also given the opportunity to ask questions throughout the project, maximising the accordance of the project with the DPA principle of lawfulness, fairness, and transparency.

The legal basis for this project relies on the informed consent of the participants, and so participants must sign to give consent to both their participation in the project and to the use of their data before starting the storytelling workshops and the questionnaire. Participants in the storytelling workshops will also be asked to sign a confidentiality agreement so that any sensitive information shared by other participants is not shared with people outside of the project, ensuring that the DPA principle of integrity and confidentiality is met.

Because participants in the storytelling workshops will give consent before they take part, some participants may disclose information in the session that they do not wish to be

used in the project. Therefore, participants will be asked to review the recordings and digital storyboards before they are analysed and remove any information they do not wish to be processed.

### **Management and curation of data**

When transferred to the UKDS, the data will be stored in a way which complies with FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles (Wilkinson et al., 2016). To achieve this, data will be stored in widely compatible formats from the start, such as *.csv*, *.txt*, *.mp4*, and *.tif*, to ensure that the data remains usable over time and across different platforms. Rich metadata will also be included, and the data will be assigned a Data Object Identifier (DOI).

### **Difficulties in data sharing and measures to overcome these**

The data will be stored on the UKDS, to allow future researchers to use it. The data will be stored according to FAIR principles, as discussed above, which should make sharing data as simple as possible.

The questionnaire data will be stored in the UKDS open access level. Due to the workshop data not being fully anonymised, this data will be stored at the controlled access level. This will make it harder for future researchers to access this part of the project, though it will still be accessible for researchers who can be trusted not to use the data improperly - it will just take some time to be granted access.

### **Consent, anonymisation, and strategies to enable further re-use of data**

Much of this has been covered already. To summarise, questionnaire data will be fully anonymised, while the workshop data will be pseudonymised and will still contain some identifiers such as participants' voices. Participants will be informed of the project and their rights both verbally and through a participant information sheet. They will then give informed consent to both taking part in the experiment and to the use of their data.

Data will be stored on the UKDS, and will be stored according to FAIR principles, making it maximally reusable for future researchers. The questionnaire data will be stored under open access, while the workshop data will be stored under controlled access due to it containing sensitive information and potential identifiers.

## **Copyright and intellectual property ownership**

According to the Copyright, Designs, and Patents Act (1988), the creators of a work automatically own copyright. This means that participants own copyright to some of the data produced, such as any videos and the words in the transcript. This could cause problems for future reuse of the data, and so participants will be asked to transfer copyright to Loughborough University at the same time that they sign the consent forms.

When creating the digital storyboards, it is important that the participants do not use copyrighted media, such as music or images, in their storyboards. To prevent this, researchers will direct participants to sources of royalty-free media they are able to use.

## **Responsibilities**

The data controller will be Loughborough University. Loughborough University will be responsible for all collection, use, and storage of the data for the project. Loughborough University will also be responsible for ensuring the rights of participants are protected and for transferring the data to the UKDS in a suitable format.

Participants can exercise their right to withdraw from the project by notifying a researcher whenever they wish to withdraw. During the workshops, they may withdraw by talking to a researcher, who will immediately allow the participant to leave and destroy any data gathered so far. During the questionnaire, participants may withdraw by exiting the online questionnaire before it is completed – any incomplete responses will be removed from the dataset. After either part of the project, participants exercise their lawful right to object to their data being used in the project, and to erasure of their data (DPA, 2018), by contacting the university by telephone, email, or letter.

**Data Management Plan Word Count: 1599**

**Total Word Count: 3262**