



Daniel Turner
Data Management Plan

Precious:
Reclaiming Value From Personal Data

DATA COLLECTION

01 Data that will be collected

I will be collecting peoples personal Facebook data. This data comprises of everything that Facebook stores in relation to individual account. Users are able to request a downloadable package that contains most of the data they have stored or created via Facebook. Examples of the type of data are below:

- **About me** - Information you added to the **About** section of your Timeline like relationships, work, education, where you live and more. It includes any updates or changes you made in the past and what is currently in the **About** section of your Timeline.
- **Ads Clicked** - Dates, times and titles of ads clicked (limited retention period).
- **Address** - Your current address or any past addresses you had on your account.
- **Chat** - A history of the conversations you've had on Facebook Chat (a complete history is available directly from your messages inbox).
- **Credit Cards** - Facebook your credit card number (this may depend on account settings)
- **Facial Recognition Data** - A unique number based on a comparison of the photos you're tagged in. Facebook uses this data to help others tag you in photos.
- **Followers/Friend Requests/Friends** - A list of people who follow you / Pending sent and received friend requests / A list of your friends
- **IP Addresses** - A list of IP addresses where you've logged into your Facebook account.
- **Logins** - IP address, date and time associated with logins to your Facebook account
- **Messages** - Messages you've sent and received on Facebook
- **Phone Numbers** - Mobile phone numbers you've added to your account, including verified mobile numbers you've added for security purposes.
- **Photos / Photo Metadata** - Photos you've uploaded to your account / Any metadata that is transmitted with your uploaded photos
- **Posts by You** - Anything you posted to your own Timeline, like photos, videos and status updates.
- **Posts by Others** - Anything posted to your Timeline by someone else, like wall posts or links shared on your Timeline by friends
- **Videos** - Videos you've posted to your Timeline

(All information and full list sourced from bigbrotherwatch.org.uk)

For a full list of the data contained in the package, along with items I will exclude from collection please see **Appendix 1: Facebook Data List.**

The data is downloaded in .zip format but when extracted resolves to numerous folders with mostly .html and .txt files or JSON and .txt files (depending on download options selected). There are obvious exceptions for uploaded media such as photographs, videos, audio etc. which will be in standard .jpg, .png, .mp4, .avi, .mp3 formats.

The combined file size of each individuals data could vary widely, depending on amount of usage, type of usage (eg. lots of photographs videos etc.) but the average user package size is approximately 500MB. It is assumed that approximately 5-10 participants will be recruited resulting in a total data set of approximately 2.5-5 GB. Data will be accessed via secure LJMU OneDrive so will not need to be transferred anywhere from this storage medium.

These initial data-sets are gathered as a means of testing the software programs developed to ensure they work effectively for various individuals. The participants providing their data will also be engaged in an ongoing dialogue with the project; participating in discussion groups, reviewing outcomes and suggesting possible approaches to developing outputs from their personal data.

02 Data that will be created

The following table briefly outlines data that will be created during this project. Each of them will be discussed in more detail below.

Ref	Data	Description
A	Artworks, visualisations, code and experiments	Data generated via iterative and final work
B	Questionnaires, Discussion Groups, Public Engagement	Data generated by participants in response to project
C	Reflective journal	Data relating to project development and reflection

A - Artworks, visualisations, code and experiments

The information contained within these data sets will be used to create visual output in a number of possible formats (e.g. .tiff, .pdf, .svg) which in turn may be processed via proprietary software such as Adobe Photoshop (.psd) or Adobe Illustrator (.ai).

The code that processes and synthesises data into visual outputs will be created in Processing which stores files in a .pde format. It is anticipated that these code files (minus all data sets) will also be made freely available online via the GitHub repository along with instructions for usage.

It is hoped that people will be able to use these programs to create artwork from their own personal Facebook data even if they are not directly participating in the study.

Processing (which is based on Java) has been used as the programming language/IDE for a number of reasons:

- It is free to download and open source
- Has a large, inclusive, active community with a great deal of support
- Designed as "a flexible software sketchbook and a language for learning how to code within the context of the visual arts"
- Has numerous additional libraries that extend the core software
- Is a language I have a great deal of experience in developing creative work with
- Can be run locally on machines requiring no internet or network connection, enabling safe access of sensitive data.

Experiments and iterative versions of software and outputs will be stored in separate folders on OneDrive to the original data sets.

Once artworks are completed they will be rendered into a physical format via methods such as printing, plotting, embroidery and screenprint etc. Whilst it is expected that visual outputs will be abstracted versions of (elements of) this data, they will still be representative of it in some way and thus have the information encoded within them.

NB. This is a central consideration of project. i.e. how happy are we to provide personal information in whatever from and display it to people*

Once the project is completed, original data sets can be deleted.

B - Questionnaires, Discussion Groups, Public Engagement

Questionnaire results will be collated and stored on the LJMU OneDrive along with recordings and notes from discussion groups.

Any public engagement (e.g. people utilising programs and providing examples of output/feedback who are not directly involved in the study) will also be collated and stored on the LJMU OneDrive.

C - Reflective journal (inc. analysis of feedback, technical + conceptual notes etc.)

The reflective journal will be where ongoing documentation of, planning for and reflection on the project is recorded. Alongside physical sketchbooks, a digital, interactive journal will be utilized to record these details. As well as informing project development the reflective journal will provide the basis for the exegesis.

The digital aspects of this journal will be created on Microsoft Teams as this is a secure, private system that can link directly to OneDrive. The ability to incorporate various apps into this platform also offers the ability to combine the various forms of media that will be required for this project.

03 How data will be collected or created

The following table gives a brief overview of how each type of data will be collected / created. More detailed explanations are outlined below:

Ref	Data	Collection/Creation	Storage
A	Personal data sets of participants	Encrypted HDD	OneDrive – Chunked, pseudonymised
B	Intermediary visual output and experimental design work	LJMU Computer using Processing	Encrypted HDD/OneDrive
C	Finished work	Various physical outputs	NA
D	Questionnaires/Discussion Groups / Feedback forms	Digital forms/Teams/Audio Recording	OneDrive
E	Reflective Journal	Digital Document	OneDrive

A - Personal data sets of participants

Participants will be given instructions on how to download and secure personal data from Facebook on encrypted, external hard drives (HDD). The extraneous data (identified in Appendix 01) will then be deleted and the remaining data transferred to the LJMU OneDrive via an LJMU computer connected to the LJMU network.

The HDD will then be returned to the participant along with instructions for how to alter the passcode. Dates and times of transfers will be logged (via pseudonymised system).

Individual datasets will be “chunked” and stored in three, pseudonymised folders on the OneDrive and do not need to be moved from here at any point during the study.

Participants will also be given guidance on how to fully remove personal information from the encrypted HDD (i.e. formatting) and how to store/access it safely depending on their preference.

B - Intermediary visual output and experimental design work

All intermediary visual output will be created from software run on a LJMU protected computer that accesses securely stored data sets on OneDrive where necessary. It is anticipated that in the early stages of development and testing (e.g. where data-parsing functionality is being explored, ideas are being initially developed) that my personal Facebook data (also stored on OneDrive) will be used.

Visual outputs that are results of participants datasets will be initially saved to an encrypted HDD. At the end of each experiment session these will be uploaded to the OneDrive and the HDD will be formatted.

It is anticipated that approximately 4-5 experiment and development cycles will be completed throughout the "experimentation" stage of the study. Each of these cycles will have their own folder on the OneDrive. Within these folders each software iteration and outcomes will be stored in within sequentially numbered folders (e.g. Experiment_01_0001).

Due to the nature of Processing files they automatically generate a folder that matches the .pde file name. The .pde file is automatically saved in here and all outputs can be directed to save in a sub folder within. Naming restrictions on the filetypes do not allow for spaces in filenames, hence the underscores.

Output filenames (for bitmap and vector files etc.) will be auto-generated within the software to contain relevant references and date creation:

- software_filename - Dataset ref - year_month_day_hourMinute.extension = **Experiment_01_0001 - A - 2020_12_08_1655.tif**

Any variable settings and parameters utilised with the program (e.g. random seed, noise, transformation settings etc.) will be exported in .txt format along with each image.

Participants will sign consent for any printed/physical intermediary outputs to be created from the generated works. These works will be referenced with the matching filenames from the output so data source, program, parameters etc. can all be traced back via the pseudonymised referencing system. Importantly, this system will also allow for tracing and removal of any work should someone choose to withdraw consent.

C - Finished work

At present it is anticipated that the finished works will be individual, non-reproducible, physical artworks that will explore various themes within the data sets. They will be outputs generated from individual data sets and referenced via a number system. This number will be linked on a document in a similar fashion to that detailed above. As the programs themselves are the 'artworks' this referencing system is anticipated only to be used for verifying/analysing/discussing results within the exegesis and with the individual participant who owns the data. As detailed above, it will also create a tracing system should anyone withdraw consent.

Consent will be gained from individual participants for displaying any interim or final work in its original or reproduced form (e.g. photograph).

D - Questionnaires/Discussion Groups / Feedback forms

It is not anticipated that original questionnaires will contain any personal or sensitive information in relation to individual data. They are simply a means of gauging peoples understanding of data usage, practice, value and their current level of interaction with it/attitudes towards it. Personal information such as name/age/contact details etc. will be pseudonymised and answers will be stored on the LJMU OneDrive.

Feedback forms from participants at the experiment stage will be stored in the same manner, referenced against pseudonymised system.

Facilitation for discussion and interaction with and between participants during the experimentation stage is expected to take two forms:

1. Ongoing, online, private discussion group on Microsoft Teams where they can post messages, images of outcomes, questions etc. This platform allows for group discussion and private conversations alongside the ability to upload various forms of media.
2. In person / virtual (e.g. Teams) discussion groups at culmination of each experiment

Content added to Teams and recorded meetings are automatically stored on OneDrive and, consequently, backed up frequently.

Audio recordings of staged discussion groups and their transcriptions will be stored in folders on OneDrive organised by date

E – Reflective Journal

References to individuals, artworks and relevant discussion extracts will utilise the naming and pseudonymising conventions outlined above.

DOCUMENTATION AND METADATA

01 Documentation and metadata to accompany the data

As the finished product for this project is a set of software programs, there is no output data as such. Metadata will be useful in relation to developing the programs, analysing results and for project management but is not necessary for publication. The use of metadata in this context is outlined in the previous section.

Documentation will be created to cover various aspects of the project, however, examples of which are outlined below:

A – Supporting Documentation for Participants

- Guide - Safely download, store and access personal Facebook Data
- Guide – Using Microsoft Teams for this study
- Guide/s – Short guide describing how to use each software experiment developed
- Documentation to be stored/distributed on private Microsoft Teams group

B – Final Programs

- Final programs will have documentation on usage.
- Accompanying guidance for production methods where relevant (e.g. screen printing, plotting, embroidery etc.) so people can follow full process if they want
- Documentation can be stored/distributed on GitHub alongside the software.

C - Exegesis

ETHICS AND LEGAL COMPLIANCE

01 Managing ethical issues

The main ethical issues with this project relate to the collection, storage and re-encoding of participants personal data. Some explanation of this is already detailed above but the main points are collated and outlined below:

Full Ethical Review

The project will be going through a full ethical review at LJMU.

Exclusion of Certain Data

As outlined in **Appendix 1** (Facebook Data List), certain aspects of Personal Facebook data (e.g. credit card details) will not be collected at the initial phase.

Detailed consent

Consent for storage and use of participants personal data in developing software to be gained at outset of study. Consent will also be gained to show interim work to people partaking in the study and supervisory team. Also to have work included in the project documentation.

A detailed consent form will be developed allowing people to choose exactly what elements of the project they are happy to participate in. For example participants may choose to have their data used for program development but no images of work developed from it to be published in any form.

Importantly, the consent form will also require participants to gain consent for any identifiable third-party data (e.g. photograph or comment of a friend etc.) that is used in an artwork.

Storage

All personal data to be chunked and stored on OneDrive.

Pseudonymisation key to be stored on separate OneDrive account.

Personal data sets to be deleted immediately at the end of the study.

Disclaimers on Released Software

Disclaimers included in software samples hosted on Github explaining that whilst people are able to generate visual outputs using personal Facebook data this is essentially a re-encoding of this personal data in a visual format. They share this information at their own risk.

02 Management of copyright and Intellectual Property Rights (IPR) issues

Exegesis

Copyright/IPR belongs to author (i.e. myself)

Programs and Code

Released under Creative Commons license (e.g. MIT X11)

Final Sample Artwork

Copyright/IPR belongs to artist (i.e. myself)

Participant Created Artwork

Copyright/IPR belongs to the participant

STORAGE AND BACKUP

01 Data storage and back up during research

Data will be stored on the LJMU OneDrive:

- 1TB capacity which should be ample for both source and generated data.
- Automatically backed up on a daily basis.
- Personal data sets will be transported from people's personal computers to OneDrive (accessed via LJMU network) on an encrypted hard drive.

Programs suitable for release (iterative or final) will also be stored on GitHub (<https://github.com/>)

- Accessible to anyone
- Backed up daily with versioning and forking recorded

02 Management of access and security

One Drive

Password changed every six weeks

Personal Data Sets

Access restricted solely to the lead investigator (Daniel Turner)

Teams

- Private Team setup for participants and supervisors.
- Separate, private channel for lead investigator and supervisors.
- Separate, private chats for individual participants.
- All relevant project information (e.g. consent forms, timelines, documentation etc) available to participants via Teams at any time.
- Iterative work and discussions released to Teams as required.

SELECTION AND PRESERVATION

01 Data of long-term value that should be retained, shared, and/or preserved

Personal Data

All participants personal data used for development will be deleted immediately on completion of project.

Iterative Visualisations

These visualisations are important in evidencing the development of the project and illustrating outcomes for programs created. Their value lies in representing conceptual underpinning of the work and as contextual reference for any further development. As such:

- Visualisations of data that are used in exegesis to be retained.
- Visualisations of data that have been authorised for sharing by participants to be retained.
- Participants to be offered digital copies of any work made using their data.

Final Sample Artwork

As with above, the value of these pieces is in providing evidence of outcomes and contextualizing project. Alongside this, each output is expected to have personal value to the person whose data it was created from. As such:

- Photographic evidence of exhibition retained
- Study copies to be retained
- Copies produced for participants given to them to do with what they wish

02 Long-term preservation plan for the dataset?

The plan is for iterative designs, output examples and findings to be curated as the study progresses. There are a number of experimentation and design cycles identified in the project timeline that have this curation process built in. Relevant work will be retained, catalogued and (in some cases) utilised for reference in the exegesis. Extraneous work will be removed at these points.

The final dataset will actually consist of a set of programs and accompanying documentation for usage that are to be made freely available. In this context:

- Backup of Final Programs to be retained on physical HDD and stored on GitHub for free access to general public.

Depending on features utilised in programs (developed in Processing) certain elements of code may need updating to work with new versions of software. If hosted on GitHub, these updates can be done by myself or the community. It is also hoped that people will create different versions of these programs independently as they will be able to 'fork' the originals, adapt and upload new versions.

DATA SHARING

01 How data will be shared

Personal datasets of participants

Deleted, so not shared

Programs and Code

Available on GitHub along with documentation and instructions for use

Physical Outputs

Photographic documentation available online and in exegesis

Exegesis

Published online

02 Required restrictions on data sharing

All referencing in exegesis to be pseudonymised.

RESPONSIBILITIES AND RESOURCES

01 Responsibility for data management

As the lead investigator I will be responsible for the data management of all assets collected or created throughout this project.

Participants will be responsible for copies of datasets retained by themselves and any artwork generated from these datasets on their own computers. They will be given guidance on how to store and access their own datasets safely at the outset of the study along with guidelines for sharing information at various stages (e.g. group discussions, upload of artworks etc.).

In the case of a data breach, the lead supervisor will immediately contact their supervisor and the LJMU Data Protection Officer (DPO@ljmu.ac.uk, 0151 904 6134)

01 Resources required to deliver your plan

Delivery of the DMP outlined here will require:

2 x Private LJMU OneDrive account:

- Account A - storage of data and outputs
- Account B - storage of pseudonymisation key

1 x Private Microsoft Teams 'Team' – For collaborative work

1 x GitHub account – For storage of externally accessible programs/code

1 x Encrypted HDD - For transfer of personal datasets

APPENDIX 1: FACEBOOK DATA LIST

In the Archive	Description	Excluded
About Me	Information you added to the About section of your Timeline like relationships, work, education, where you live and more. It includes any updates or changes you made in the past and what is currently in the About section of your Timeline.	
Account Status History	The dates when your account was reactivated, deactivated, disabled or deleted.	
Active Sessions	All stored active sessions, including date, time, device, IP address, machine cookie and browser information.	
Ads Clicked	Dates, times and titles of ads clicked (limited retention period).	
Address	Your current address or any past addresses you had on your account.	
Ad Topics	A list of topics that you may be targeted against based on your stated likes, interests and other data you put in your Timeline.	
Alternate Name	Any alternate names you have on your account (ex: a maiden name or a nickname).	
Apps	All of the apps you have added.	
Birthday Visibility	How your birthday appears on your Timeline.	
Chat	A history of the conversations you've had on Facebook Chat (a complete history is available directly from your messages inbox).	
Check-ins	The places you've checked into.	
Credit Cards	If you make purchases on Facebook (ex: in apps) and have given Facebook your credit card number (this may depend on account settings)	x
Currency	Your preferred currency on Facebook. If you use Facebook Payments, this will be used to display prices and charge your credit cards.	
Current City	The city you added to the About section of your Timeline.	
Date of Birth	The date you added to Birthday in the About section of your Timeline.	x
Deleted Friends	People you've removed as friends.	
Education	Any information you added to Education field in the About section of your Timeline.	
Emails	Email addresses added to your account (even those you may have removed).	x

Events	Events you've joined or been invited to.	
Facial Recognition Data	A unique number based on a comparison of the photos you're tagged in. Facebook uses this data to help others tag you in photos.	
Family	Friends you've indicated are family members.	
Favorite Quotes	Information you've added to the Favourite Quotes section of the About section of your Timeline.	
Followers	A list of people who follow you.	
Friend Requests	Pending sent and received friend requests.	
Friends	A list of your friends.	
Gender	The gender you added to the About section of your Timeline.	
Groups	A list of groups you belong to on Facebook.	
Hidden from News Feed	Any friends, apps or pages you've hidden from your News Feed.	
Hometown	The place you added to hometown in the About section of your Timeline.	
IP Addresses	A list of IP addresses where you've logged into your Facebook account.	x
Locale	The language you've selected to use Facebook in.	
Logins	IP address, date and time associated with logins to your Facebook account.	
Logouts	IP address, date and time associated with logouts from your Facebook account.	
Messages	Messages you've sent and received on Facebook.	
Name	The name on your Facebook account.	
Name Changes	Any changes you've made to the original name you used when you signed up for Facebook.	
Networks	Networks (affiliations with schools or workplaces) that you belong to on Facebook.	
Notification Settings	A list of all your notification preferences and whether you have email and text enabled or disabled for each.	
Pages You Admin	A list of pages you admin.	
Pending Friend Requests	Pending sent and received friend requests.	
Phone Numbers	Mobile phone numbers you've added to your account, including verified mobile numbers you've added for security purposes.	x

Photos	Photos you've uploaded to your account.	
Photos Metadata	Any metadata that is transmitted with your uploaded photos.	
Physical Tokens	Badges you've added to your account.	
Pokes	A list of who's poked you and who you've poked. Poke content from our mobile poke app is not included because it's only available for a brief period of time.	
Political Views	Any information you added to Political Views in the About section of Timeline.	
Posts by You	Anything you posted to your own Timeline, like photos, videos and status updates.	
Posts by Others	Anything posted to your Timeline by someone else, like wall posts or links shared on your Timeline by friends.	
Privacy Settings	Your privacy settings.	
Recent Activities	Actions you've taken and interactions you've recently had.	
Registration Date	The date you joined Facebook.	
Religious Views	The current information you added to Religious Views in the About section of your Timeline.	x
Removed Friends	People you've removed as friends.	
Screen Names	The screen names you've added to your account, and the service they're associated with. You can also see if they're hidden or visible on your account.	
Spoken Languages	The languages you added to Spoken Languages in the About section of your Timeline.	
Status Updates	Any status updates you've posted.	
Work	Any current information you've added to Work in the About section of your Timeline.	
Vanity URL	Your Facebook URL (ex: username or vanity for your account).	
Videos	Videos you've posted to your Timeline.	