

# Data Management Plan

## Personally Identifying Information

Where possible, our strong preference is not to collect or receive personally identifiable information (PII) such as name, photographs of faces, home addresses, credit card numbers, date of birth, email addresses, etc. Where such information is held, it is classified as *restricted* or more likely *reserved* under the University's information security procedures. The procedures below are designed to meet the higher *reserved* standard.

We appreciate that the anonymity of data is not limited to consideration of personally identifying information (e.g., name, date of birth, address). Instead we take the view of considering what data a potential adversary might hold and be able to match against our data (???; ???).

## Quality Assurance of Data

The lead author for a paper is responsible for the QA process for the data in that paper, and the data manager for the overall project/dataset is responsible for collating QA information and communicating it to the relevant researchers, or external partners. We will conduct a set of independent replications of key results (e.g., with a simple linear regression in place of a more complicated econometric specification). We will also assess the match between our analysis and publicly available industry or regulator aggregate and summary statistics (e.g., the fraction of people defaulting on their card in any month, etc.). Finally, by linking data sets with public data, we can assess the data quality. For example the prevalence of cash advance charges on credit cards correlates with the fraction of children receiving free school meals in the public domain School Census data, at the level outbound postcode sectors (all but the last two letters).

In addition, we describe specific QA for specific datasets. Financial services in the UK have statutory duties with regard to the collation and maintenance of accurate records under the European Credit Directive and as part of their broader obligations under the Data Protection Act. Thus Barclays and Lloyds have their own internal quality assurance process, as do the credit card providers who have shared data with us in collaboration with the UK Cards Association

and the data aggregator Argus. The police has a statutory duty to report statistics (e.g., recorded crime, stop and search) to the Home Office.

## **Backup and Security of Data**

### **Storage**

Data for the majority of datasets are held on servers physically at Warwick. These servers are fully encrypted at rest, rendering data inaccessible in the event of physical theft. These are further backed up to University of Warwick file stores. All data transferred off-site for backup are encrypted beforehand, preventing data access within the backup system. Backups are maintained under standard University of Warwick information services policies. Storage on removable media (e.g., USB sticks, external hard drives) or laptops is not permitted.

For police data that includes identifying information and that requires level 3 vetting to access, the data will not be stored on a server, but on a single machine. Hard drives on this machine are at rest encrypted, with the sensitive data held on virtual drives with an additional layer of encryption. These virtual drives are only mounted and made accessible when a level three vetted researcher is working on the data. The machine is stored in a physically secure room, and only level 3 vetted individuals have login credentials or encryption passwords.

### **Transfer**

To protect against data interception during transfer, data are only ever transferred over encrypted channels, or, data are encrypted before being placed on an unencrypted channel or device, with PGP keys or separately communicated symmetric keys. Where possible, we prefer to use secure FTP to transfer data. Alternatively, encrypted data can be shared by the warwick.files service. Where use of a physical device is absolutely necessary, data must be encrypted. Sharing of data unencrypted data by email attachment is forbidden. Use of Google Drive, Dropbox, etc., is forbidden.

### **Access**

Data are only accessed by a limited, named set of researchers. Strong passwords are enforced and two factor authentication used. Physical access is restricted by locked office or server room doors. Logins are logged. Password protected lock screens are used on all data access computers, protecting the data while machines are on but the operator is away.

### **Data Provided by Industry Partners**

ESRC Research Data Policy does not specify requirements for handling data provided by 3rd parties (i.e., industry partners). Here we follow the Engineering

and Physical Sciences Research Council (EPSRC) policy framework on research data. Ethical and legal responsibilities are undertaken by the company collecting the data. Data are held under conditions of the specific company agreements. Company specific legal and ethical procedures will be undertaken and conformed to on a case-by-case basis. Additional ethical and legal considerations, that may extend past 3rd party responsibilities, are undertaken on a per dataset basis within the University of Warwick. This is compliant with EPSRC guidelines, specifically: “Research Organisations are not expected to assume responsibility for the preservation and management of third party research data not generated within their own organisation” [clarification of Expectation VII]. When data are used to support research findings (becoming research data under EPSRC definitions) then metadata, as defined by the EPSRC, will be published alongside journal articles in order to support research replicability. Additionally, meeting EPSRC Expectation II, published research papers will include a short statement describing “how and on what terms any supporting research data may be accessed” noting that this data are covered by the clause in the provided clarification document stating that access is restricted due to “compelling legal or ethical reasons [that] exist to protect access to the data”. For confidential and/or sensitive data, metadata will be limited to a researcher contact and a statement indicating that data access was under legal agreement and that “compelling legal or ethical reasons exist to protect access to the data”. UK Data Protection Act compliance responsibilities are undertaken by the company collecting the data.

## **Data Collected by Us**

Any personal data collection will require and be governed by University of Warwick Humanities and Social Sciences Research Ethics Committee, in addition to any other funder requirements. Data will be held for 10 years from date of last use/access by a third party [EPSRC Expectation VII]. Data will not publicly accessible [EPSRC Expectation II, Expectation VI]. Meta-data will be publicly published, including access statement detailing the route to access [EPSRC Expectation V, Expectation VI]. UK Data Protection Act compliance responsibilities are undertaken by University of Warwick.

## **Expected Difficulties in Data Sharing**

We will provide full contact details for data collected by third parties as part of the meta data [EPSRC Expectation V, Expectation VI]. Thus other researchers will be able to request the same or additional extracts. For data collected by us, data will be freely available and will be deposited at the UK Data Archive. We will use standard CSV file formats with accompanying README files detailing the column headings. For confidential and/or sensitive data collected by us, sharing is limited to metadata and a statement indicating that data access was under legal agreement and that “compelling legal or ethical reasons exist to protect access to the data”. Confidential or sensitive data will not be publicly

available [EPSRC Expectation II, Expectation VI].

## **Copyright/Intellectual Property Right**

The data provided by third parties remains the property of the industry partners and these rights are not transferable.

## **Responsibilities**

All researchers on the project will share responsibility for the data. Mullett is the data license holder. Mullett is responsible for data security, processing, quality assurance, archiving, and for providing researchers with the appropriate level of data access given their vetting and research projects.

## **University and Other Relevant Bounding Policies**

Warwick works with confidential and sensitive data every day from a wide range of sources (e.g., credit card records, medical records, police and security services data). As employees of Warwick all researchers are bound by:

- The University's obligations under UK Data Protection Legislation
- Locally by University data protection policies
- The University research ethics committees specialising in the use of human data in the social sciences
- The professional bodies for Psychology and Economics in the UK
- The ethical requirements of the scientific journals in which they publish

Relevant links are:

- All lab members should be aware of University policies and their obligations under those policies
- British Psychological Society Code of Conduct
- University HSSREC
- EPSRC standards