<p align="center"># Safe Data Access Professionals Quarterly Meeting</p>

<p align="center">Wednesday 4th December, The University of Manchester</p>

<p align="center">## Minutes</p>

Attendees (in person)

Richard Welpton (Chair) (*The Health Foundation*), Sarah Al-Adely (*University of Manchester*), Carole Morris (*National Services Scotland*), Simon Parker (*Cancer Research UK),* Martin O'Reilly (*Turing Institute*), James Dainty (*HMRC*), Will Crocombe (*consultant*), Elaine Mackay (*University of Manchester*), Anthony Macken (*Central Statistics Office, Ireland*), Alix Taylor (*UK Data Service*), Valerija Kolbas (*UK Data Service*), Mary McDerby (*University of Manchester*), Emily Griffiths (*University of Manchester),* Rachel Thompson *(University of Manchester)*

Apologies

Aida Sanchez (*UCL*), Beate Lichtwardt (*SSHOC*), Shayla Leib (*ONS*), Olly Butters (*University of Newcastle*).
Telephone unavailable.

Opening

Emily Griffiths opened the meeting. Richard Welpton thanked everybody for attending, and especially thanked Emily Griffiths and Rachel Thompson for organising the local arrangements.

Updates

Cancer Research UK – Simon Parker

- Launched Secure Data Environment, used by externally funded project, more projects coming in next year
- Secure Data Environment is being made available throughout the charity
- Research Data Strategy for the charity is under development
- Training slides to accompany SDC Handbook available (on SDAP website)

UK Data Service – Alix Taylor and Valerija Kolbas

- Team are reviewing older projects in the Secure Lab or those that have gone beyond original purpose. This is to ensure all projects are compliant with the Digital Economy Act
- Need to increase capacity (hardware) and introduce new software
- User experience testing (for data access, including applications access data) to begin in 2020.

National Services Scotland – Carole Morris

- Customisable virtual machines can now be requested by users, though this is not default access yet
- Getting ready for machine learning and AI. Carole to brief again in future meetings.
- Researchers can now request to have their own software installed
- Imaging data now available; working with researchers to develop access to these data as a new service, including how to identify cohorts etc., metadata etc.
- April 2020: joining Public Health Scotland
- Research Data Scotland to be introduced in 2020: this will include a 'front door' for the research community for administrative data access

HMRC Datalab - James Dainty

- Provided an overview of the HMRC Data Lab
- Computers upgraded to high-spec
- Overview of capacity
- New software (Stata 16), pause for IT hardware upgrades will occur in 2020
- 10 new projects have recently been approved
- Discussion about the use of standalone PCs

Central Statistics Office, Ireland – Anthony Macken

- CSO's recently introduced accreditation process for research organisations and their researchers has been renamed as a registration process. The change has been made to ensure that no organisation could use this process as a means to gain a competitive advantage over another organisation.
- For a similar reason to the renamed process for registration, the terms "Research Organisation" has been clearly defined on all RMF application forms.
- Older projects due for renewal: researchers are now being asked not just to explain why they require a project renewal, but in some cases, may also be asked to demonstrate progress on a project before a renewal can be granted.
- Sample sizes: CSO is considering allowing researchers, on a case by case basis to state the size of the sample which they require for their research project. This will apply only to person-based administrative datasets.

University of Manchester (1) - Rachel Thompson and Elaine Mackay

- Spoke about their use of cloud services, proof of concepts. These are in design phase at the moment.
- Allowing external access (to Manchester) researchers due in the future
- Setting up a Data Sharing Board
- Looking at sharing data through an integrated structure (presented later in the meeting)
- Currently using the infrastructure with 'Cloudy with a chance of pain' dataset
    o   Using an app to create the data, expressing pain given the weather
- Other proofs of concept include
        o   anonymisation
        o   data from wearable devices going into the Safe Setting

University of Manchester (2) – Mary McDerby
 Introduction to Project S

- Secure data service predicated on the research data lifecycle
- Examining resource and cost models
- Year 2 of a 5-year project
- Highly restricted data service to be available next year

Manchester Foundation Trust Hospital - Sarah Al-Adely

- Involved in a project to store data at the Trust and to make available for research
- Functionality includes anonymisation, storage, SDC, building relationships between university and hospital
- Transferring data to the university for use by researchers

<u>The Health Foundation – Richard Welpton</u>

- Recruiting for 2 posts: Senior Data Manager for Networked DataLab, and Data Systems Environment Manager
- Issues with storage, reaching capacity too quickly. Project to reconfigure 30TB storage to take place in 2020.
- Partnership agreed with HDR UK, likely to lead to increased access in 2020
- External researchers will be allowed to use facility, researchers from Turing Institute to begin accessing data in 2020
- GitLab has successfully been installed and allows researchers to share non-disclosure statistical code

Other items arising

- There was an explanation of the **Digital Economy Act** accreditation
  - Output checkers to be accredited, discussion about checking
  - Safe Researcher Training test discussed
  - New ONS Assured Organisational Connectivity
- New ESRC-funded **SafePods**: Scotland have 6 installed already, Manchester University will receive one in 2020
- There was an update about the new **DWP Employment Lab**. There are few details of this, but it will work on a 'remote execution' basis (researchers sensing code in for DWP staff to run on data and return the results)
- There was some discussion about **data curation and how to provide a 'data as a service'** concept, which would include providing curated data, catalogues and metadata. The Digital Curation Centre and the metadata catalogue based at Oxford University were mentioned as exemplars.
- The Health Research Authority is due to publish **advice on data collected from wearable devices**
- The concept of **'demonstrating impact'** was discussed by group members, in particular recognising that the services we provide make much research possible that couldn't be undertaken otherwise
- There was a discussion by the group about projects undertaken in the secure data environments, in particular, **how researchers define projects and their end dates**, and a discussion about why some projects may follow a seemingly long duration.
- The group talked about '**over-anonymisation' and 'classification creep'**: this occurs as new secure facilities are created, then existing data sources that have been anonymised and considered sufficiently 'safe' to release via download with conditions become reclassified as 'secure' and access is restricted. **To be discussed at a future meeting.**
- The group discussed service **funding models** and in particular, whether upfront charges should be levied on users, and what the consequences of these charges might be.

Afternoon presentations: Safe Settings

- Gary Leeming from University of Manchester presented their 'Highly Restrictive Data Service', which is a secure data facility established at the institution to enable researchers to store confidential data they collect throughout their researcher. Slides accompany these minutes.
- Simon Parker from Cancer Research UK provided a live demonstration of the Cancer Intelligence Secure, which is powered by AIMES UK
- Martin O'Reilly from the Turing Institute presented the Turing's schematic for providing access to data based on the sensitivity of the data. Slides accompany these minutes.
- Valerija Kolbas from the UK Data Service presented plans to put together a set of standards for Safe Settings, inviting contributions from other participants. It was agreed that a separate meeting would be arranged early in 2020 to discuss this further.

Closing

The meeting closed at 4pm and the Chair thanked everybody for attending. The next meeting will be held at HMRC in Canary Wharf, London, on 18th March 2020.