Safe Data Access Professionals Quarterly Meeting

Thursday 26th September 2019

Essex University Business School

Attendees (in person)

Richard Welpton, (The Health Foundation) (Chair); **James Scott** (UK Data Service, local organiser); **Yannis Kotrotsios** (Cancer Research UK); **Joshua Symons** (NHS Digital); **Christine Woods** (UK Data Service); **Elaine Mackey** (University of Manchester); **Anthony Macken**(Central Statistics Office, Ireland); **Mary McDerby** (University of Manchester); **Beate Lichtwardt** (on behalf of SSHOC); **Emily Griffiths** (Greater Manchester Connected Health City); **John Ipe** (Office for National Statistics); **Aida Sanchez** (University College London)

1) Introduction

The Chair thanked everybody for attending this September meeting of the Safe Data Access Professionals, and welcomed new participants. The Chair also thanked James Scott and UK Data Service for the kind and generous help making the local arrangements for the meeting. We apologise to Alastair Currie (HMRC) who had intended to dial-in – due to technical problems we were unable to offer tele-conference facilities.

2) Round-table updates (all services)

UK Data Service Secure Lab

- UK Data Service are undergoing accreditation under the Digital Economy Act. The accreditation consists of two assessments: Security and Capability.
    - The capability assessment includes ensuring sufficient procedures are in place for undertaking SDC of outputs, ingesting data, de-identifying data, communicating to researchers and other operational matters.
    - Audit is expected in October 2019, and will be conducted by ONS on behalf of the UK Statistics Authority
- UK Data Service is undergoing a re-organisation, and the Access and Support teams will be merged

The Health Foundation

- The Networked DataLab (NDL) will be launched in October 2019. This initiative is aimed to support Local Authorities with the linking, storage and access of data collected about individuals that they provide services too. The NDL will have a research strand to it.
- GitLab has been installed to enable researchers to share code with each other within the Secure Data Environment. A pilot scheme is running until the end of December 2019.
- New sources of data continue to arrive, particularly from NHS Digital, CPRD and NHS England.

Office for National Statistics

- Training for output checkers is due to commence at the end of 2019.

Cancer Research UK – Cancer Intelligence

- The team are preparing for their next NHS Digital Data Security and Protection Toolkit (DSPT) accreditation
- AIMES as a platform for hosting their Secure Data Environment is working well. Presentation to be made at the December 2019 SDAP meeting in Manchester

University of Manchester Data Safe Haven

- Introduced by Mary McDerby
- Data Safe Haven for University of Manchester:
  - Consists of e-infrastructure, high performance computing (HPC), cloud-based
  - For use by University of Manchester researchers who are collecting sensitive data as part of their research - will be obliged to deposit their data into the Data Safe Haven
  - Based on research-data lifecycle model: creation of data, various data 'touch points' and archiving
  - Looking into how external collaborators can work with this
- Other University of Manchester news: will receive a new ESRC-funded Safe Pod to enable remote access to government data. 10 librarians are to be trained to support researchers who use it

Social Sciences & Humanities Open Cloud (SSHOC)

- SSHOC introduced by Beate Lichtwardt
- Formerly known as EUDAN
- Purpose is to make controlled data available internationally
- Standards on Safe Rooms to be published in due course

Central Statistics Office, Ireland

- More sources of data now available for researchers to request
- Centralising of key tasks continuing to progress
- Currently remote access security based on keyfobs: planning to move to 'soft' fobs (e.g. mobile phones)
- Researchers can access remotely only from their institutions; plans to enable them to access from broader IP ranges are under consideration

Centre for Longitudinal Studies/UCL

- Introduced by Aida Sanchez
- Currently deposits data with UK Data Service for providing access to birth cohorts
- Local secure servers used (UCL safe haven) for preparing the data, including linked data e.g. to HES)
- Key challenge involves decides whether data are classified for access via End User Licence, Special Licence or Secure Lab access

Imperial College

- Preparing for ISO27017/18 (cloud) accreditation
- ISO27001 accreditation undergoing renewal
- Bid made for Digital Innovation Hub
- Consideration of combining different secure environments into one central service

NHS Digital

- Move to cloud has been successful;
- Mental health data from 220 providers: processing times reduced from 5 days to 16 hours, tied to data processing platform
- Data Access Environment: new analytical software going in, Bring Your Own data issue to consider
- Awarded two Digital Innovation Hub awards

Connected Health Cities, Manchester

- Project funding has ended, and service is winding down
- However, a new European project may begin which will draw from the CHC experience

3) Other topics discussed

During the previous session, a number of topics arose which the group discussed, including:

- Describing how we manage secure environments - from setting up, running, etc.
- Standards of depositing syntax (from an SDC point of view)
- Sharing/accessing health data internationally
- Is there a gap for semi-secure infrastructures?
- Elaine Mackay mentioned that the new ICO Anonymisation Code of practice (first draft) should be available from December 2019; and also that the Anonymisation Decision Making Framework will be published as a book
- Elaine Mackay talked about new data penetration tests that have been developed

4) Discussion about acceptable outputs for release

The group were joined by other UK Data Service Support staff. The following issues were discussed:

- What outputs actually constitute a dataset (e.g. obviously very large tables)?
- How to manage such requests

The group discussed a number of issues around this:

- Practicalities in terms of time and resources taken to make an assessment
- The risk that a dataset is released which the original data supplier would not wish to have released
- Whether such an output would be published in e.g. a journal or report
- Should the principle of using a secure environment be that analysis takes place in the secure environment? Therefore, should requests for derived data intended for further analysis outside the environment, be refused?
- Do output checkers have the skills, time, assessment criteria that statistical organisation deploy before releasing published results?

Generally, the group agreed that large tabular outputs that were clearly derived datasets ought not to be released; a point about training researchers about this issue was raised

5) Researcher training and assessments

Richard Welpton and Yannis Kotrotsios presented the first module of a new joint CRUK-THF training course for researchers. The assessment exercises for this module were also presented to the group

The training covers the following topics:

- Introduction to data; why are some data more sensitive than others; Data Access Spectrum, safeguards and the GDPR; classifying data
- The Five Safes as a data governance framework
- Safe People
- Safe Outputs (five or six modules)

The training is on the SDAP website (https://securedatagroup.org/training2/)

Felix Ritchie also asked for the Secure Researcher Training to be linked via the website, this is also available

6) Meeting closed

The meeting was closed at 16:00, and the Chair thanked everybody for their attendance and contributions.