

Pan-UK Data Governance Steering Group

26 June 2024

Reminder

The UK Health Data Research Alliance is convening a **Researcher Registry Information Governance and Policy Working Group** chaired by Professor Felix Richie. The group's objective is the co-development of a common researcher verification standard and supporting its adoption. Please register to attend one of the workshops via the Eventbrite links below.

Workshop 1: [18th July 13:00 - 14:30](#)

Workshop 2: [22nd August 13:00 - 14:30](#)

Workshop 3: [5th September 13:00 - 14:30](#)

Summary of Key Discussion Points

Chair: Andy Boyd

Welcome & updates for the Steering Group: Speaker - Cassie Smith

Transparency Standards

[\[https://zenodo.org/records/8262453\]](https://zenodo.org/records/8262453) which are as follows:

1. Open access application form and guidance
2. Transparent application process and criteria
3. Clear website navigation
4. Consider target audience
5. Regular review of website content
6. Transparency of data use and auditing

HDR UK hosted an Alliance Transparency Showcase in May to celebrate the work of the 19 organisations that were awarded funding to implement the Transparency Standards. Congratulations to the poster winners, **DATAMIND**. Abstracts from the funded organisations



have been published in the International Journal of Population Data Science. (<https://ijpds.org/issue/view/37>).

GUARDS principles

Workshops were held with Steering Group members to refine the principles, with GUARDS acting as a set of overarching information governance principles for the data science community. Public consultation is underway, in the form of a survey co-produced with HDR UK Public Advisory Board (PAB) members. Following this consultation, the principles will be updated with the public feedback for wider engagement with the Alliance alongside the submission of a journal paper with the finalised principles and commitments.

A note of congratulations to Jillian Hastings Ward

Jillian, a former PAB member, has been awarded an MBE for her outstanding service to patients and participants in the field of UK genomics. Congratulations Jillian!

UPRN linkage: Speakers – Andrew Boyd, Pia Hardelid

What is UPRN? UPRN is an acronym for Unique Property Reference Number. It is assigned by local authorities, made available via the Ordnance Survey, and facilitates discrete grouping such as care homes or university accommodation.

Kid's Environment and Health Cohort. A cohort that is aiming to establish a novel linked education, demographic, health & environmental resource for wider child health research. Using UPRNs in this research allows environmental aspects to be investigated with more precision.

ECHILD. The Education and Child Health Insights from linked data project is led by Katie Harron and Ruth Gilbert and is the first linked data set spanning spending, education, and health care. It aims to use UPRN data to derive and validate household groupings within the original dataset.

Governance of using UPRN data. UPRN is available and mappable to a full address and is a unique identifier. Whilst it is personal data, UPRN data alone (unlinked) can be considered as relatively low sensitivity. Furthermore, there is a general expectation that the data is already in use and is routinely shared in the public domain. Once UPRN is linked to other identifiable data it must be treated with the appropriate governance.

Managing the governance risk of using UPRN data: UPRN data falls into two categories; small data and big data, depending on quantity of data and compute required for processing. Small data is easily processed within a TRE, and big data, as often is the case with environmental data, requires bespoke computing environments and skills.

Model 1 – single source TRE setting for ‘small’ data. The UPRN data is pseudonymised and then linked with the health data. This flow suits small data and is regularly carried out within TREs.

Model 2 – ‘split’ file TRE for ‘small’ data. In this model there is an additional step in the process where a third-party data processor receives the UPRN data, similar to the model operated in the SAIL databank. This additional step means that the TRE operators do not need the UPRN data and mapping, providing an additional risk control measure.

Model 3 – multi-party processing for ‘big’ data. In the first step of the model, a data owner (for example, the NHS) would send the health data to the TRE where it is deidentified and minimised. The data owner would then encrypt and mask the UPRNs associated with the health data to send to a geospatial environment (for example, the Met office). This model includes an additional step to further mask the UPRNs by providing UPRNs in the location and timeframe of interest that are not in the cohort. The geospatial environment can then derive the environmental data of interest. This approach requires the data owner to share the encryption key with the geospatial environment and the masking cases with the TRE so that these cases can be removed at the point of analysis.

Discussion points:

- UPRN vs post code information. UPRN provides much more information than post code so is essential for this type of research.
- Process automation – This current process is costly and time-consuming but there are existing projects looking to automate some aspects, where possible (specifically, disclosure control).
- Public involvement – work with the HDR UK Public Advisory Board is being undertaken to produce a learning module for researchers and the general public.
- The use of TREs and data minimisation – Data minimisation is an important risk mitigation strategy, the need for minimisation may depend on level of encryption and masking.

Safe Programmes: Speaker – Ed Humpherson

Office for Statistics Regulation - The UK Statistics Authority (UKSA) is a non-ministerial department and is responsible for implementing the Statistics and Registration Act. It also holds TRE accreditation responsibilities and the Researcher Accreditation Panel.

Underneath UKSA, there are two executive offices. The Office for National Statistics (ONS) and the Office for Statistics Regulation (OSR). Differentiation between the two is essential because the OSR has a responsibility to regulate the ONS. The OSR is responsible for setting the standards for official statistics and ensuring that ONS and any other part of central government that produces statistics complies.

Data Sharing and Linkage for the Public Good. OSR's goal is to ensure that statistics maximise and serve the public good. Data should be linked to enable greater insight. The report found lots of strengths and progress, including the work done by ADR UK and HDR UK. Consistent public engagement builds social confidence in data sharing and linkage. Therefore, a follow-up report is being conducted and the OSR is planning to publish this report in mid-July with three main areas of exploration:

- The Five Safes
- Safe Programmes
- Transparency of outputs.

The Five Safes Framework. The five safes are extremely positively regarded and widely used both in the UK and internationally. However, the implementation of the five safes does present challenges for researchers. Specifically, in the application for data access process. It should be considered whether data owners and the UKSA are following legislative requirements or if they go beyond what is required.

Safe Programmes. The 'Safe Projects' aspect of the Five Safes Framework could be updated from projects to programmes or purpose. Researchers commented that they are required to make very specific use cases for data access at points in their project where it is very difficult to be so specific. Therefore, the safe project phrase should be updated to programme to encourage approvers to be more accepting of overall research aims. With the other four of the five safes in place, the overall risk is manageable, and it is acceptable for approvers to require less information on the project details.

Publishing outputs. Researchers reported two main points of feedback. The back-end cost and time it takes for outputs to be produced and a risk adverse culture around outputting data that needs to be published. To address these problems, we need to avoid publication bias that gives preferential treatment to significant results. We also should be using intelligent transparency, based on the idea that the government should publish its analytical evidence to support the public accountability wherever possible and every time analysis is used or voted publicly.

Discussion points

- Group members are receiving increasing feedback on the difficulty of output checking and disclosure controls. There is a general lack of standardisation. However, increased issues with output checks could be a symptom of more projects progressing to the stage where output checks are necessary.
- The culture towards output checking in TREs has been very risk adverse. We need to encourage TREs to adhere to the legal framework and encourage TREs to not go beyond the requirements without justification.
- HDR UK's Public Advisory Board has stressed on the importance of data minimisation as a safeguard to build public trust.
- Changing to a safe programme approach allows for broader permissions but there is still a need for specific permissions e.g. who has approval to be conducting the work.

Suggested agenda items for the next steering group meeting based on these discussion points:

- **Output checking standardisation.**
- **Data minimisation.**

Future Data Services: Speakers – Richard Welpton and Felix Richie

What is Future Data Services about? The data landscape has exploded over the last few years. The Economic and Social Research Council (ESRC) has been investing in data services infrastructures for some time. We now know it as the UK Data Archive and it's part of the UK Data service. ESRC works closely with organisations like the ONS and other research councils.

Future Data Services themes:

- Discovery and Data Curation
- Data access, user support and training
- Technology
- People, organisation, and culture
- Legal, ethics, public engagement, and impact
- Discovery and data curation

Data Access and Culture. Data access is not as seamless as it should be. ESRC has received feedback from researchers and data owners on the data access process. One of the main issues was the lack of predictability for time frames of access to data. This highlighted a problem of culture within TREs.

The Culture problem.

- Awareness - ESRC found that some operational staff were not aware of access to data as an objective within the overall research process.
- Approach – Data access processes were undertaken sequentially where some could be undertaken in parallel to reduce time frames for access.
- Aptitude – Individuals were not confident in making decisions on access to data and consulted more senior operational staff, maintaining security but prolonging access to data.

How do we change the culture of data access from 'default closed' to 'default open'. Many organisations agree with the recommendations by the ESRC but view the changes as aspirational. How can we manage these problems within the Pan-UK Data Governance Steering Group? One recommendation is that we move from data-centred to user-centred, which would reframe attitudes from “what we can do with this data?” to “what researchers want to do with the data?”

Discussion points on the Steering Group and cultural change:

- The use of service metrics could encourage TREs to address risk adverse culture. The use of these metrics is becoming more common, but an agreed publication method is required. Maintaining public benefit also needs to be considered when streamlining data access.
- Training for operational individuals within TREs could increase confidence in decision making. However, this approach might require funding.



Open forum: Speaker - Cassie Smith

ICO case studies: Some steering group members, including HDR UK, have been asked to produce case studies for forthcoming ICO guidance on anonymisation principles. Updates to follow.

DPDI Bill. The Bill is dead having not been passed in the “wash up” period of the outgoing government.

Potential future agenda items.

- **Article 89 of GDPR – Scientific research and minimisation.**
- **Output checking.**
- **TRE Data Access culture- change in perspective.**
- **Review of Five Safes Framework- adjustments to ‘Safe programmes’?**
- **Governance of UPRN data**

It was discussed that another Frontiers meeting could be held to delve into these issues in greater depth.

In attendance:

Andrew Morris Cassie Smith Andy Boyd David Seymour Edel McNamara Yemi Macaulay Rachel Brophy Beth Gilbert Lara Edwards Doreen Tembo Ester Bellavia	HDR UK
Paola Quattroni Stephen Burrows	HDR UK Health Data Research Alliance
Pia Hardeid	University College London
Ruth Gilbert	University College London
Sara Jane McAteer Munisa Hashimi	HDR UK Public Advisory Board
Felix Ritchie	University of the West of England (UWE)
Ed Humpherson	Office for Statistics Regulation (OSR)
Rhys Nadin	Office for National Statistics (ONS)
Alex Bailey	Medical Research Council (MRC)
Richard Welpton Joe Cuddeford	Economic and Social Research Council (ESRC)
Ruth Blackburn	University College London
Alison Knight	Health Research Authority (HRA)
Nadia Azimikorf Peter Wright	CPRD/ MHRA
Balint Stewart	DARE UK
Carole Morris	Public Health Scotland (PHS)
Garry Coleman Narissa Leyland	NHS England
Jim Squires	Association of the British Pharmaceutical Industry (ABPI)
Maeve Groot Bluemink	Our Future Health
Nick Maltby	Genomics England
Paul Jackson	Research Data Scotland (RDS)
Gary Ricker Colin Wilson	Office for Life Sciences (OLS)
Claire Edgeworth	NHS North of England Commissioning Support Unit (NECS)
Tim Hubbard	ELIXIR
Frances Burns	Northern Ireland Health and Social Care