

International Journal of Population Data Science

Journal Website: www.ijpds.org



Generation Scotland – Transparency in the data access process

Archie Campbell^{1,2}, Alex Wood², Robin Flaig^{1,2}, Daniel McCartney², Heather Whalley^{2,3}, and Cathie Sudlow^{1,2,4}

¹Usher Institute, University of Edinburgh

²Centre for Genomics and Experimental Medicine, University of Edinburgh

³Department of Psychiatry, University of Edinburgh

⁴Centre for Clinical Brain Sciences, University of Edinburgh

Background

Generation Scotland (GS) is large family-based cohort study established as a resource for research into the genetic and environmental determinants of physical and mental health. Participants completed a baseline questionnaire and clinic visit in 2006–2011.

24,084 Participants

5,501 Families

- **NHS data linkage** 22,402 (93%)
- **Genotype array data** 20,026 (83%)
- **DNA methylation data** 18,876 (78%)

In 2019 funding was obtained to expand the existing GS cohort. By 2025 NextGenScot aims to:

- Recruit **20,000 new participants** online
- Expand eligibility to **12–17 year olds**
- Collect saliva samples by post for genotyping
- Linkage to NHS and other administrative data

Details on how to join at www.gen.scot

Participants consented to linkage to NHS Scotland medical records which offer longitudinal phenotype information. Linkage is conducted using the Community Health Index (CHI).

Introduction

GS receives 30–40 requests for access to data and/or samples every year, and has been included in over 380 publications. Research proposals are subject to review by the GS Access Committee. Details of the application process terms and conditions for access are available on the GS website, and now on the new research access website.

The previous GS access process was based on an application form in Word, downloaded from the main GS website. We are replacing this with a new online research access system, adhering to transparency standards¹, helped by HDR Alliance Transparency funding.

After researchers publish their results, a link to the publication and a lay summary is currently posted on the GS website, with a spreadsheet listing papers. We will now make this information available and searchable for researchers, participants and the public on the new research access website.

Methods

We are developing a website, using in-house staff, with HDR Alliance Transparency contributing to their salaries. The system coding uses Django (a Python tool) linked to a MySQL database, a platform widely used for such projects, enabling faster development in a secure environment.

Transparent access to GS data for registered users

- View summary of data available
- Query the summary data to see if it fits research requirements
- Select data requested from data dictionaries
- Get a detailed quote for access charges
- Complete application form showing the project complies with the Five Safes²
- Upload supporting documentation (e.g. protocol, DPIA, ethics, etc)
- Track progress of their application review, approval and delivery
- Researchers will be required to submit a lay summary of their application

GS participants, researchers and the public can view:

- List of approved applications with lay summaries
- Summaries of ongoing projects which will link to eventual publications
- Searchable list of publications

Results

The new system is undergoing internal testing, using data imported from previous and ongoing access requests (May 2024). We will then pilot it with selected external researchers and GS participants (June 2024), with a plan to launch it for public access in July/August 2024.

By adding a participant/public access area, with lay summaries of projects and publications, GS participants will be able to see how their data is being used for public benefit, as can potential new volunteers or other interested parties.

Conclusions

We were already planning to develop a new online application system, but the Transparency grant funding has enabled us to extend the functionality beyond a simple form-filling system. The new access website will give researchers more information about the application process, the data and samples available, see what other research has been approved in their area of interest, and track the progress of their project from enquiry through to publication.

By adding a participant/public access area, with lay summaries of projects and publications, GS participants will be able to see how their data is being used for public benefit, as can potential new volunteers or other interested parties.

Notes

The URL for the new website is www.gsaccess.org

The main GS website is at www.generationscotland.org

References

1. Transparency Standards <https://zenodo.org/record/8262453>
2. Five Safes <https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/>

