Job Description

1. Job Details
   Job Title: GS:21CGH Data Manager (AD Scale)
   School/Support Department: Molecular and Clinical Medicine / Medical Genetics
   Line Manager: Science Manager, Medical Genetics

2. Job Purpose
   The main area of responsibility of the Data Manager is coordination and development of all IT aspects of the GS:21CGH project. The data manager has the role of specifying a complete IT infrastructure (hardware and software) for the capture, storage and dissemination of genotype and phenotype (clinical) data on the 21CGH project (part of Generation Scotland). This infrastructure must be designed so that it is scalable and suitable for rolling out on to larger population studies such as the Scottish Family Health Study (also part of Generation Scotland). This ensures that the data manager plays a key role in IT infrastructure building not just for 21CGH, but for Generation Scotland as a whole. The Data Manager also designs & implements IT infrastructure for the WTCRF Genetics Core at the Western Hospital in Edinburgh. This will integrate with the new Laboratory Information Management System (LIMS).

3. Main Responsibilities
   • Clinical data collection system: Provide front-end IT solutions, support and training for the remote collection of clinical data by research nurses in a variety of healthcare settings. This involves selecting appropriate software and hardware to support remote data collection. This also involves developing data entry forms and piloting these using Welcome Trust Clinical Research Facility (WTCRF) research nurses before field use by 21CGH nurses. Training must be given to the 21CGH nurses in how to use the technology provided. When data is being collected by the research nurses, this must be managed daily. On call support will need to be provided by the data manager during the clinics’ open hours Monday to Friday. Ensure all clinical form data is received and validated.

   • Automatic entry of clinical data into database: Develop, pilot and maintain an automated data loader programme for entry of the data captured above into the phenotype database. Develop, pilot and maintain a graphical user interface to enable administration and management of all captured data.

   • Storage of project data: Design and maintain a database storage network in order to serve the research interests of other studies looking to analyse the datasets. The database storage must meet the needs of the security and patient confidentiality. To this end it will require the use of regular maintenance and backup procedures. All phenotype and genotype project data shall be stored here. Additionally, confidential information will be stored on a separate PC with a regular CD backup procedure. Work with WTCRF staff to ensure all project blood samples are received. Co-ordinate genotype data structure and handling, and enter genetic data into the genotype database.

   • Dissemination of data to researchers: Develop, pilot and maintain an IT based data delivery solution to enable researchers to analyse the genotype and phenotype information contained in the database storage network over the world wide web. Collaborate with institutions within Scotland and Europe in order to ensure that the datasets collected are in a useable format for researchers on other studies both nationally and internationally. Institutions include the NHS Information and Statistics Division (ISD), the National E-Science Centre (NEsc), the Finnish Genome Centre (Biomedicum) and the World Health Organisation.

   • Design and implement an IT infrastructure for the WTCRF Genetics Core at the Western Hospital in Edinburgh that will integrate with the new Laboratory Information Management System (LIMS). Initially assisted Science Manager with the selection of a suitable LIMS system. Once the LIMS manager is in post, assist him/her with ongoing maintenance of the LIMS database, dealing with any infrastructure or database problems which may arise.
• Website administration and maintenance: In association with the Informatics Research Development Manager, administer and maintain the project web site (www.generationscotland.org). This includes web page changes (e.g. dissemination of reports and scientific literature), renewal of domains and setting up of email addresses.
• Development of IT systems for other research projects: There is the possibility of setting up data capture collection and dissemination systems for other research studies in the future.
• Ongoing research and development into research IT systems, and the linkage of NHS and research patient data.

4. Planning and Organising
All planning and organising of the IT infrastructure and IT related matters is the remit of the data manager. This has been long term planning due to the length of the project contract. It has been necessary to work with the 21CGH team in order to coincide with the timescale of the ethics submission and patient recruitment.

Over 1st year: (i) Meet key project stakeholders; (ii) Gather project requirements; (iii) Decide on IT System Architecture
(iv) Develop IT System Infrastructure specification; (v) Build Clinical data collection system

Over 2nd year: (i) Build application to enable automatic entry of clinical data into database; (ii) Build database storage network; (iii) Pilot IT infrastructure; (iv) Roll out IT infrastructure and train nurses; (v) Maintain and support IT infrastructure.

Over 3rd and 4th years: (i) Develop Genotype database; (ii) Develop web system for research data access; (iii) Ongoing IT research and development.

5. Problem Solving
Designing and building an IT infrastructure for collection and storage of genotype and phenotype data that can be scaled up to handle a large genetics bank of 50,000 samples and upwards; Dealing with all resulting support and maintenance issues both in the field with the nurses and at the server back end; Designing and building an IT data delivery solution for researchers that uses internationally recognised standards, in collaboration with similar studies in Europe and industry; Designing and building an IT infrastructure to store all genotype data from the WTCRF genetics core and implementing this in conjunction with the LIMS provider; forward planning rollout of IT infrastructure; Collaborating with GS Informatics group to formulate Scotland-wide research IT infrastructure; Policy decisions on major strategic decisions would be discussed with GS Principle Investigators/line manager.

6. Decision Making
Taken independently: Decide on architecture of IT infrastructure for 21CGH eg. what infrastructure to use or build and where; Decide on detailed specifications of IT infrastructure eg. Hardware/software; Decide on maintenance and support level of IT infrastructure; Resolve support issues with hardware in the field
In collaboration with others: Deciding on which technologies will be adopted across the whole of Generation Scotland; Building systems to enable linking of patient identifier information (Dundee); Dividing of IT tasks across Generation Scotland.
Referred to manager: Things that need to be decided by the principal investigators (Strategic Management Group);
Dealings with the press.
Level of Direction Given: Orientation and advice; Approval of IT proposals.

7. Relationships and Contacts
Internal: (i) Generation Scotland Team (Project Manager, Informatics Research Development Manager and Science Manager); (ii) Genetics Core staff in the WTCRF, Western General Hospital, Edinburgh; (iii) Professors & Senior Academics at University of Edinburgh
External: (i) Nurses at medical practices – provide IT training and support; (ii) Mackenzie Institute in Dundee – IT collaboration; (iii) Robertson Centre in Glasgow – IT collaboration; (iv) NHS Information Services Division – IT linkage
(v) IBM Life Sciences – IT collaboration
8. **Knowledge, skills and experience needed for the job**

The post holder should have experience of taking a project through the whole software lifecycle – requirements capture, design, development, support. Good interpersonal skills are important, particularly in dealing with end users. A relevant degree and at least two years of either academic or commercial experience in an appropriate area are essential. Key skill sets are: Database Design, ideally using an enterprise class database eg. Oracle, SQL Server, MySQL; Experience of writing automated data loader programmes; Familiarity with designing data entry forms; Experiencing of evaluating and selecting commercial off the shelf software and hardware, for supporting remote data collection and management; An understanding of the definition of data standards and specifications; networking skills & ability to train users in new technology; wide knowledge of IT solutions to enable IT architecture level decisions.

9. **Dimensions**

The post has no direct reports. The data manager works directly with two AS2 grade staff, one AS3 and one Professor of Medicine. Also the data manager will be working directly with research nurses across Scotland in the piloting of the IT infrastructure. The data manager is part of the GS Informatics group, which is a decision making team made up of key informatics staff from the Universities of Edinburgh, Glasgow, Dundee and Aberdeen, as well as from the NHS Information Services Division and the National E-Science Centre. The data that will be stored in the IT infrastructure will be disseminated to researchers funded by 21CGH and to those on other studies. The data is collected to serve as a bio-bank for any studies to use as controls in the future, so the number of researchers accessing this data is potentially very large. These researchers will be direct customers of the IT infrastructure that the data manager builds.

10. **Job Context and Any Other Relevant Information**

Generation Scotland (www.generationscotland.org) aims to determine the genetic basis of common diseases by creating an ethically sound, population and family based sample collection in Scotland in order to identify genetic contributions to individual risk of disease, disease prognosis and response to treatment. GS has so far contributed to winning over £6.4million in grant awards. GS:21CGH will collect phenotype and genotype information from around 2000 participants in one year, and will feed into the larger GS:SFHS study which will collect 50,000 samples over 5 years.