

Effectively Storing and Analyzing Biological Data at UK Biobank

UK Biobank is a major medical research initiative based in Manchester, UK. The company's goal is to build the world's most detailed information resource that can support a diverse range of research aimed at improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses and to promote health throughout society. UK Biobank chose Thermo Scientific LIMS to track, store, manage and report on the biological data of 500,000 people as part of one of the largest medical research projects ever conducted.

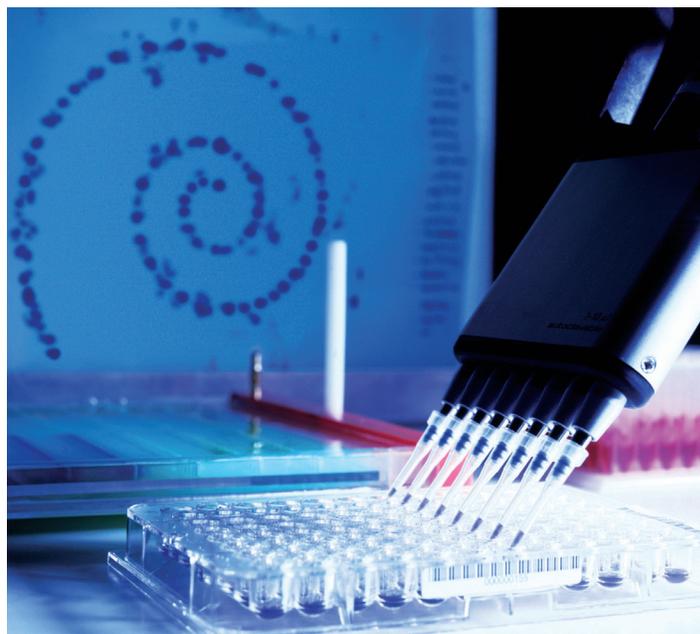
UK Biobank's operations and procedures

The relationship between a nation's health and poverty is well known, but the solutions for closing the gap between health provision for rich and poor are not so easy to find. A great deal of knowledge is available from different sources, but there is no comprehensive system for the systematic collection of data which can be used to establish the true impact of disease on a nation's prosperity, and vice versa. One aspect of this is being addressed by UK Biobank.

Over the next few years the UK Biobank is building a uniquely rich resource for investigating why some people develop particular diseases while others do not. During an initial recruitment period of three-and-a-half years, 20 assessment centers were set up around the UK. The organization recruited just over 500,000 participants by mid-2010. All information collected is stored centrally and updated as each participant's health status is monitored.

The aim was to recruit as many participants as possible and collect as much detailed information from them as possible. The number of participants, responses to questions and quality of measurements are critical to a favorable outcome for this project. To accommodate for the vast scope of this project, a comprehensive Laboratory Information Management System (LIMS) to collate, store, sort and retrieve relevant data in a useable format was required. In addition, the LIMS needed to be capable of yielding real-time, dependable analysis and reports and providing secure access to data at any given time.

In an attempt to address the needs of this challenging and complex project, UK Biobank selected Thermo Scientific LIMS to store and track the tremendous amount of data generated from the processing



and analysis of samples. In addition to answering a variety of lifestyle and medical questions using automated touch screen technology, participants donated blood, urine, saliva and RNA samples which were processed at a central processing center to yield more than 15 million individual aliquots. All processing and analysis was tracked by Thermo Scientific LIMS. For each aliquot the LIMS records a reference to its parent, its 2-D barcode ID, its rack barcode ID, its position on the rack, the location of the rack in the store and the aliquot contents, among other things. In total, UK Biobank is holding several hundred million data points within the underlying database.



Format of the study

Following consent, participants were asked to give small samples of blood, urine, saliva and RNA for long-term storage and analysis and have some standard measurements such as blood pressure, height (standing and sitting), weight and bioimpedance, body fat/waist and hip circumference, hand grip strength, bone density and lung function. Participants were also asked to complete a confidential health, lifestyle, memory, work and family history questionnaire while their routine medical and other health-related records will be followed by UK Biobank over the next 30 years.

Eventually, UK Biobank will allow fully approved researchers to access data and samples to study the impact of lifestyle, environment and genes on the progression of illnesses. By analyzing answers, measurements and samples collected from participants, researchers may be able to work out why some people develop particular diseases while others do not. This will help researchers to understand the causes of diseases better and to find new ways of preventing, diagnosing and treating many different conditions.

Thermo Scientific LIMS

First commercialized in 1998, Thermo Scientific LIMS is designed to address applications that require more flexibility than a traditional enterprise LIMS. Flexible and robust, with automated plate handling and easily configured extensions, Thermo Scientific LIMS is the LIMS of choice for many organizations.

Thermo Scientific LIMS includes patented workflows technology with a flexible and intuitive interface to graphically map laboratory workflows of the sample life cycle. With its built-in instrument integration, the

system offers productivity gains right from the outset whereas the flexibility of the software offers the option to build extensions in order to interface to other systems. The system is compatible with Microsoft® Office while being designed, developed and supported within an ISO 9001/TickIT environment.

UK Biobank chose Thermo Scientific LIMS as it is designed to meet the unique requirements of R&D laboratories. Without a clearly targeted and appropriate information management system in place, researchers would be unable to access and use the enormous information resource created by the UK Biobank project.

“The flexibility, adaptability, scalability and stability of Thermo Scientific LIMS, combined with Thermo Fisher Scientific’s commitment to biobanking operations and long-established experience serving the health sciences and pharmaceutical industries, make Thermo Scientific LIMS an excellent choice for the data management of this particular project,” said Darren Jackson, LIMS manager for UK Biobank. “At the height of participant recruitment activities, UK Biobank was collecting samples and data from six concurrently operating remote assessment centers, six days each week. Approximately 600 participants attended the assessment centers daily, donating key data and about 5,000 biological samples in the process. Laboratory processing of the samples provided yielded approximately 20,000 aliquots stored on 200 racks each day. Thermo Scientific LIMS was instrumental in enabling such high-throughput activities to be supported.”

Prior to joining UK Biobank, Darren Jackson gained over 10 years of experience in designing, implementing and administering Thermo Scientific LIMS systems in a variety of laboratory

environments. On joining UK Biobank, a year after commencement of the study, Darren inherited a LIMS that had been configured to support initial participant recruitment operations and was responsible for further developing and administering the system as laboratory operations intensified and diversified.

Data collection

During the initial recruitment phase of the UK Biobank study, daily participant data and samples were transferred securely to the UK Biobank coordinating center. Following sample processing, multiple aliquots (up to 36 per person) were produced and stored in an automated -80°C working archive and in a back-up liquid nitrogen store at a geographically distinct location for security. During assessment and sample collection, each participant was issued with a Universal Serial Bus (USB) memory key which acted both as an identifier and as a back-up temporary data storage device.

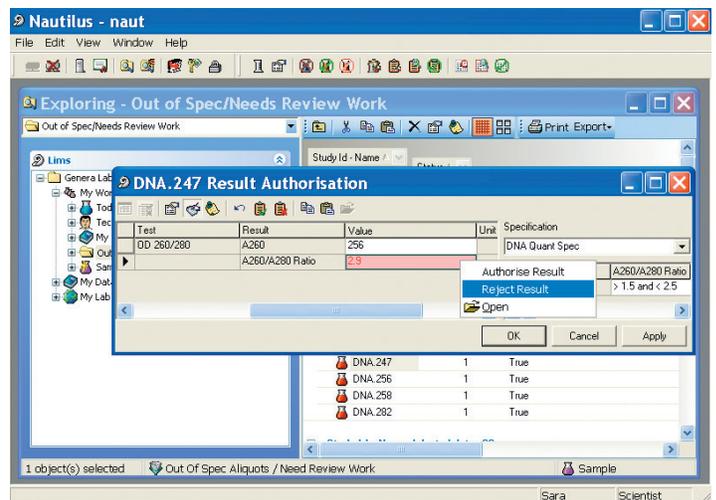
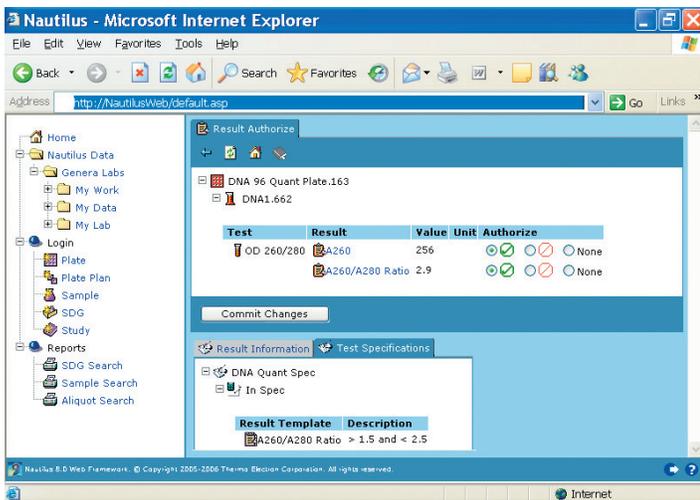
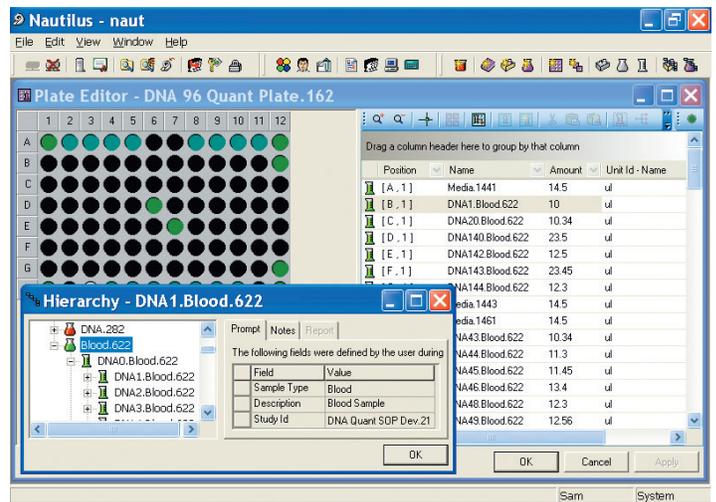
Barcoded vacutainer tubes were used to collect blood, urine, saliva and RNA samples at the assessment centers using pre-prepared racks. Following the collection of a complete set of samples from a participant, computers verified that each participant had completed all of the required tests and the unique barcode on each vacutainer tube was

then scanned into the assessment center IT system thereby linking each vacutainer with a unique participant identification number. This process allowed for a link to be generated between the participant data from the assessment center with the start of the laboratory data structure in the central Thermo Scientific LIMS.

When the vacutainers arrived at the central laboratory, they were scanned and compared against a LIMS data file generated from data provided in the assessment centers. The vacutainers were then processed in the laboratory using automated systems. The processing times and temperatures of all operations and associated laboratory technician identifiers were all logged in the LIMS. In the future any samples required for research for a given participant(s), will be identified in LIMS and retrieved from racks containing the required tubes using their barcoded identifiers.

Benefits achieved with Thermo Scientific LIMS

Thermo Scientific LIMS was initially tested for efficiency and accuracy, at UK Biobank, in a three-month pilot study, which took place between March and June 2006, and recruited 4,000 participants. During the pilot study, the system was proven in storing location and linkage information between participants and samples/aliquots. It also maintained a record



After collecting the data, results can be reviewed and authorized on the web or within the Thermo Scientific LIMS client.

of the volume of samples used and the volume remaining. This has triggered replenishment from the back-up archive and has helped guide resource access decisions for depleted samples. The integrated pilot study demonstrated that automated interfacing and validation of data from robotic pipetting and formatting workstations worked well with no problems encountered. Thermo Scientific LIMS' built-in process validation prevented human-related errors in data transcription.

Though Thermo Scientific LIMS offers flexibility in workflows required by R&D organizations, UK Biobank has specifically configured the LIMS to follow certain fixed workflows such that all samples follow the same protocols for testing and storage. The flexibility of the solution will prove vital for the effectiveness of the long-term study, since certain parameters and requirements are liable to change as processes evolve. To appropriately handle and archive the vast amount of data generated by the UK Biobank study, Thermo Scientific LIMS has developed into a central repository of an automated system that receives samples, fractions them into appropriate vessels for testing, and then analyzes, tracks and stores all data relevant to those samples. Data collection, resource management and data processing has been largely fully automated and all key UK Biobank study laboratory data stored in a central database. During the course of the initial participant recruitment phase, the LIMS was routinely used to provide daily updates and management reports to enable laboratory processing activities to be easily tracked. It is anticipated that, eventually, authorized researchers will be able to access relevant LIMS data, as required, to support their work.

Next steps

Following the conclusion of the participant recruitment phase, UK Biobank will now further develop the study, over the coming years, by inviting selected participants to undergo follow-up studies. The Thermo Scientific LIMS will be further developed to support such ongoing operations.

The LIMS will eventually serve as part of a comprehensive data warehouse for researchers who will need to access and query key data and samples, analyses and results.



Conclusion

With the help of Thermo Fisher Scientific and its Thermo Scientific LIMS solution, UK Biobank will enable scientists to gain unique insight into the genetic and environmental causal factors associated with a wide range of debilitating diseases, providing vital information needed to work on future preventative and curative measures.

Find out more at thermofisher.com/IntegratedInformatics

Australia/Asia Pacific +61 8 8208 8200

China +86 800 810 5118

France +33 160924800

Germany/Eastern Europe

+49 6103 408 1110

India +91 22 6742 94 94

Latin America and the Caribbean

+1 781 933 4689

Netherlands/Benelux +31 76 579 5555

Spain/Portugal +34 914 845 965

United Kingdom +44 1619423000

USA/North America +1 800 395 5467