Nuffield Consultation on Biological and Health Data

Submission by

Professor Neil Lawrence
Professor of Machine Learning
Department of Computer Science
Professor of Computational Biology
Sheffield Institute for Translational Neuroscience
University of Sheffield
N.Lawrence@sheffield.ac.uk

1) Do biomedical data have special significance?

Clearly it does, in that it enables specific predictions that we may be able to make about an individual's life outcomes. Knowledge of which would have a direct effect on that individual's emotional state, their family and their quality of life. However, to a large extent it is just a distillation of the effects we see when we look at commercial data, or social networking data on a large linked scale.

2) What are the new privacy issues?

The new issues relate to our ability to link across disparate data sets and draw inferences which are strongly predictive about and individual's life outcomes. The ability to link, for example, loyalty card data with genomic data, gives a relatively complete picture of the environmental and biological background for an individual. Holders of this data may be able to make relatively accurate predictions for future life outcomes for this individual.

3) What is the impact of developments in data science and information technology?

Already we are seeing that inferences can be made about an individual's commercial behaviour (leading to targeted advertising) which may distress that individual. For examples, prediction of pregnancy in an teenage mother before that teenager's family knows. This can be achieved by monitoring commercial behaviour. Once this information is cross linked with biomedical information and the relevant set of algorithms are developed, many more situations such as this are likely to arise.

4) What are the opportunities for and the impacts of, the use of linked biomedical data in research?

Very large advances in understanding. Relation of phenotype of disease to environmental causes on population wide scales.

5) What are the opportunities for and the impacts of, data linking in medical practice?

Improved care, improved monitoring of care, better patient outcomes. Better monitoring of patients in the community. Intervention with patients before diseases have taken hold. Early diagnosis of dementia, Parkinson's, motor neurone disease etc. Massive reduction in healthcare costs.

6) What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and healthcare?

Better integration of lifestyle with health is the main benefit. But, strictly speaking this is not moving outside healthcare. There are particular dangers with commercialisation of this process. Monetary incentives for an individual company are not necessarily aligned with the societal incentive for reducing the cost of healthcare and improving the quality of life for individuals. Ensuring the balance between societal benefit and profiteering at the expense of individuals will be a major regulatory challenge. When this balance goes wrong (as we see with direct marketing of drugs in the United States) the costs of healthcare rise and the quality of healthcare reduces.

7) What legal and governance mechanisms might support the ethical linking of biomedical data?

A major re-examination of the data legislation, taking place over a number of years (decades even), is probably required. As technology evolves, the right solution to the challenges evolves. Just as in the motor industry legislation evolved from 'red flag' legislation to 'the highway code' as roads and cars changed. This process took sixty years.

My own preference would be to return the ownership of all data (commercial and biological) to the individual, but to allow companies to use the data under a licensing agreement. This agreement could be terminated by that individual (electronically) under certain conditions. An analogy would be high street banking services. The money in the bank is owned by the individual and can be withdrawn by the individual. The individual stores their money with the bank for mutual benefit, but can terminate the contract. Currently this is not the situation for data. We are unable to terminate this relationship in a convenient fashion. There is very little information in the public domain about how this data is actually being used. The problems we will face with biomedical data are perhaps already with us, in the form of commercial data that is already stored about us. Regulatory frameworks need to be put in place for these types of institutions. One could even envisage 'data mutuals' where individuals choose to combine together to share their data for mutual benefit under regulated frameworks.