

Royal Society – Written evidence (LSI0119)

The Royal Society is the UK's national academy of science. It is a self-governing Fellowship of many of the world's most distinguished scientists working across a broad range of disciplines in academia, industry, charities and the public sector. This includes many researchers working in health life sciences. The Society draws on the expertise of the Fellowship to provide independent and authoritative scientific advice to UK, European and international decision makers.

While our sister Academy, the Academy of Medical Sciences, is better placed to address the breadth of this inquiry, the Royal Society's recent work on data and machine learning offers insights into proposals within the recently published Life Sciences Industrial Strategy for the handling of healthcare data and potential opportunities for data-driven technologies within the life sciences. The Society has recently conducted work into the needs of a 21st century data governance system and the potential of Machine Learning – a form of artificial intelligence that allows computer systems to learn from examples, data, and experience – upon which this response draws. We would welcome the opportunity to come and speak further with the Committee about these issues.

Our report *Data Management and Use: Governance for the 21st Century*, published jointly with the British Academy, started from the standpoint of a taking a *connected* approach to the governance of data and its uses. New ways of using data and the interconnected nature of digital systems mean that governance frameworks and mechanisms designed for one purpose or application may have implications for its use in another. For example, transport data may inform health choices, or commercial data may be used to target public services. There is great scope for benefit here but also challenges arising from common underlying themes such as privacy, consent, bias and quality. As different sectors grapple with these challenges, it is likely that there is much to be learned from each other. This is relevant to the proposal in the strategy to establish mechanisms to ensure that health data is accessible to researchers and streamlining ethical and governance approaches to facilitate this.

Our report on *Machine Learning: the power and promise of computers that learn by example* set out a number of examples of how machine learning might be used in healthcare to provide more accurate diagnoses and more effective healthcare services, through advanced analysis that improves decision-making.

One example of this function comes from breast cancer diagnosis. Breast cancer diagnoses typically include an assessment by pathologists of a tissue sample, in which doctors look for certain features that indicate the presence or extent of disease. A machine learning system trained on tissue images was able to achieve a higher accuracy than pathologists, by finding and utilising features of the image that were predictive but had not previously been used in the pathology assessments¹. In doing so, the system was able to help doctors more accurately assess a patient's prognosis.

¹ Beck A et al. 2011 Systematic analysis of breast cancer morphology uncovers stromal

Moving forward, the potential for machine learning algorithms to assist doctors is substantial. Tasks such as extracting features from complex data sets like images, ECGs, and other monitoring devices; or spotting patterns indicative of health or illness in individuals from medical records, wearable devices; or combining information from disparate sources to reach diagnoses and treatment decisions, are all well-suited to machine learning approaches. With access to the right kind and volume of training data, machine learning algorithms would be expected to perform well in many of these settings.

The Life Sciences Industrial Strategy recommends that a new regulatory, Health Technology Assessment and commercial framework should be established to capture for the UK the value in algorithms generated using NHS data. Our report noted that, in areas where there are datasets unsuitable for general release, further progress in supporting access to public sector data could be driven by creating policy frameworks or agreements which make data available to specific users under clear and binding legal constraints to safeguard its use, and set out acceptable uses. The UK Biobank demonstrates how such a framework can work. Government should further consider the form and function of such new models of data sharing. It may also be useful to consider where there are opportunities to feed practices and lessons from one context to another. For example, the Administrative Data Research Network (ADRN) is a model which allows accredited researchers to access de-personalised administrative data for social and economic research.²

How to extract value from public sector data will involve debates about how to best promote and distribute the benefits of data management and data use fairly across society while ensuring acceptable level of risks for individuals and communities. It will also require the creation of appropriate mechanisms to apportion value, which will be a social and technical challenge and one that needs to consider how to balance asymmetries of power between different actors

Failure to ensure that models are in place to provide appropriate access to, and use of, these types of data sets, may result in missed opportunities for the UK. For example, the care data centralised records system, which would have seen GP patient records opened to analysis by the National Health Service (NHS) and some third parties, could have provided an invaluable research resource and an important nationally strategic data set.

This report also notes that, for the benefits of data availability to be fully realised, data from research needs to be produced in a way that makes it accessible, so others can find it; intelligible, so it can be scrutinised; assessable, so its reliability can be judged before use; and usable by others. Publishing data in this way can also help increase the impact of research. For example, in a study

features associated with survival. *Sci. Transl. Med.* 3, 108. (doi: 10.1126/scitranslmed.3002564)

² The UK Government handles 1.5 billion transactions with business and citizens annually and analysis of this administrative data can help reduce the cost of public services, increase understanding of socio-economic issues and help make better policy.

of cancer microarray data, the copublication of publicly available data was found to be associated with a 69% increase in citation of the original publication. As data management and data availability become an ever-more integral part of science, the need to bring in specific expertise in handling or processing data, and in preparing it for release, will have implications for the allocation of research funding. While resource costs, such as staff costs, can be considered as part of funding applications to research councils, guidance on the extent to which applications for funding may cover data handling is not clear; while some schemes may offer this, it is not clear that this is always the case.

16 October 2017