UCL

Who owns the data? Open data for healthcare

UCL BRIEFING – SEPTEMBER 2015

AUTHORS

Dr Patty Kostkova
UCL Computer Science
p.kostkova@ucl.ac.uk

Dr Olivia Stevenson
(Acting) Head of Public Policy
Office of the UCL Vice-Provost (Research)
o.stevenson@ucl.ac.uk

Raquel Velho
UCL Science & Technology Studies
raquel.velho.12@ucl.ac.uk

KEY FINDINGS

- Technical advances have made it easier to colllect and analyse information from multiple open sources of data, but this raises challenges for the future of data management. Public mistrust in the use of regulated Electronic Patient Records is in stark contrast to more relaxed attitudes to private user-generated data harvested by social media corporations through wearable technologies.
- It is important to consider safety and privacy of individuals and not just security of data. Striking the right balance between individuals' rights to privacy and the benefits of secondary uses of data for health research is vital.
- There are significant challenges around data transparency – particularly concerning trust, privacy and accountability. Strong disclosure and notification mechanisms to inform the public if data violations occur are important for ensuing accountability and building trust.
- Enhanced public engagement with different populations is needed to support on going dialogue on the balance between access to individuals' data, and the benefits for population health, and to provide platforms for public dialogue.
- Linking unclear raw data to health records, demographic data and genetic information can offer opportunities to uncover population health patterns, predict long-term conditions and identify intervention points.
- User privacy and data ownership remains an under-explored territory from a regulatory and policy perspective.

Introduction

The potential for life-saving information that can be derived from open data is of increasing interest to researchers, policymakers and service providers worldwide. New forms of disease knowledge, including causes and patterns of disease among populations, are being generated from large medical datasets that provide opportunities for more accurate diagnostics, and improved delivery of healthcare and treatments. However, there are serious concerns over Big Data (BD) in healthcare, in particular relating to issues of responsibility and accountability, and around data confidentiality. This briefing discusses the challenges of open data and **how public policy can support a balanced agenda that safeguards personal information whilst enabling the use of data to improve public health.**

DEFINING OPEN DATA

Open data is publicly available data without restrictions of copyright or other mechanisms of control which are structured for usability, but which protect private information. Open data defies a single definition, which has implications for the development of universal policy. Health data poses certain challenges - particularly concerning privacy and identity - that require sensitivity and at times restrictions.

Opening up clinical data for health research

Benefits

As developments in technology increasingly connect people, processes, data and objects online, the opportunities for personal data sharing increase, offering a number of potential benefits:

Mining and analysing large data sources for health research can increase the knowledge base. The priority for any developments in opening up clinical data for health research should be to focus on particular target groups who could benefit most (e.g. patients or healthcare practitioners). Researchers should recognise that for the majority of patients the benefits will be a better understanding and treatment for specific diseases and increased access to information that enables benefits to personal care.

More efficient ways of working for healthcare practitioners: for example a reduction in the time it takes for GPs to provide repeat prescriptions.

Provides information on social inequalities in population health at global, national and local scales: the sharing of large amounts

of data in this way can help researchers to identify accurately geographical areas and target groups for health interventions.

Concerns

There are a number of potentially negative consequences that might arise from open data in health research, including:

Increased health interventions without clinically proven outcomes: there is a risk that research generated from clinical data will be used to aggressively identify under-served patients and disease areas for intervention ahead of the science evidence (as in the case of cervical screening occurring annually in the US).

The re-identification of individuals from detailed data releases: combining de-identified (e.g. pseudonym or coded) datasets significantly increases the risk of re-identification of individuals. There should be opportunities for researchers to feed into the development of a national and international minimum standard for the use of linked datasets for research. Consideration should be given to safety and privacy of individuals, and not just security of data.

Noise risk generated by large data: the noise risk generated by an increasing ecosystem of people giving data signals and searching for different queries can lead to data overload. The quality of analysis, such as data mining, machine learning algorithms, as well as the quality of data is important. There is a need for research that seeks solutions to develop ways to filter on information important for healthcare and which designs secure computer ecosystems.

Merging data from different sources to deliver healthcare benefits

Multiple sources of data

There is not simply one type of data – as a resource, it can come from vastly different sources, have entirely different attributes, and be manipulated in a variety of ways. In the clinical sphere, the amount of computerised patient data is increasing because of the adoption of computer and database management systems for maintaining patient records. Several other sources also contribute to healthcare landscapes; in the US for example, these include: claims and cost data; pharmaceutical R&D data; patient behaviour and sentiment data generated both inside and outside of a healthcare context.

With the increasing use of mobile and wearable devices, there are unprecedented levels of data sharing and collection by corporations largely free from regulation. There are also issues around secret or highly confidential data that cannot be published or made readily available. **Good data sharing standards that ensure confidence in data quality and evidence informed policy can help to encourage making the full range of data available for research.**

Analysing data

Merging data from different sources requires a strong integrated IT skills base and the development of a coherent information architecture (e.g. intercloud) capable of bringing datasets together. It is important to simplify the technical barriers to information sharing within organisations to ensure a comprehensive strategy to capture and distribute data between all appropriate organisations.

Technological advances allow data from a range of sources to be analysed without data merging, providing there is a secure data infrastructure available. **Designing data architecture and governance policy that provides clear strategies for sharing data sources across organisations is essential to break down silos and ensure partnership working.**

Transferring vast amounts of data requires storage capacity and can be time consuming. However technological advances have increased the potential for working with data and running queries remotely. There are a number of technological pitfalls that arise, such as how personal data and/or initial source data can be identified and the capability of technologies to combine datasets from different formats, systems and across countries (e.g. data harmonisation). The latter is important because only sources of standardised data will generate increased insights and not all data can be harmonised.

Balancing access to data with patient privacy

Public attitudes to healthcare

Changing public attitudes to healthcare delivery and the shift in GP-patient relationships has led to public concerns for healthcare services. Greater use of digital systems in consultations has significant implications for the delivery of healthcare and the GP-patient relationship – for example, an app used to monitor glucose levels can be combined with traditional medical records. This could help to inform clinical decisions and patient treatment.

However, the use of open data for healthcare has failed to gain public trust. Perversely, citizens appear less concerned with the use of potentially private and unregulated health data collected directly by MedTech manufacturers* through tracking/wearable devices and social media companies.

Patient privacy

It may be helpful to review what is meant by patient privacy in order to ensure standards are fit for purpose, in the context of technological and open data developments. Common public fears around open data coalesce in worries that personally identifiable data may be misused or inappropriately shared and that data may be inaccurate or contain information potentially harmful to individuals. Similarly, the reluctance to share information stems from a fear of data being used against individual patients to impact on their insurance cover and future access to insurance, rights to benefits and to invade personal privacy.

The expected traditional standards of consent, protection and privacy need to be balanced with rights to use data, under what circumstances and trust. Developing improved approaches to patient consent and risk-based assessments of clinical data usage for research is a priority. Public engagement that emphasises to different populations the required balance of access to individuals' data and the benefits for population health is also needed. The challenge for policy is to ensure an appropriate level of anonymisation of health data for public safety that still allows meaningful usability. This requires guidance on data sharing, privacy and access that is demonstrably secure but also transparent to patients and citizens, in order to ensure trust in the system.

* For a description of MedTech, see EU Medical Devices Directive (2012, pg. 5): http://ec.europa.eu/health/medical-devices/files/meddev/2_1_6_ol_en.pdf

Developing Big Data and open data for health

It is important that public policy engages with developments in active data and knowledge sharing. BD and resultant knowledge should be shared for the purpose of delivering better health outcomes. Failure to engage fully with the information-based culture already underway risks policies for data sharing stagnating, with the potential for broad impacts consequently reduced. Clear policies are needed to govern research access to clinical data sources and facilitate their use for evidence-informed learning in health. There are a number of issues to be addressed:

Public and citizen engagements: public awareness campaigns about the benefits of sharing data can have far reaching effects, and should be based on expert scientific and empirical methods. Such outreach will ensure a feedback loop from the public towards policy development and align impacts to produce greater public benefit. The priorities should be the development of ongoing dialogues between the public and private sectors, and building public trust in data use for research.

Transparency: data transparency requires coordination and commonalty of purpose. It is important that open data and data transparency policies are fit for purpose, and can manage the links between BD and open data for health research. Transparency efforts should focus on making stored data usable, searchable, and actionable for research, and provide data catalogues that index the suitability, sustainability and scalability of data at national and international levels.

Clarity on use of data: the right to withhold or withdraw consent for data use is not well understood, and there is insufficient clarity on how data can be used. The challenge is to ensure ethical and informed data use policies are generated. Consent is complex, but the gold standard should be appropriate informed consent provisions that integrate a flexible opt-out clause, to protect privacy and encourage public trust.

Training and education: citizens should be equipped with technological, ethical and efficiency skills that enable future workforces to take advantage of the open data revolution. Creating centres of excellence with a focus on training scientists in the use of open-source tools for data analysis should be a priority. The teaching of ethical and moral codes should be embedded within training and education frameworks. Further, training should take place to enable communities to apply data to solve local problems.

Data structures: common coding and interoperability standards in information architecture, based on principles for achieving consistent, coherent applications and databases, should be developed to benefit patients, practitioners and clinicians at every level. The development of a systematic approach for how patients and researchers access open data is important and will require greater synergies between IT developers and researchers to develop complementary research agendas.

Data safe havens: the development of independent not-for-profit data registries should take account of research needs to ensure that health BD is provided in machine-readable formats for the purpose of research e.g over an 'API' (application program interface). However, data architectures should ensure that researchers can receive data in an anonymous format that does not violate privacy protections.

BACKGROUND

This briefing has been developed from a roundtable discussion held at UCL in 2014 as part of the **UCL Festival for Digital Health**, with representatives from industry, policymakers, healthcare practitioners and patients. The roundtable included academic participants with expertise in: computer science, engineering, medicine and health service delivery.

The Festival for Digital Health 2014 was a cross-disciplinary initiative chaired by Dr Patty Kostkova (UCL Computer Science) and hosted by the **UCL Grand Challenge of Human Wellbeing**. It was organised in partnership with the **UCL Institute of Biomedical Engineering**, **UCL Centre for Behavioural Change**, **i-sense** and **UCL Advances**.

PARTICIPANTS IN THE UCL ROUNDTABLE DISCUSSION

Peter Knight, Deputy Director Research Information and Intelligence, Department of Health
Edward Fottrell, Lecturer at the UCL Institute for Global Health
Ben Goldacre, Research Fellow in Epidemiology, LSHTM
Ralph Sullivan, GP and Chair, Health Informatics Group, Royal College of General Practitioners
Emma Ross, Senior Consulting Fellow, Chatham House Centre on Global Health Security
Corinne Marsolier, Director, Cisco Consulting Services, Life Sciences, Health and Care
Patty Kostkova, Principal Research Associate, UCL Computer Science, and the Chair 1st UCL Festival for Digital Health
Graham Hart, Dean, UCL Population Health Sciences
Angela Sasse, Professor of Human Centred Technology, Head of the Information Security Group and Director for Industrial Liaison, UCL Computer Science
Rachel McKendry, Director i-sense, the EPSRC IRC in Early Warning Sensing Systems for Infectious Diseases
Phil Koczan, Chief Clinical Information Officer, UCL Partners
John Tooke, Vice-Provost (Health), UCL
Helen Brewer, Postgraduate Fellow, Parliamentary Office of Science and Technology
Simon De Lusignan, Director RCGP Research and Surveillance Centre

UCL Public Policy