phgfoundation

making science work for health

What is citizen generated data?

Historically data about our health has been predominantly produced in a healthcare environment. In today's digitised world, however, individuals are now generating increasing amounts of health-related data outside the healthcare setting, either intentionally through the use of health tools including fitness trackers or home monitoring devices, or passively through environmental sensors and online activity. There is growing interest in harnessing this citizen generated data (CGD) for informing health-related predictions and clinical decisions. This briefing note explores what CGD is, how this data could be utilised for health purposes, and the considerations arising as the volume, velocity and variety of CGD continue to grow rapidly.

In brief

- Health-related CGD may include data produced specifically for health purposes, or for lifestyle goals, or simply by our interactions with a 21st century digitally enabled world

- Increase in CGD has been stimulated by many factors, including the ubiquity of smartphones, our growing reliance on the internet, and growing interest in the 'quantified self'

- Sources of CGD related to health include smartphones, health devices, internet of medical things, online activity, environmental tracking and direct-to-consumer tests

- The use of CGD is being explored for a range of purposes including using social media data to investigate food poisoning outbreaks and data from online forums to gain insights into mental health

- CGD has the potential to empower individuals, enable citizen science, personalise healthcare and support prevention of disease, but raises questions for health systems, policy makers and health researchers

How are citizens generating heath-related data?

The ubiquity of smartphones, our increasing reliance on the internet, and growing interest in the 'quantified self' and the associated proliferation of wearables, self-monitoring devices and direct-to- consumer health products are all stimulating the expansion of CGD. Health-related CGD may include data produced specifically for health purposes, or for lifestyle goals, or simply by interacting with a 21st century digitally enabled world. Sources of CGD related to health include:

Smartphones

This is currently the most dominant and rapidly expanding area of innovation in digital health. Countless smartphone health apps aimed at improving physical and mental health through monitoring, education or encouragement are available. Apps may utilise user-inputted data (e.g. self-reported mood), or data gathered through the phones in-built sensors:

- Gyroscope (e.g. activity and sleep)
- Camera (e.g. diet and mood)
- GPS (e.g. location and environment)

Health devices

A large number of devices for measuring various aspects of an individual's behaviour and physiology have emerged in recent years. These include wearables with built-in sensors (e.g. heart rate monitors), implantable sensors, and self-testing devices. The devices are typically designed for specific disease management (e.g. diabetes and other chronic diseases), fitness tracking or health monitoring (e.g. fertility checkers).

Internet of medical things

Some internet-connected devices can network with other devices, smartphones, computers, or web platforms – in an 'internet of things' (IoT) communication system. This typically allows data to be stored, analysed, displayed, and even annotated by the user.

Online activity

Internet searches, online shopping, social media activity, patient platforms, discussion forums and website usage could potentially provide proxy measures for an individual's health. The data could also be used for analysing population health, including disease surveillance, predicting infectious disease outbreaks and tracking their spread[1].

Environmental tracking

Smart devices, CCTV, shopping transactions, utility usage and environmental sensors are all potential sources of data about behaviours that could be relevant to health.

Direct-to-consumer (DTC) tests:

Tests marketed directly at consumers are allowing individuals to access information about their genetics and blood biochemistry. Improvements in sequencing technology and our understanding of genomics will probably result in an expansion of DTC tests and information available.

The rise of citizen generated data

The growth in consumer-facing health technologies and digital devices, and the resulting citizen generated health data raises important questions for health systems, policy makers and health researchers

The potential

Empowering individuals: to better manage their health by giving citizens more information about their behaviour, health and lifestyle and to inform behaviour change or other interventions

Enabling citizen science: by crowdsourcing data from digital and online platforms to inform research and enrich datasets for health research

Personalising healthcare: by better understanding the extent of variability within individuals and the differences between individuals by capturing health data on an ongoing basis and outside of the healthcare setting - rather than only at limited points in time and during clinical interactions

Supporting prevention: through more comprehensive and intensive data capture that may help identify early warning signs of declining health

Citizen generated data in action

Besides individuals gaining information about their own health through wearables and data-driven digital tools, the use of CGD is currently being explored for a range of purposes:

- **Social media data** – as a tool to investigate food poisoning outbreaks[2] and mental health decline[3]

- **Sensor technology** – projects are underway to predict declining health and inform clinical decisions by monitoring activity in the home

- **IoT** – technology is integrating data from self-monitoring devices and apps with data from health records in pilot projects for prevention services

- **Wearables and apps** – some hospitals are working towards integrating data from digital tools into their electronic health records (EHRs)

- **Online forums** – work has taken place to examine the use of big data analytics on online-user generated content for insight into mental health[4]

The questions

Health system impact

The changing dynamics of how and where health data are generated are creating an ecosystem where individuals could generate and hold a greater volume of data related to their health. How will the health system need to adapt? How will the increasing availability of data impact on citizens interactions with - and expectations of - the health system?

Data utility

What types of CGD would be useful to collect for individuals, healthcare and research, and are these datasets of sufficient quality to inform these applications? If CGD were to be collated for healthcare, how and by whom should these datasets be analysed and interpreted?

Data integration

Using CGD to improve insight into health will require disparate sources of data to be linked. There are some aspirations for citizens to be able to add information into their EHRs from sources such as wearable devices. How will a health system, which is currently in the process of undergoing widespread digitisation, cope with integrating data from a vast number of devices, different sources, and varying formats? Are there implications for patient safety?

Data and device regulation

How will the General Data Protection Regulation (GDPR) impact on processing CGD? Will expanding the scope of health data to include CGD create barriers to optimal data use for health applications? Should devices and software that are generating or processing CGD for health applications be regulated as medical devices?

Privacy protection

Could CGD be accessed by industry or others in ways that could breach personal privacy and violate human rights? What safeguards or protections might need to be put in place to protect vulnerable adults or children?

User acceptability and user appetite

What is the user acceptability for CGD to be made available for healthcare or research? Do users want rights over the use of CGD even if it is in the public domain? What is the long-term user appetite for health, lifestyle and fitness monitoring, and what factors motivate or discourage continued use of monitoring devices?

The fundamental question is whether, and how in practice, CGD can be used to advance our understanding of health and disease, support better healthcare and improve citizen health.

1. Yang S, Kou S C, Lu F et al. **Advances in using Internet searches to track dengue**. PLoS Computational Biology. 2017; 13(7).
2. Harris J K, Hawkins J B, Nguyen L et al. **Using Twitter to Identify and Respond to Food Poisoning**: The Food Safety STL Project. Journal of Public Health Management and Practice. 2017; 23(6): 577–580.
3. Reece A G, Danforth C M. **Instagram photos reveal predictive markers of depression**. EPJ Data Science. 2017; 6(1).
4. Smith J, Bartlett J, Buck D, Honeyman M. **Investigating the role of public online forums in mental health**. Kings Fund. 2017.