CHATHAM HOUSE

The Royal Institute of
International Affairs

Strengthening Data Sharing for Public Health

23 April 2015

Introduction

Recent public health crises, such as the Ebola outbreaks in West Africa, have shown that averting or mitigating the impact of global health emergencies relies on efficient and high-quality data sharing across national borders. However, there is at present no global framework or operational guidance for systematic sharing of public health surveillance data. A number of specific barriers to data sharing have been identified, but in order to achieve good practice these barriers must be overcome collectively.

The *Strengthening Data Sharing for Public Health* project is an initiative of the Centre on Global Health Security at Chatham House. It seeks to inform and support global policy and action to shift the norms towards a model where data are shared as openly as is possible and appropriate, by developing guidance on how to create the right environment for, and achieve good practice in, data sharing.

This document is a summary of discussions that took place during a high-level Strategy Roundtable convened at Chatham House on 23 April 2015. The aim of the meeting was to stimulate a high-quality dialogue among experts from a broad range of data sharing environments to help inform the project and its strategy going forward. The specific objectives of the roundtable were to consider the best approach for the project, how it would take shape, whom it should involve, and which areas of data sharing are most appropriate and necessary to tackle to ensure success. The meeting was held under the Chatham House Rule.[1]

General discussion

The meeting opened with summary presentations of two research papers,[2,3] written to inform discussion at the meeting. One considered the barriers to public health data sharing and the potential means of overcoming them, while the second outlined key data sharing lessons from other sectors.

Which data should the project address?

Central to discussions was the need to specify the project's definition of 'public health data' and, in so doing, to determine the parameters of the project: whether its focus should be on research data, routine surveillance, or a mixture of the two. With other key actors already focused on sharing research data, participants considered that it would be most helpful to focus on surveillance, which is of great value during a public health crisis. However, the point was made that the boundaries between research and surveillance data are often blurred: there are many examples of research data being used as a surveillance resource to inform public health action and response, and of surveillance data being used for research purposes.

The project should be clear about what types of 'sharing' will be included – whether it will be from the local level to the global level, sharing across borders, or between different sectors (e.g. from public health to the private sector, and vice versa). It was emphasized that the greatest need lies in the data sharing that

---

[1] When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed.

[2] Sane J, Edelstein M. *Overcoming Barriers to Data Sharing in Public Health: A Global Perspective*. Chatham House Centre on Global Health Security Research Paper; [Online] London: The Royal Institute of International Affairs. 2015. Available from: http://www.chathamhouse.org/sites/files/chathamhouse/field/field_document/20150417OvercomingBarriersDataSharingPublicHealthSaneEdelstein.pdf [Accessed 31 May 2015].

[3] Brack M, Castillo T. *Data Sharing for Public Health: Key Lessons from Other Sectors*. Chatham House Centre on Global Health Security Research Paper; [Online] London: The Royal Institute of International Affairs. 2015. Available from: http://www.chathamhouse.org/sites/files/chathamhouse/field/field_document/20150417DataSharingPublicHealthLessonsBrackCastillo.pdf [Accessed 31 May 2015].

must occur directly between countries, and from the country level to the global level, particularly during public health emergencies.

There was brief consideration of whether raw or aggregated data would be most suitable for inclusion. However, it was felt that the specific public health question being addressed would largely determine the most appropriate format of the data on a case-by-case basis. It was recognized that not all data are equal, and different data have different sensitivities that require different management processes.

There was general consensus that the next step should be the development of the project definition of public health data, to refine the project's scope and clarify it for all interested parties.

The value of data sharing

Participants proposed that the project should include an element that explains the value of public health data sharing to those producing and using the data, and to the wider community. It was generally felt that one of the best ways to do this would be to identify and disseminate case studies that illustrate the benefits of data sharing. It was also considered important to highlight situations where negative outcomes were expected but did not occur. It was suggested that time and money should be invested in identifying what the next large-scale data sharing success might be, and in ensuring that this is publicized widely in order to stimulate the debate on data sharing.

It was mentioned that in places where data sharing is not the norm, it would be beneficial to communicate that data sharing is well established and that the project is principally attempting to find new ways to make this a fairer, more systematic and more timely process. In such areas, it would be important to engage key stakeholders at the policy and operational levels in order to ensure that guidance coming from the project is owned, endorsed and implemented.

The lack of awareness of existing data sharing platforms was identified as a problem, for instance with data being readily available on the World Health Organization (WHO) website but not necessarily actively publicized. An analysis of gaps in the knowledge and communication pathways concerning data access was considered to be required.

Incentivizing data sharing for the individual

Collectors of data may be reluctant to share them for fear of forfeiting benefits, such as credit from publication. Providing incentives to share data was a significant point of discussion throughout the meeting, and there were good examples, including from other sectors. In the area of genomics, recognized as a leader in data sharing, the mandate for sharing has come both from geneticists themselves and also from funding agencies, which require – as a condition for receiving funds – researchers to articulate a data sharing policy . In the field of public health, DataFirst, a South African data service dedicated to making survey and administrative micro-data available to researchers and policy analysts, offers data-cleaning and improvement services to encourage sharing.

One of the most commonly cited barriers to data sharing was the pervading 'publish or perish' paradigm, whereby those who collect data are pressured to publish academic papers in order to build and maintain their careers. Although journal editors had indicated they would not want data that could benefit public health to be held back for reasons of publication, there is currently no process to ensure verification and appropriate attribution in situations where data are made available prior to their official publication. An alternative norm to 'publish or perish' – 'publish data or perish' – was suggested. It was noted that there

are now methods for data collectors to publish data sets in peer-reviewed journals, and that both the *New England Journal of Medicine* and *The Lancet* are in discussions about data peer review and online publication, with a view to improving ease of access to and distribution of health data.

Just as incentives would be needed to drive public health data sharing, swift and effective penalties would be required in the case of data misuse. Further discussion was needed on the nature of such penalties, and on how and through whom they might be enforced.

Building and maintaining trust between data producers and data users

Health is deeply personal and has implications that extend to other spheres. This must be taken into account. The need to build trust at all levels was therefore a recurrent theme during the discussions: at the level of the individual who provides confidential personal data, in respect of ethical management and protection; at the level of those who produce and share research and surveillance data, in relation to appropriate secondary use; and at the level of those who gain access to such data with expectations of high quality. It was important to recognize the different sensitivities, such as the potential restrictions on trade and travel and the risk of de-anonymization, that exist with health data when seeking to further global public health data sharing.

To assist with building and maintaining consistency and trust between health data producers and users, it was suggested that a certification system covering organizations that have appropriate data management practices in place could be an appropriate way forward. Certification would necessarily entail the development of standards on data quality and management, and of data sharing agreements to which all certified bodies would adhere; in order to maintain certified status, these would be subject to regular mandatory checks.

The introduction of a certification process would require the creation of a group, most likely at the international level, dedicated to developing and maintaining data practice standards. The International Council for Science: Committee on Data for Science and Technology (CODATA) was mentioned. CODATA's sole purpose is to promote the sharing of data, and its activities include certifying data repositories – which must meet legal and ethical stipulations – as an incentive for maintaining data quality. The Data Seal of Approval, a Dutch initiative, also certifies digital research data repositories and includes data curation among its activities. These have not yet dealt with health data, but have developed around social science data, which has some similar sensitivity issues.

The role played by in-country ethics review boards was also noted, and the value of involving such boards throughout the project emphasized. These vary greatly, which can have an onward impact on the sharing of data across borders, and they also reflect a variety of different perspectives that may be encountered during data sharing negotiations.

There was mention of the need for measures to ensure the trust and protection of individuals providing confidential data, specifically around the failure of de-identification of health records, or anonymization. However, this aspect was not covered in depth.

Data sharing frameworks

Consideration was given to the possible formal mechanisms that could be put in place to ensure effective and efficient public health data sharing. The distinction was made between enforceable laws and broad, flexible legal guidelines and principles that complement national legislation. Although legally binding

structures were generally considered important at international and national levels, their value with regard to wider public health data sharing was not so clear. Some participants thought a legally binding mechanism would be necessary, others suggested it would be unrealistic, and the consensus was that the creation of a set of guidelines and principles would be more appropriate for this project. There is potential to build from the International Health Regulations and from the Pandemic Influenza Preparedness (PIP) Framework (which promotes virus sequence sharing and benefit sharing) to develop a new framework of evolving guidelines and principles specifically around public health data sharing – which could be endorsed by the WHO and other international bodies.

National sovereignty is considered a rigid barrier to data sharing at an international level, as seen in a number of recent disease outbreaks. The EU, with its public health data access policy spanning 28 countries, was seen as a valuable case study for cross-border data protection and data sharing processes. The EU was able to overcome national sovereignty on data ownership by defining case-based surveillance data in a way that made clear it could only be owned by the individual and not by the state. The European Centre for Disease Prevention and Control (ECDC) – the primary EU body responsible for health data control and processing – is thus entrusted by individuals. Participants noted that this system made the negotiation around sharing data much simpler and required EU member states to share their national data automatically with the ECDC.

Several practical formal data sharing options were presented to the roundtable, with a focus on avoiding case-by-case agreements – which can be time-consuming and convoluted. These options included developing a standard set of global licences that could cover a number of producer-user relationships (national to global, private to national, etc.). Although the widespread adoption of data sharing licences can take several years, uptake by key international bodies could accelerate this process. Another option mentioned was the introduction of a patent pool, as used by various industries, to deal successfully with intellectual property issues. A patent pool is an efficiency measure whereby at least two companies agree to cross-license patents relating to a particular technology.

It was suggested that checklists could be a simple way to ensure effective and efficient data sharing between two parties, when accompanied by a context-specific memorandum of understanding. Such checklists would allow data producers to provide guidance on the data and ensure adequate quality, and for data users to identify why they want the data and how they intend to use them.

Public health data from other sectors

It was noted that private companies collect and produce increasing amounts of data that are suitable for use in public health surveillance and action. Furthermore, the public health community is only just beginning to appreciate the potential of social media for picking up new significant public health events before they are identified by routine surveillance and analysing them as they happen. However, these data are not yet widely incorporated into the broader surveillance landscape, they are not nationally owned, and there are accompanying intellectual property issues. Questions arise over ownership (for example, does the individual own the personal data they provide on Facebook, or does Facebook?), and over how using such data for public health may differ from commercial usages such as tailored advertising. Participants felt that while these novel methods of surveillance are exciting, they could distract from the larger purpose of the project. Nevertheless, it would be necessary to take them into account for the project outputs to be relevant and useful.

It was pointed out that guidance to inform the ownership question – if a private entity collects data in a country, who owns the data and how can the data be used? – would be useful both for those producing and for those using data.

Trade and travel

Participants recognized that an important area in public health data sharing is trade and travel, and the manner in which public health problems can lead to major economic consequences for both countries and private companies. While the WHO rarely recommends trade and travel restrictions during times of public health emergencies, governments and companies operating in the affected area do not generally heed this advice, and the consequences can be far-reaching. This is particularly apparent with food-borne disease outbreaks, where better data sharing could quickly identify the affected area, leading to more appropriate restrictions on food exports when necessary. Investment banking and hedge funds have shown interest in infectious disease control for business reasons.

Ministries responsible for tourism were identified as important actors to engage. They are affected by the reporting of public health data, as any ongoing outbreak might have implications for leisure and business travel and thus corresponding effects on local economies.

The benefits and solutions for health data sharing extend beyond the realm of public health. The business case for good public health data sharing is strong, and it will be important for the *Strengthening Data Sharing for Public Health* project to engage and work with stakeholders involved in trade and travel to explore their risk perception and understanding of surveillance, and to reduce any risk of disproportionate economic impact of actions taken.

Engaging regional stakeholders

The view was expressed that a central agreement applicable to all players could be most expedient, but an alternative opinion was that this could overlook the nuance required for context-appropriate negotiations.

Many cross-border data sharing networks are already in place across geographic regions, and it was emphasized that convening roundtable meetings of regional stakeholders to examine regional concerns surrounding data sharing policy would be vital to the success and value of the project.

Field epidemiology training programmes, which tend to have a regional focus, have been essential in facilitating current regional networks. Regional surveillance, epidemiological expertise and data sharing tend to rely on individual relationships, putting data sharing measures that are in place regionally at risk when individuals move on. Without formal guidance and processes, data sharing practices can change depending on the personnel involved or as a result of top-down interference. By inviting strategic and operational-level staff to discuss the individual barriers to data sharing from a regional and country-specific context, future roundtables should serve to stimulate an open dialogue on the topic while also allowing for the development of regional frameworks. A consolidated global framework could potentially be developed thereafter.

When discussing regional networks, it is clear – particularly following the West Africa Ebola crisis – that strong institutional capacity is essential for both country and regional surveillance and response. It was mentioned that system strengthening and capacity-building in low- and middle-income countries will be key to any successful data sharing policy because weak institutions will neither trust data from others nor produce good quality data for sharing.

Although long-term plans should ensure continued success, it was suggested that in the short term a focused group of technical experts, under the auspices of the WHO for instance, could consult with local and regional institutions to train local data producers, simultaneously introducing the effective data sharing principles that this project is aiming to produce.

Moving the *Strengthening Data Sharing for Public Health* project forward

There was clear consensus towards adjusting the project plan to accommodate a two-pronged approach: holding roundtables on specific themes and in specific regional hubs; and meanwhile pursuing a parallel, iterative process – through pilot implementation and key stakeholder feedback – to develop and test an evolving, dynamic set of data sharing guidelines and principles based on what is already collectively known and understood and so is useful in practice.

The project should focus on outreach, incentives, high-quality and trustworthy data, and data management that respects the rights of producers and increases access to users. The project's broad approach – involving key experts outside the public health field, and incorporating dimensions such as the economics of trade and travel that are often lacking in public health data sharing discussions – will ensure that it remains relevant, up to date and forward-looking.

Next steps determined from the meeting and follow-up discussions:

- **Establish essential principles for data sharing.** Based on the Strategy Roundtable and research conducted to date, Chatham House will produce a draft set of essential principles for public health surveillance data sharing for review by Strategy Roundtable participants and other key stakeholders. These will include practical examples for implementation, and are intended to cover critical actions involved in sharing data within regional cross-border networks, north–south collaborations, and locally in-country.

- **Create pilot guidance document.** The essential principles for data sharing, with feedback from key stakeholders, will form the basis of an outline for guidance on data sharing. This guidance will be produced as a first iteration and piloted before the end of 2015. Subsequent iterations will be produced and tested prior to launch towards the end of 2016.

- **Initiate pilots.** The draft guidance will be piloted within a number of data sharing contexts globally. Such contexts could include a high-income country setting, a low-income country setting, and a north–south collaborative setting. Receiving feedback from these pilots will allow further refinements to suit the needs of users globally.

- **Convene targeted thematic and regional roundtables.** As the pilot guidance is developed, areas of data sharing requiring expert input and meriting targeted thematic roundtables will be identified. Such examples may include the international data protection legislative landscape, the economics of trade and travel, and the role of 'big data' in public health surveillance. Regional roundtables will engage stakeholders at a local level.

The intended outcome of the project will be a functional, flexible and systematic framework through which to create the right environment, and achieve good practice, for public health data sharing. It is anticipated that the successful products of the project will be endorsed at both the national and international level. However, it was emphasized that while the leading international authority in public health is the WHO, it is not necessary to wait for endorsement by the WHO before proceeding with

implementing effective data sharing practices. Data sharing should be recognized and promoted as a common good that is happening now.