

Research Paper

Emma Ross

Centre on Global Health Security | April 2014

Perspectives on Data Sharing in Disease Surveillance

Background

Public health surveillance is the systematic collection, analysis and interpretation of health data and the timely dissemination of the resulting information to those who need it for planning, implementing and evaluating public health action.

Having data, and putting it to use, is important for effective surveillance at local, national, regional and global levels; and for informed decision-making to implement and evaluate disease control programmes.

However, knowledge of the global distribution of disease, particularly infectious diseases, is patchy at best, making public health decision-making more difficult and uncertain. Part of the reason is that not enough data are collected and/or reported. But the failure to clean, analyse, and use data that are collected in countries with a high burden of diseases – data of interest to both the affected countries and the global health community – is also considered a significant problem that thwarts the potential of surveillance data to inform public health planning and action.

Some of the surveillance data being discussed in international forums about data sharing is data collected by small research groups in developing countries. One major research funder estimates that 90 per cent of data collected in projects it funds are never cleaned or analysed and that for the 10 per cent that are, it often takes a painfully long time for any papers to be published. Data management and analysis plans are often only half-conceived. Some donors say this is a frustrating, expensive and inefficient way of funding public health research. While data collection, analysis and publication are more established in developed countries, there are sometimes methodology problems there as well; and in both developed and developing countries, data sets of interest are present in varying quality, often untapped and locked.

One way to unlock the potential of this data, to be used and to expand the ability of data sets that are of satisfactory quality to inform public health decision-making, is for data sharing to become the norm. While it is becoming increasingly common to share data from public health surveillance, it is not yet as well established as it is in some other scientific fields such as genomics and physics.

There is a growing demand for surveillance data, a growing recognition of the potential of data sharing to advance public health, and agreement that more can be done to open up access to this health-related data. Advances in information technology have made it technically more possible to share; and there is greater willingness to discuss it, a wider movement for open access science, and a push to step up progress in global health specifically. Furthermore, some data-sharing initiatives in surveillance have demonstrated their benefits. A new generation of research networks spanning several countries has sprung up over the last 10 years and many of them engage in data sharing. Some will last, while others will fade, but what has proved important is inclusiveness and equal participation.

The advent and growth of more informal, more open, sources of data for disease surveillance have shown the potential of advances in technology to increase data sharing in this field and have increased the appetite for it. In the absence of progress in sharing of more official purpose-collected surveillance data, alternative informal sources of surveillance data such as ProMED-mail and

HealthMap have carved out a role for themselves in global disease surveillance. This has led to a sense among some that a moment of opportunity now exists. This opportunity, combined with a mandate from revision of the International Health Regulations that officially lets the WHO accept information from sources other than countries, has resulted in a louder call for more widespread data sharing in public health.

This paper was prepared to inform the Chatham House workshop on 'Data Sharing in Disease Surveillance: Experiences and Vision', held on 12–13 February 2014, and was based on interviews with more than 25 stakeholders in disease surveillance data sharing, representing primary data producers, secondary users, research funders, international organizations and other relevant groups. As the workshop was held under the Chatham House Rule and most interviewees attended as participants, comments have not been attributed.

Why share data?

Proponents say large-scale data sharing in disease surveillance would greatly enhance and accelerate understanding of diseases and their patterns, whether through the sharing of routine surveillance data or through real-time sharing of outbreak reports. Some examples have demonstrated this. For instance, routine data sharing across the WHO-coordinated Global Influenza Surveillance Network (see Box 1) contributed to the confirmation of the SARS outbreak in 2003; while the sharing of more informal surveillance data through ProMED-Mail and the Global Public Health Intelligence Network is credited with its detection, and the global response that followed, by flagging an outbreak of severe respiratory disease in Guangzhou, China, long before the government reported it officially. These two informal surveillance networks served as a 'safety net', detecting a disease outbreak that either had not been identified as a global threat in the country where it was occurring, or was not promptly reported.

Box 1: The Global Influenza Surveillance Network and data sharing

The Global Influenza Surveillance Network, which has been sharing laboratory diagnostic and biological data on influenza since 1952 in order to monitor flu activity across the world, has enjoyed a high level of data sharing for many years. Real-time sharing of data is part of the terms of reference for network membership, which consists of government-nominated institutions in 111 countries. The reporting system is web-based and the aggregate surveillance data are publicly available on the internet.

Part of the success of the network's data-sharing efforts have been attributed to the fact that its data-sharing culture and practices took root long before discussions around legal issues, standard operating procedures and benefits sharing appeared on the global health political agenda. Data-sharing success is also put down to the public-health spirit of the members trumping concerns about benefits or career advancement through publication, and to a sense of certainty that the network and its continued functioning are an asset to the world.

Another example of the benefits gained from the sharing of routinely collected data can be seen in research conducted by the ALPHA network to assess the effect of HIV infection on pregnancy-related mortality. Maternal mortality is a rare enough event that researchers from a single site

would not have been able to answer the question alone, but the sharing of data across the network revealed the elevated risk of HIV infection for pregnant women.

Other potential benefits of data sharing in disease surveillance cited include advancing innovation, reducing duplication, enabling further analysis and smarter investments. From the point of view of the data producer specifically, some benefits of sharing include learning new uses for the data, gaining insights from other researchers' interpretations of the data, becoming aware of biases in the fieldwork, the enabling of comparisons, career development, and in some cases, the availability of funding. For governments of countries with weak data analysis and interpretation capacity, potential benefits include more valid information to guide disease control programmes and priority setting, reducing the need to rely on modelling to guide policy, and external funding and technical support.

Visions of data sharing in disease surveillance

There are various visions of what type of data should be shared and how openly it should be shared. They range from sharing nearly all anonymized data in freely available unsupported internet databases (open access), to increasing the frequency of/normalizing informal sharing between trusted groups, and expanding beyond data sharing to knowledge sharing and investing in capacity-building so that researchers and governments in countries with weaker data collection, management, analysis and interpretation systems can gradually become self-sufficient in this area.

There are some who support the idea that every data set should be made openly accessible to everyone on the internet, anonymized, fully documented and packaged with metadata, methodology, coding and any other information necessary so that secondary users do not have to be in direct contact with the data gatherer in order to understand how to use it. Proponents of this vision say the US National Health and Nutritional Examination Survey (NHANES) should be the model.

Some agree that universal open access to anonymized surveillance data of all kinds is the ideal, but acknowledge that there are a number of reasons, discussed later in this paper, why that is not currently feasible.

Others feel that open access should be the default, but with exceptions, such as when it would not be worth the time and expense to make data available, or when the sensitivities are too great.

Most others feel that open access is not appropriate for most public health data and that it will never be feasible. It has been proposed that the policy should be that surveillance data be opened to the widest possible group without violating legal, ethical or privacy concerns, but agreeing what that translates to in practical terms could be challenging.

Most data producers and secondary users agree that there are gradations of openness that are legitimate. They argue that efforts to increase data sharing in disease surveillance should adopt a mixed approach and that thought needs to be given to what the reasonable constraints should be data sharing in these cases. Most agree that different types of data require different levels of sharing, depending on the question being investigated and/or the sensitivities surrounding release

of the data. There are arguments related to social justice, sovereignty, economics and privacy that many consider to be valid and important to discuss. Some see the current period as a transitional time for access to data.

Most believe it is reasonable that a few data sets be openly available on the internet, but that in other cases secondary users should have to apply to the data producers to discuss usage, with the expected answer being yes and the discussion being about how. In those cases, sometimes it should be collaboration, or sometimes secondary users could be invited to come to the data producer's centre to use the data, but not take it away.

For aggregate national surveillance data, most believe that data sets should be as open as possible, with some advocating open access in publicly available repositories and others arguing it should be made easily available to bona fide researchers or others who can show the need to know and negotiate terms. This is considered the easiest type of data to share. Experts say the sharing of individual case data is more problematic. In addition, it is argued that community-based surveillance data are intrinsically less shareable than data from surveillance based on a single interaction, and that the need to sustain the relationship with the community being studied breeds a different attitude to data sharing than in surveillance based on, for instance, screening people for infections responsible for ongoing outbreaks before they board an aircraft.

Others argue that embarking on an exercise to understand what data are available, and making the results of that investigation available to everyone, would be a huge step. It would then be easier to determine which datasets are best to target for sharing, once it is established which public health questions are the priorities to answer. Some argue that thought also needs to be given to whether some data sets are worth sharing, whether because their use is limited, because the audience is narrow, or because it would be too costly and problematic to resurrect and repair, and that those should be weeded out because the effort and expense to make them available are not warranted.

Policy-level moves

There has been some activity by research funders and others in a position of influence to move the data-sharing agenda forward on a policy level.

One prominent example is the 2011 joint statement emerging from a workshop of major public health research funders, international organizations, large research studies and journals on how to increase the sharing of public health research data. However, it has been noted that the statement, which has 19 signatories to date, is heavy on institutions from the global North but light on institutions from the South and that more progress can be made in that area.

It set out principles and goals for data sharing in the field and announced the establishment of working groups to further them. Projects being undertaken by the working groups include a capacity-building report to investigate what training was available for data management in low- and middle-income countries and how to expand it, a data discovery exercise to investigate the potential for cataloguing what data are in databases, the building of an evidence base on stakeholder views and expectations regarding data sharing, and research on what incentives are needed to increase data sharing.

Major funders such as the Wellcome Trust, the Bill & Melinda Gates Foundation and the US National Institutes of Health (NIH) are now applying data-sharing requirements to large grants, with many requiring applicants seeking significant sums of money to submit data-sharing plans and expecting data to be made available in a timely manner to the broader research community, sometimes in specific repositories. The level, mechanics and timing of access to data are usually negotiated on a case-by-case basis, and funding to make it possible can be added to the budget. Funders have reported that grantees sometimes push back on some details, but those elements are usually negotiated successfully enough to achieve data sharing.

In another notable move, the humanitarian aid group MSF announced a data-sharing policy in 2013, raising hopes that more players in the humanitarian crisis field might also be ready to share data on surveillance.

The Human Genome Project: a data-sharing pioneer

The publicly funded Human Genome Project, launched in 1990, is considered a pioneer in data sharing in the health field, with scientists working on the project agreeing, against their professional grain, to put their data in the public domain before publication in a scientific journal. Sequencing centres agreed that every evening, the data produced that day would be made freely available on the internet.

Commentators argue that what made it possible was that a large amount of public money was being invested and the ethos of openness, framed as a moral obligation to open access to what is a shared human heritage, was agreed upon by representatives of sequencing centres from around the world at a key meeting in Bermuda in 1996. Also, a fairly powerful dual incentive was used to increase the willingness to share data – peer pressure and concern about funding. Everyone involved in the project was to share, and it was made clear that scientists who did not share would not receive funding again.

Agreement and compliance were further facilitated by the threat of a competing private enterprise effort planning to patent 300 genes of clinical importance and charge a subscription for access to its data. There was a shared determination in the consortium to ensure that access to this science be freely available. The Human Genome Project has been credited with changing the culture in the field of genomics and the sharing of data is now the accepted norm of what is expected in genomic research.

The Human Genome Project is credited with having had a precedent-setting effect of breaking the cycle of the data generator being the only one who analyses the data. The success of data sharing in genomics has spurred funders to consider to what extent other health data can be opened up, particularly publicly funded large community resources, giving hope that the same can be achieved in disease surveillance.

Challenges in disease surveillance data sharing

The sharing of public health data is considered to be more complicated and sensitive than the sharing of genomics data, however, and many experts agree the path of data sharing in disease surveillance must be trodden carefully if the effort is to succeed. There are several relevant sensitivities around this practice and obstacles need to be addressed and incentives devised before it becomes widespread.

Barriers include the following.

Data quality

The quality of disease surveillance data is variable and data quality is a key challenge for data sharing. The public health information infrastructure is weak in many countries and the quality of routinely collected data, especially in low-income countries, is often poor, so there are questions over how shareable and useable they are.

Data producers are often reluctant to share poor-quality data because they are embarrassed by it and fear criticism of the data. The limited value of poor-quality data is equally a barrier to data sharing for secondary users.

For instance, vital registration data systems are weak in many countries and there are sometimes no data on total births and deaths, let alone accurate reporting of cause of death. Some experts argue that discussions of aspirations for data sharing need to keep this in perspective and subject to a reality check. However, there is a divergence of opinion on how much money should be spent to get high-quality data. Some argue that it depends on the question being investigated. For instance, if it is a broad-brush question on what kind of diseases are present, there is not much need to drill down too far. However, if the aim is to determine whether a vaccine that has been introduced is having an impact, then data would need to be of high quality. The trouble arises when data are collected for purpose A but a secondary user is interested in question B and uses the data without regard for their provenance or quality, which may not be good enough to answer question B and may even lead to the user's analysis yielding a wrong answer.

Data preparation burden and capacity

Documentation and management of databases are often missing steps in disease surveillance. Researchers often realize when they have collected the data how much work is entailed, they lack the expertise and it takes months or years for them to get around to cleaning the data. And they usually only clean the most important variables that they want to look at and leave the rest because they are running out of time and money. Duplicates, inconsistencies and other data management problems are commonly detected when case-based data are shared.

This is mainly a capacity issue and several experts argue that funding is needed to increase data curation capacity where it is weak. It often takes a considerable amount of work to make data, especially older data, shareable. Many institutions, including ministries of health, do not have the technology or capacity to handle the volume of health information that should be shared. Some data

generators report that the biggest hurdle they face in sharing data is editing their data for international use. Since they originally envision using the data themselves, they do not document as thoroughly as they would for secondary users, and it is not a small job to document 20–30 years' worth of legacy data to modern standards. Some producers argue that if they handed over the data in the raw state in which it is kept at some surveillance sites, they would need to spend too much time on the phone explaining what they meant, eccentricities in coding, and other anomalies. And data in that state would, rightly, not be accepted into a global data repository anyway. It is argued that decisions need to be made on what data are worth the work and who should pay for that work. Some argue that it should become standard for researchers to thoroughly document their data as it is collected, even just so that new staff in their own teams can understand it.

Concerns over how the data would be used

Some of the reluctance to share data stems from concerns among data producers that secondary analysts might misuse their data and yield results that are inconsistent with the way the data were collected. This often means that secondary users need to interact with the data producer, rather than pulling the data from the producer and interpreting it alone, even if metadata are provided, as advocates of data sharing urge. Many argue that it is not just a matter of courtesy but also a matter of understanding the context in which the data were collected, and some argue that this is one reason why open access is not appropriate for all surveillance data.

The problem is not as simple as gathering all the data available and putting it to use. It sounds tempting, but some experts argue data sharing can go into overdrive and produce dangerous conclusions. This can happen when, for instance, routinely collected data are used for time trend analysis by taking a bit of data from one country collected in a certain year and combining it with data collected a few years later from a neighbouring country, and from a third country collected five years later, to come up with a picture of what is happening over time in a region. This conflates time and place and doesn't control for what else might be going on in those countries, such as changes in the prevalence of other diseases or nutrition status. It is not possible to control for those variables because the data are not there for that purpose, but because there is such a dearth of data, it is a common temptation for secondary users of data to take data that compares over place and use it to compare over time, which often produces erroneous answers. To address this, experts urge, data sharing should proceed with care and include the person who generated the data to ensure they are used properly. Some researchers charge that when poor-quality data being used for a purpose that was never intended are incorporated into global studies, an unwarranted aura of respectability may attach to them.

Some noted that controversy over the 2010 Global Burden of Disease Report (GBD) is relevant to the discussion of disease surveillance data sharing, as some researchers who contributed data say they were baffled by the findings and have been unable to find an explanation as to why the results were unrecognizable to them from the data they provided. However, it is understood that data sharing was problematic for the GBD 2010 because agreements between the analysts and some data providers prohibited them from sharing that data with third parties. For the next GBD report, provisions have been made to expand the ranks of expert collaborators, and this is expected to widen access to the data.

Concerns over data ownership, professional recognition and participation

Public health researchers the world over make their careers out of the data they collect and are often reluctant to give that away because that is where their power lies. They are usually only willing to consider sharing data after they have published their analysis. Concerns over data ownership and professional recognition explain some of the resistance to data sharing on the part of data producers in both high-income and low- and middle-income countries. In fact, one major funder reports that when there are funder meetings with population researchers and funders say they expect researchers to clean their data, publish quickly and make the data available in a reasonable timeframe, the biggest block comes not from African or Asian researchers, but from their Northern collaborators, usually from Europe or the United States, who see their comparative advantage diminished because the data would be made available to others who are not collaborating. They attempt to slow down data sharing so that they can have exclusivity over the data for longer.

However, there is wide agreement that concerns over data ownership, professional recognition and participation are a fundamental issue hampering the willingness of data gatherers in low- and middle-income countries to share data. Much of the time, data in developing countries, particularly Africa, are collected in a research context, externally funded, and data gatherers are being asked to share the data they collect. Data producers there say they are not entirely comfortable with donors insisting on data sharing without funding that would allow them to analyse and publish themselves, and that there is a major concern among researchers in developing countries that they are expected to be a data-gathering factory for developed-country analysts. They would rather have technical and financial support to build their own analytical and publishing capacity, and get due professional recognition and credit for their contributions, through co-authorship, than be paid to simply hand over their data.

Some data producers say that efforts to advance data sharing will not succeed unless this issue is confronted head-on. That is because it appears the push for greater data sharing results partly from the mismatch between where most of the data are and where most of the analytical talent is. Low- and middle-income countries, particularly in Africa, have lots of data that researchers in developed countries want to tap, but they tend not to have much power or great research machines behind them. Producers say that successful large-scale data sharing requires African agreement and that data producers there will not engage enthusiastically in the practice until they can see that it benefits everyone, and that can share in the process as active participants.

This exposes the long-term capacity issue of strengthening the ability of epidemiologists and researchers in developing countries to analyse their own data. Some funders say their grant application system accommodates data management and analysis, but that grant applicants are not good at asking for it. Some apparently fear that if they add costs their applications might not get approval. The funders report that in general, grant applicants are better at conceiving what information they would like to collect, but then do not pay much attention to its subsequent treatment, such as who is going to enter it, manage it, annotate it and analyse it, even though funding is available to them if they include it in their budget.

Some publicly funded research projects are building data analysis capacity. One example is the USAID-funded Demographic and Health Surveys, which collaborate with countries in the

publication of the data they collect. Some countries have become self-sufficient in the publication of their survey results (see Box 2).

Box 2: The DHS project and data sharing

The Demographic and Health Surveys (DHS) project, set up in 1984 by USAID, is held in high regard as an example of good practice in public health surveillance data sharing. The project has conducted more than 300 nationally representative and internationally comparable household surveys in more than 90 low- and middle-income countries. Data are usually collected every 4–5 years, although some countries survey more frequently. DHS data are in high demand and more than 1,000 secondary analyses have been published.

Capacity-building within the host country is a key priority of the programme, which was designed to be a capacity-building exercise for data collection. Responsibility for executing a DHS resides with a single implementing agency, which can be any governmental, non-governmental or private-sector organization such as a national statistical office, family planning organization, ministry of health, university, government research group or private research group. A DHS contractor provides technical assistance at every level, from survey planning and design, and recruitment and training of surveyors, to data analysis and publication. This person also acts as a quality control officer to ensure survey procedures are consistent with the technical standards set by the DHS. Data belong to the country and can be used for a multitude of purposes, such as time-trend analysis, country comparison, informing policy decisions, resource allocation and identification of intervention strategies. Some countries have graduated from the DHS and are now self-sufficient in the conduct of their surveys.

In order to participate in the survey, countries must agree to produce a report. Although data sharing is not mandatory, it is strongly encouraged in order to make the data available to researchers across the world. Currently, all countries share their aggregate data and 98–99 per cent share their individual-level data, which is standardized by the DHS and stored in its archive. There is open access to data tables and survey methodology on the internet after publication of results and anonymized individual data are made freely available to legitimate users upon application. Moreover, an independent contractor manages the data archive.

The data-sharing success of DHS, the longest enduring internationally comparable survey of its type, is attributed to the fact that the project respects country ownership of data and allows governments to own the process, including the timing of release of data and the publication of the reports. Other hallmarks cited include trust and confidence in the way the archive is managed; the innovative user-friendliness of the database and the web and smartphone platforms; and quality, consistency, objectivity and predictability.

The biggest challenges are reported to be navigation of the political processes of government, and management of in-country and external interest groups that want to be included in the survey, both of which can threaten to depress data quality, present harmonization issues and delay implementation.

Another example of increasing data sharing through capacity-building can be seen in INDEPTH, which coordinates an international network of health and demographic surveillance systems sites operating across 49 sites in 20 low- and middle-income countries in sub-Saharan Africa, Asia and Oceania. The network was for years criticized for not making its data available, but after securing donor funding to increase its analytical, publishing and sharing capacity, it recently launched an open-access online data repository.

Capacity-building is also taking place within other research networks. For instance, in some cases, the network secretariat works out how to analyse the data submitted from its members, and then holds a workshop to teach analysis methods to the statisticians and analysts from the surveillance sites and plans publications with co-authorship.

But experts say that while data analysis capacity-building is a challenge that has to be addressed more widely, data-sharing efforts must in the meantime ensure that data gatherers are rewarded appropriately for their contribution. Their argument is that groups outside a country need a more engaged role in data analysis, one of true partnership, and that telling developing-country researchers that they are ‘helping humanity’ is not enough.

Professional recognition, participation and capacity strengthening are seen as elements of benefit-sharing when it comes to data sharing, and the framing of this as a fairness issue echoes the concerns Indonesia expressed when it stopped sharing influenza virus samples with the World Health Organization in 2007. The Jakarta Declaration on Best Practices for Sharing Avian Influenza Viruses and Resulting Benefits addressed the need for developing countries to share in the benefits, particularly strengthening of laboratory and surveillance capacity and affordable access to vaccines resulting from the sharing of influenza information, data and biological specimens.

Resentment over perceived inadequate exchange

There is a sense among some that the spirit of data sharing – a fair exchange – is not always respected and some cite this as an element of benefits sharing and a significant barrier to the willingness to share. It is widely agreed that for data sharing to work, secondary users must also share their data sets, metadata and methodology, and other relevant information. The sharing of analytical software is seen as equally important in discussions.

Some complain that too often secondary users of their data, whether within their network or external, do not fully reciprocate. In exchange for sharing their data, data producers say they want access not only to all the data that the secondary user generates on the back of their primary data, but also to the methodology being used, so that they can replicate the study and adapt the methodology to expand their own research and ask locally relevant questions that may not be of interest to an international institute but that will enable them to address local problems in-country. And they want the swap to occur when they hand over their data.

Some see this as an ethical or equity issue – who contributes the most and who gets the most out of it? Is it really a global good? Others contend that the effort to increase data sharing has a political tinge to it; part of the push for data sharing comes from a desire to detect potentially dangerous outbreaks earlier, and to some extent the demand for low-income countries to step up surveillance and share data is the rich world keeping an eye on the poor world.

Standardization and infrastructure

One significant practical challenge to widespread data sharing in disease surveillance is the lack of standardization and infrastructure, or sustainable repositories for holding the data. Standardization is considered a key to success and this is thought to be difficult in disease surveillance because the data come in all shapes and sizes. However, there are standardized databases in disease

surveillance. There are also standards for certain data, such as using the International Classification of Diseases to record cause of death, or internationally accepted case definitions, but adherence is a problem. Some experts believe that the standards will need to be flexible to fit various objectives and that standardization bodies may need to be involved in taking the data-sharing agenda forward.

Political sensitivities around government data

The sensitivities around sharing government data are slightly different from those related to data collected by research groups. They include political and economic concerns that affect tourism, which is usually the first to suffer, and trade, especially when it comes to sharing outbreak data. Government sensitivities tend to be fairly disease-specific, but some data analysts say there can sometimes be general concern on the part of government data providers that analysis might expose disparities within the country, making it look bad. There are also often political sensitivities around a country sharing its immediate surveillance data with another country. Governments can also be quite protective of GPS data that show where national boundaries are, which is relevant for surveillance activities that involve mapping the geographical distribution of diseases. In some cases when analysts at WHO headquarters have difficulty gaining access to government data it is the WHO regional office that is holding data back from WHO headquarters, often out of protectiveness towards the country supplying it with the data.

What works?

Despite the barriers, data gatherers and users report that some characteristics and practices do work, to varying degrees and in various contexts. Incentives need to be tailored to the specific concerns of the data producer being asked to share.

Characteristics and practices that make data sharing work include:

Trust – Dialogue, trust and confidence between data producer and secondary user;

Value – Secondary users showing they will add value and be responsible users;

Collaboration – Collaboration with and inclusion of the primary data producer;

Professional reward – Data producer recognition, perhaps through co-authorship or other professional credit;

Exchange – Either knowledge-sharing or some other benefit, such as enhanced access to products resulting from research;

Money – Funding and capacity-building for data management to enable data producers to make their data shareable;

Alarm – A collective external alarm, such as an outbreak or epidemic;

Legal force – Legal requirements to share data, such as under the International Health Regulations;

Global goals – A global eradication effort;

Donor demands – Requiring data sharing as a condition of grants; and

Public shaming – As in the case of China, which began sharing its data on SARS after being publicly shamed for its lack of transparency.

Conclusion

Many data sharers agree that the quest to normalize data sharing in disease surveillance needs to proceed carefully and relatively slowly in order to successfully navigate the sensitivities and resistance currently being shown, and that the first thing is to recognize people's fundamental interests and address them.

Many recommend proceeding with a stepwise and mixed approach; several data sharers say that pressing immediately for open access to all data in disease surveillance is likely to result in failure.

Some say the first step should be to identify the impediments to widespread data sharing. One suggestion is to convene an international group – with representation of the different data fields and barrier groups in both developing and industrialized countries, to engage in a process to articulate the case for data sharing and produce guidelines or a framework for doing so in disease surveillance.

Other suggestions include:

- Developing methods for evaluating data-sharing initiatives, to leverage the aspects that are characteristic of success.
- Articulating a fairness statement on data sharing between countries and between researchers.
- Drafting by funders of a model agreement, or set of agreements, that primary data producers and secondary users can use to spell out the roles and responsibilities on both sides. This could allay some of the fears among data producers that funders want to wrest the data out of their hands to give to others to work on. It would enable funders to show they recognize the larger role that producers want to play and could help producers determine what would be legitimate to ask for in their agreements with secondary users.
- Examining successful initiatives to illustrate the benefits and impact of data sharing, derive best practices and determine the value of conducting more aggressive pilot studies of data-sharing initiatives.
- To make data more shareable, not embarking on, or funding, data collection that is not going to result in analysis, interpretation and reporting.

As for what data sharing in disease surveillance might look like in the future, many believe it is likely to remain fragmented, owing to the diversity of the data types involved and varying skills of

those who collect it. Some natural groupings might emerge to simplify the picture, but it is considered that those are likely to be disease-specific.

Another factor that could have an impact on the current model of data sharing is the advent of mobile phone health apps that allow citizens to send data to researchers and participate more actively in surveillance. It was argued that app users have shown extraordinary willingness to share their health data, raising the question of whether such interactive surveillance could become an increasingly important source of health surveillance data in the future.

Whatever the future model of data sharing, some experts say that secondary users may always feel that they are not getting enough data, quickly enough, in good enough shape. However, it was argued that data-sharing practice is moving in the right direction. It may not be progressing as fast as desired but there is greater openness and insight into the need for sharing than there was 10 years ago and it may be reaching a tipping point.

But for many, the ultimate vision for data sharing, and the goal it is trying to address – improved public health – is a world where epidemiologists, public health workers and investigators can ask and answer questions about health in their country without the work having to be carried out in external organizations. The best policy, some say, is to strengthen decision-making systems and public health career structures in developing countries so that talented people in low-income countries want to make their careers in analysis of health information and countries have the capacity to plan for themselves, decide their own priorities and address their own problems.

About the author

Emma Ross is a consultant at the Chatham House Centre on Global Health Security, advising the Centre on strategy, project development and communications. She is also associate editor of the 20th revision of the Control of Communicable Diseases Manual, to be published later in 2014 by the American Public Health Association. In addition, she helps non-profit organizations involved in medical research, public health or global health develop and articulate their messages and work. She was previously a long-time medical correspondent at the Associated Press, then went to the World Health Organization headquarters in Geneva as news team leader to advise the director-general and head news operations there. She has a Master's degree in journalism from Northwestern University in the United States.

Independent thinking since 1920

© The Royal Institute of International Affairs, 2014

Chatham House (The Royal Institute of International Affairs) is an independent body which promotes the rigorous study of international questions and does not express opinions of its own. The opinions expressed in this publication are the responsibility of the author.

Chatham House, the Royal Institute of International Affairs, does not express opinions of its own. The opinions expressed in this publication are the responsibility of the authors.