

Centre on Global Health Security | May 2017

A GUIDE TO **SHARING THE DATA AND BENEFITS OF PUBLIC HEALTH SURVEILLANCE**

**CHATHAM
HOUSE**
The Royal Institute of
International Affairs

ACKNOWLEDGMENTS

The Chatham House project team is grateful to all those who contributed their time and expertise to the *Strengthening Data Sharing for Public Health project* (listed here in alphabetical order by last name):

Clement Adebamowo, Obijiofor Aginam, Sibounhom Archkawongs, Ray Arthur, Kareem Ayoub, François-Xavier Babin, Angela Ballantyne, Daniel Barth-Jones, Claire Bayntun, Eric Beda, Jesse Berns, Michael von Bertele, Louise Bezuidenhout, Martin Bobrow, Arnold Bosman, Peter Bradley, Richard Brown, John Brownstein, Sergio Brusin, Susan Bull, Gian Luca Burci, Plácido Cardoso, Lisa Carlson, David Carr, Samba Ceesay, Denise Chrysler, Richard Coker, Jennifer Cole, Yali Cong, Mike Conway, Carlos Correa, Steve Crossan, Foday Dafaë, Angus Dawson, Sara Davies, Aboubacar Sidiki Diakité, Zacharie Tsala Dimbuene, Scott Dowell, Kevin Dunbar, Claudia Emerson, Mosoka Fallah, David P Fidler, Jane Greig, Philippe Guerin, Debra Hanna, Habib Hasan Farooqui, Dionisio Jose Herrera Guibert, Anne Huvos, Chikwe Ihekweazu, Kashef Ijaz, Ilesh Jani, Christa June, Unni Karunakara, Stephen Keenan, Marie-Paule Kieny, Daniel Kim, Woo-Joo Kim, Ann Marie Kimball, Gareth Knight, Francis Kombe, Jeffrey Koplan, Kwadwo Koram, Naomi Korn, Kendall Krause, Shengjie Lai, Fessou Emessan Lawson, Lisa M Lee, Marlo Libel, Nigel Lightfoot, Louis Lillywhite, Katherine Littler, Sarah Macfarlane, Larry Madoff, Maimuna Majumder, Mwele Malecela, Madhav Marathe, Vicki Marsh, Sophie Mathewson, Leonard Mboera, David Mbulumi, Sheila Mensah, Laura Merson, Etienne Magloire Minkoulou, Ali Mokdad, Yves-Alexandre de Montjoye, Abdulsalami Nasidi, Hilary Njenge, Charles Okot, Jennifer Olsen, Chi Onwurah, Moe Ko Oo, Charles Oppenheim, Emma Orefuwa, Marisella Ouma, Paba Palihawadana, Elizabeth Pisani, Veerasak Punyapornwithaya, Andrew Roddam, Guénaél Rodier, Wendy Rogers, Emma Ross, Cathy Roth, Charles Rotimi, Kim Sampson, Osman Sankoh, Sam Scarpino, Fabrizio Scrollini, Abul Kair Mohammad Shamsuzzaman, Meghana Sharafudeen, Dan Shoukimas, Richard Smith, Peter Kim Streatfield, Laura Streichert, Bakary Sylla, Antony Taubman, Jemee Tegli, Titus Tei, Mike Thorn, Paulina Tindana, Sok Touch, Kumnuan Ungchusak, Effy Vayena, Nguyen Dang Vung, Sean Ward, Tyson Welzel, Richard Wilder, Julia Wilson, Lynn Woolfrey, Khadija Yahya-Malima, Terdsak Yano, James Zeere, the West African Network for Infectious Disease Surveillance (WANIDS), the Mekong Basin Disease Surveillance (MBDS) network, and staff at Savannakhet and Mukdahan provincial and border health departments.

Thanks are due to the Bill & Melinda Gates Foundation for providing the financial support to pursue this project.

The Chatham House *Strengthening Data Sharing for Public Health* project team:

David R Harper, Michael Edelstein, Asha Herten-Crabb, Matthew Brack and David L Heymann.



Introduction	3
Explaining the guide	4
Principles for sharing the data and benefits of public health surveillance	7
Building trust	8
Articulating the value	12
Planning for data sharing	16
Achieving quality data	20
Understanding the legal context	24
Creating data sharing agreements	28
Monitoring and evaluation	32
Additional resources	36



INTRODUCTION

Sharing public health surveillance data improves and protects public health. It will be an essential element in achieving the UN Sustainable Development Goals (SDGs), particularly SDG 3 – to ensure healthy lives and promote well-being for all at all ages.

This guide has been developed as part of the Chatham House Centre on Global Health Security project *Strengthening Data Sharing for Public Health*. Sharing public health surveillance data enables regional collaborations, capacity strengthening and insight into public health system performance, leads to overall improvements in risk management, and enhances public health responsiveness. Informal data sharing arrangements can often produce the desired public health outcome, and should be encouraged where appropriate. Sometimes, however, more formal agreements are required. This guide is intended to facilitate both informal and formal data sharing, and to be used when a need to share public health surveillance data has been identified or when obstacles to sharing have been encountered. It aims to help create the right environment for data sharing, to facilitate good practice in addressing technical, political, ethical, economic and legal concerns that may arise, and to ensure to the greatest extent possible that any benefits arising from the use of the data are shared equitably.

The guide accompanies the project website – datasharing.chathamhouse.org – and presents summary information and case studies on the seven data sharing principles that all stakeholders should consider when data are to be shared. Further information is available via the website, which also allows users to customize content according to their situation.

The principles have been developed through a series of expert roundtables and consultations, and their release supports the May 2016 call to share public health surveillance data, issued by the International Association of National Public Health Institutes (IANPHI) on behalf of experts convened by Chatham House. The call is published online at <http://www.ianphi.org/news/2016/datasharing1.html>.

EXPLAINING THE GUIDE



WHAT

This guide provides information to facilitate data sharing aimed at improving and protecting public health. The result of a series of expert roundtables convened by Chatham House, it sets out seven principles that address key issues related to data sharing. The principles aim to help create the right environment for data sharing, and to facilitate good practice and encourage ethical sharing to the highest achievable standards; and can also help to identify opportunities for capacity-building. Underpinning all seven data sharing principles are four ethical principles: social value, respect, justice and transparency.



WHO

The guide is intended primarily for those involved in the process of sharing public health surveillance data for the purpose of improving and protecting public health. This includes public health agencies, ministries of health, NGOs, the private sector, academic institutions, multilateral organizations, publishers, funding bodies and others. For convenience, the key stakeholders involved in data sharing can be grouped as follows:

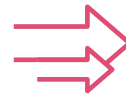
- **Data providers** – those who generate public health surveillance data, either from the community, the healthcare system, or from non-health sources.
- **Data recipients** – those who interpret and use data generated by others.
- **Data sharing facilitators** – those who facilitate sharing between data providers and recipients.

It is common for organizations to both provide and receive public health surveillance data. In addition, they can play a role in facilitating data sharing. It is therefore possible for an organization to belong to more than one group at the same time.



WHEN

In public health, data sharing is most successful when a clear need is identified. This need may arise in situations that range from planned sharing of routine surveillance data to sharing in response to public health emergencies. Forward planning, where possible, is most efficient, so that data sharing systems are already in place if an emergency arises. This guide is intended primarily for those who identify a need to share public health surveillance data, or for those who encounter obstacles when engaged in data sharing arrangements, or when there is a need for a more sustainable, transparent or equitable sharing arrangement. It should be noted that under the International Health Regulations (2005), there is a legal requirement to report cases of certain diseases and certain public health events to the World Health Organization (WHO). Routine disease surveillance serves as the foundation for effective emergency response. However, there are additional considerations when sharing data during public health emergencies. For example, although the principles set out in this guide still apply, there may need to be a greater emphasis on timeliness. WHO has published guidance on sharing data during public health emergencies that is complementary to this guide.



HOW

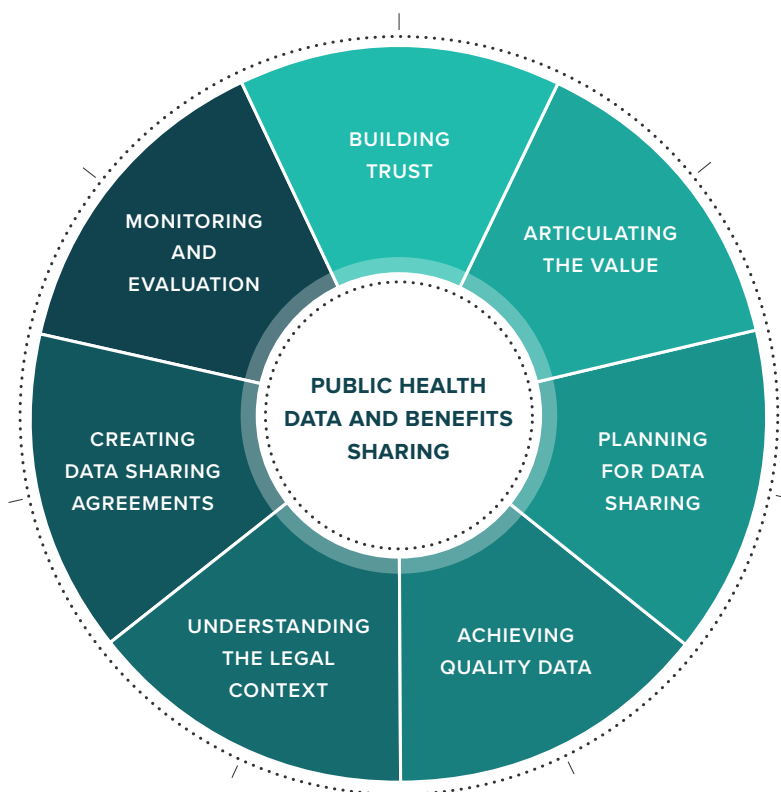
This guide sets out seven key principles that should be taken into account when a need to share public health surveillance data has been identified. Each principle is accompanied by a case study, together with key points to consider when applying the principle. The principles are not hierarchical, nor do they follow a specific order. Not all parts of this guide are applicable in all circumstances, and stakeholders will need to consider what is appropriate for their particular data sharing activity.

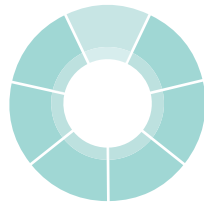
The project website that this guide accompanies – datasharing.chathamhouse.org – provides additional resources, including a model agreement, and allows users to customize the key points to consider according to their specific context.



PRINCIPLES FOR SHARING THE DATA AND BENEFITS OF PUBLIC HEALTH SURVEILLANCE

The seven principles illustrated below are intended to help create the right environment for data sharing and achieve good practice in data and benefits sharing. The principles are not hierarchical, nor do they follow a specific order. The principles are also intended to ensure that data sharing is conducted in a fair and ethical manner for all those involved, and in a way that contributes to the building of surveillance capacity.





BUILDING TRUST

Trust facilitates successful data sharing, which in turn reinforces trust. There are two aspects of trust in public health surveillance data sharing. First, sharing should be done in a transparent manner, and the communities from which the data originate should know how the data are collected, analysed, used and protected. Second, trust-building measures between stakeholders, at the personal or organizational level, help create an environment conducive to data sharing. Trust is built when the purpose of data sharing is made clear, and when those involved in the process know each other, understand each other's expectations, and carry out their commitments as agreed. Trust increases the likelihood of equitable benefit sharing and further collaboration, and improves core surveillance capacity through the creation of surveillance networks. It can be very hard to build trust, but very easy to lose it.



CASE STUDY

BUILDING AND MAINTAINING RELATIONSHIPS

The Mekong Basin Disease Surveillance (MBDS) network was established in 2001 in Southeast Asia to enhance cross-border outbreak investigations, public health response and communication using surveillance data sharing protocols. The MBDS, which originated as a response to a serious cross-border cholera outbreak, is now built on multilateral agreements among six regional governments: Cambodia, China (Yunnan and Guangxi provinces), Laos, Myanmar, Thailand and Vietnam. The initiative began locally, with support from the Rockefeller Foundation as well as several development and UN agencies.

Following on from the local initiative, the health ministers of each MBDS member country signed a memorandum of understanding to provide an agreed framework for the governing structure and processes of the network. Since its launch, teams from member countries have worked together on several outbreak investigations, and have improved cross-border disease surveillance. By means of knowledge and technology transfer and regularly planned joint activities such as workshops and outbreak simulations, MBDS has also benefited overall surveillance capacity in the region.

Through MBDS, outbreak investigations in the Mekong Basin are performed with standardized tools, and surveillance and outbreak investigation data are routinely shared. MBDS members meet annually, together with public health professionals, network board members and policymakers, to discuss challenges, achievements and future plans, and to identify and address possible problems. Experiences arising from the network are also routinely shared with similar networks globally.

MBDS coordinators attribute the success of the network to the excellent relationships of those involved in this cross-border surveillance network. Relationships have been established over time, and are consolidated through regular meetings. However, the emphasis on interpersonal relationships means the network is vulnerable to staff turnover. To mitigate this risk, it is essential to maintain trust and communication between participant countries at the local, provincial and national levels.



BUILDING TRUST



KEY POINTS TO CONSIDER

Have the expectations of all stakeholders affected by the data sharing been taken into account?

The purpose of sharing data should be explicit and transparent. If data providers are coerced into sharing data (for example by another organization or by a higher level within their own organization), trust can be more difficult to establish and maintain. Data sharing works well when those directly involved in the process know each other; personal connections and joint activities such as workshops and joint publications help create an environment that facilitates sharing. Other stakeholders' perceptions, including public perceptions, can also play a part in the success of data sharing. It is therefore important to consider public engagement and the protection of the population from which the data originate, in particular when public health surveillance data are crowdsourced and in cases where anonymization is difficult.

How has the data sharing agreement been developed?

A combination of personal interaction and supporting documentation can help to create the right environment for data sharing. The level of trust between parties will affect the type of agreement they develop, and its terms. Agreements should be developed and implemented collaboratively. The terms of the agreement should meet the expectations of all parties, be realistic and be achievable. The process should be transparent, and should give each party a clear understanding of what is expected of them, and what they can expect of others. An agreed scope of accountability, and respecting and delivering on the terms of the agreement, are essential components of trust.

Will there be equitable benefit sharing?

In addition to direct public health benefits, data sharing can also help create a trusting environment and improved capacity at the local,

national or regional level. In addition, it can create opportunities, for example for scientific publications and the ability to analyse pooled surveillance data to inform public health decisions. Such positive externalities require careful consideration and collaboration between the parties, and should be embedded in the agreement. Consideration should be given to the fact that the use of certain software packages for analysis or for making the data ‘open’ can lead to unfair access to data, since both rely on information technology expertise and facilities that may not be available to all parties.

Does the agreement reflect the rights and responsibilities of each party equitably?

There should be transparency in the use of data, and accountability that allows redress if data misuse has occurred. Any alternative data uses that have not been considered or that do not form part of the initial sharing agreement should be discussed and agreed between the parties prior to any such use. The data and

resulting information should be accessible to all parties equitably. Restricting providers’ access to their own data – such as by failing to share analytical results or to involve providers in a meaningful way – harms trust, reduces the value proposition of sharing, and compromises long-term access to data. Where one party does not adhere to the agreed terms, other parties should be empowered to terminate the agreement.

Have parties agreed on frequency and means of communication?

Planning appropriate communication by phone, email and, in particular, face-to-face meetings can help build a lasting and trusting relationship, as can joint approaches to evaluating the sharing process and the resulting data.



ARTICULATING THE VALUE

Data sharing should inform and improve public health action. It enables regional collaboration, capacity strengthening, insight into public health system performance, and better risk management and control of infectious diseases. However, concerns over, for example, loss of rights and/or the potential for misuse of data can increase the risk of data providers being reluctant to share, or of stakeholder backlash against any data sharing agreement made. When initiating data sharing, it is important to make the purpose of doing so clear, and to articulate the benefits and risks, so that all stakeholders are able to understand the value of sharing and how the data will be used. Equitable sharing of the benefits is also important. Making explicit potential benefits, such as improvements in public health and collaboration opportunities, can encourage data sharing.





CASE STUDY

FAILING TO UNDERSTAND REAL OR PERCEIVED RISKS

The care.data programme was an initiative from the publicly funded healthcare system in England that aimed to collate health and social care information securely from different healthcare settings – such as GP practices, hospitals and care homes – into a central database. The objectives were both to improve patient care and efficiency and to further medical research. The programme was controversial from the time of its announcement, with concerns about a lack of transparency in how data were to be used, as well as the potential risks to the confidentiality of patients' personal information. It was decided to close the programme down after reviews identified inadequate public consultation early in the process, as well as the need for more effective dialogue with the public on the use of personal data.

The failure of care.data was at least in part a result of an overriding focus on the potential benefits of the programme without due consideration to public perception and likely concerns. There was also a failure to recognize the significance of the relatively weak safeguards for personal data that existed in public

healthcare systems in England. The public are concerned about the use of personal data without their explicit consent. For example, studies have shown that there is a clear 'red line' for many people regarding the sharing of personal data with commercial organizations seeking to profit from the data, or for reasons that do not contribute directly towards improved public health. They are also concerned about uncertainty associated with the potential future use of their data.

The value of data sharing has to be weighed against the real or perceived risks, and communicated properly to create the right environment for sharing to take place. There is widespread agreement on the benefits of a system for sharing healthcare data, but care.data did not adequately address the real or perceived risks to those whose data were to be shared.



ARTICULATING THE VALUE



KEY POINTS TO CONSIDER

Have the public health benefits that will result from the data sharing been made explicit?

When initiating any data sharing activity, it is important that the public health purpose for doing so is made clear, along with the need it aims to address. This reduces the risk that the sharing may be perceived as having little benefit to those from whom the data originate.

Have any public health actions intended to be taken as a result of the data sharing been made explicit?

All stakeholders should be aware of what the public health benefits are, and how they can be achieved. Options for actions resulting from the data sharing should be discussed with stakeholders when the data sharing agreement is created.

All stakeholders should
be aware of what the
public health benefits are,
and how they can
be achieved.



Are potential risks and benefits clear to all stakeholders?

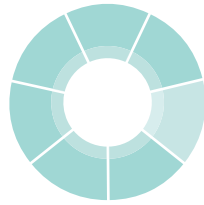
Both potential risks and benefits to participating organizations and to the communities from which the data originate must be taken into account. In addition to improved public health, benefits include academic opportunities, publications and capacity-building. Risks include data misuse, and inadequate safeguards concerning data privacy, consent and security.

How will the costs of data sharing be covered?

Sharing public health surveillance data has cost implications in terms of human, technical and material resources. Ensuring that costs for data sharing can be met in an equitable and sustainable manner should increase the likelihood of effective data sharing, and can help build capacity.

In addition to improved public health, benefits include academic opportunities, publications and capacity-building.





PLANNING FOR DATA SHARING

Public health surveillance data should be collected with potential sharing in mind. Sharing is most successful when it addresses a need, and this should be identified at the outset. Sharing should also meet the expectations of all parties. Planning for data sharing should extend to all steps of the data management cycle: collection, processing, analysis, preservation, access, reuse, and disposal. Management of data to the appropriate technical and ethical standards requires a workforce with the necessary skills and capacity. As such, data sharing can provide an opportunity for capacity-building, in line with the requirements of the International Health Regulations (2005).



CASE STUDY

CREATING A NATIONAL PUBLIC HEALTH INDICATORS DATABASE

The Public Health Outcomes Framework (PHOF) was published by the UK Department of Health in 2012 to help promote transparency and accountability across the public health system. A wide range of public health indicators are shared in the public domain. The aim is twofold: to assist local leaders in planning, developing and implementing strategies to improve health and wellbeing; and to provide the wider public with measures of how well their local services are supporting them.

Public Health England (PHE) has responsibility for analysing and presenting the indicator data for the PHOF through a publicly available web tool. The data come from a wide variety of sources, and include data already published by other organizations (e.g. under-18 conceptions) through to those requiring calculation from individual-level datasets (e.g. mortality from causes considered preventable).

Various data sharing agreements are in place that allow PHE to provide data for indicators where the data are owned by other organizations. The policy nature of the PHOF is a good driver for these agreements, which in general

have been achieved through planning and positive discussions with partner organizations. Indicators are generally updated annually, as part of a quarterly cycle, and are available for users to download and reuse under the terms of the Open Government Licence.

In order to ensure the quality of the data published via the PHOF web tool, the process for each indicator requires an independent quality assurance of the analyses. Standard templates are used for completeness and consistency, and further checks on uploaded data are undertaken on a test version of the web tool before publication.

In addition to quality, the methods used for each indicator are carefully considered. In the most recent review of the indicators, a standard set of criteria were determined against which the indicators were assessed. These included the validity, timeliness, availability and construction of each indicator. Only indicators that met these criteria and had detailed metadata were considered in the review.



PLANNING FOR DATA SHARING



KEY POINTS TO CONSIDER

Has the allocation of responsibilities been considered for each step of the data management cycle?

The data management cycle includes collection, processing, analysis, preservation, access, reuse, and disposal. There are implications for all steps of the cycle, and each should be considered by both those providing and those receiving the data, with attention given to what resources are required and who is responsible for providing them. In some instances, a step-by-step approach to the data management cycle can be an opportunity for skills and knowledge transfer between those involved.

Are the data managed in a way that facilitates sharing?

Data should be collected in a recognized format, using software available and familiar to all parties, providing a description of the data (metadata), signposting other parties to the data, and storing confidential information using

mutually agreed standards (including security standards). The possibility of later data reuse should also be considered, with attention paid to matters such as the technical and legal issues associated with linking datasets, and how data will be discovered and accessed in the future.

Have the necessary steps been taken to protect the population from which the data originate?

The necessary steps include adequate anonymization and secure storage. When considering adequate anonymization procedures, the possibilities of data linkage using other readily available sources should be taken into account. Where consent is required, this should be managed in a way that balances public health benefits with individual privacy concerns. This may require a managed access procedure, with terms and conditions for access agreed between parties. Requirements of national and international law must be observed when sharing across national borders. For example, there may



be specific restrictions for disaggregated data containing confidential or personal information. Aggregated and anonymized data are subject to fewer legal restrictions, providing data have been anonymized to acceptable standards.

Have the relevant technical standards been considered and applied?

While no universal standards have yet been developed for public health surveillance data, there are existing clinical and functional standards that are relevant. There are also existing standards for the steps of the data management cycle. These standards should be identified and used where appropriate to ensure effective data use for public health action.

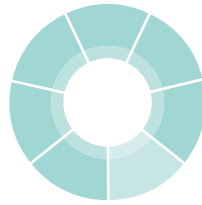
Are the necessary resources available to manage data sharing to acceptable quality standards?

These include financial, technical and human resources. If the planning of data sharing

identifies gaps in resources and capacity, measures to address them should be implemented as part of the collaboration between the parties and with any data sharing facilitators. This has the potential to both improve capacity and establish trust between parties. Mechanisms to promote longer-term, sustainable data sharing should also be considered alongside more immediate goals.

Which existing data management models have been considered when planning for data sharing?

There are existing data management models that can provide useful information on planning for data sharing, data management and how to develop agreements. It is useful to refer to such models when approaching new data sharing arrangements, to reduce duplication of effort and the risk of isolated approaches that inhibit wider collaboration and data reuse.



ACHIEVING QUALITY DATA

Data systems can be evaluated for a number of key characteristics, including relevance, accuracy, timeliness, accessibility, interpretability and coherence. Sharing data enables feedback and should therefore improve quality. Technical and human resource factors influence data quality; standardization and automation should make sharing easier, more efficient and more effective; while high-quality data provision requires a skilled workforce to develop, manage and evaluate the surveillance systems from which the data arise. There are opportunities to improve quality as those involved in data sharing work to achieve higher standards. However, this work should not have a negative impact on sharing: quality needs to be balanced with timeliness.



CASE STUDY

LESSONS FROM GOOGLE FLU TRENDS

Public health agencies conventionally use trends in laboratory-confirmed cases or primary care consultations for influenza surveillance. This can result in incomplete data and in notification delays. In 2008 Google began publishing data to help estimate trends in influenza activity and predict the onset of the yearly influenza season ahead of official influenza surveillance systems. The system, named Google Flu Trends (GFT), was based on the pattern and trend analysis of search terms, combined with a modelling algorithm. GFT aimed to allow real-time analysis at national or regional level, and provided data for up to 29 countries. Early evaluations indicated that GFT could accurately predict the timing and extent of the influenza season.

In the spring of 2009 an unexpected strain of influenza A (H1N1) virus emerged in Mexico and the US, outside the normal winter seasonal pattern. The GFT algorithm underestimated what was in fact the first wave of a pandemic, and was not able to detect it earlier than conventional surveillance. The algorithm was subsequently revised. In 2013, at the height of an influenza-like illness outbreak, GFT overestimated the proportion of doctor visits for influenza-like

illness by a factor of two compared with US Centers for Disease Control and Prevention (CDC) data. The algorithm was again updated. The discrepancies were reported in the scientific press to have eroded the faith the public health community had in GFT, and in digital disease detection more widely. The algorithm was updated again in 2014, but Google ceased publishing GFT data in August 2015.

The GFT experience does not negate the value of digital disease detection, nor the use of big data for public health surveillance, particularly since combining GFT data with other surveillance data has shown great promise. However, it does highlight the need to consider whether the quality of data is adequate for the intended sharing purpose, and whether the most important aspects of the surveillance data (in this instance timeliness and accuracy) have been properly considered before sharing. GFT appears to have suffered as a result of a perception that it could improve – and possibly replace – conventional surveillance, even though its use had been primarily experimental and was intended to be complementary to existing systems.



ACHIEVING QUALITY DATA



KEY POINTS TO CONSIDER

Are the data of adequate quality for the intended purpose?

Data do not have to be ‘perfect’, but should be of good enough quality for the intended purpose. Sharing may improve data quality through a feedback mechanism. However, setting unnecessarily high standards could inhibit data sharing in certain contexts. It is important to consider whether sharing some data could be better than holding back the data entirely.

Have the relevant technical standards been considered and applied?

There are existing standards for the steps of the data management cycle, and these should be identified and used where appropriate. This can reduce duplication of effort, help to ensure that minimum data quality requirements are met, and allow for data comparison across organizations and countries. Where specific standards are applied, they should be explicitly detailed. Organizations specializing in data sharing have

described essential elements of data quality such as relevance, accuracy, timeliness, accessibility, interpretability and coherence.

What key attributes of the data are most relevant to the intended use?

When surveillance data are shared, it is important to identify which characteristics such as relevance, accuracy, timeliness, accessibility, interpretability and coherence are most relevant to the intended public health purpose, and to prioritize improvements in these areas where possible. Third-party organizations specializing in data sharing can help address shortfalls in data quality.



What aspects of the sharing process can be automated?

Automation can reduce the risk of error, improve efficiency and decrease human resource requirements. Stakeholders involved in data sharing should identify relevant parts of the process where automation is appropriate, feasible and cost-effective.

What resources are necessary to achieve the required data quality?

Sharing data of sufficient quality requires a skilled workforce and technical capacity. Stakeholders should ensure that resource requirements are met in order to ensure data quality is sufficient to achieve the intended public health purpose.

Data do not have to be ‘perfect’, but should be of good enough quality for the intended purpose. Sharing may improve data quality through a feedback mechanism.



UNDERSTANDING THE LEGAL CONTEXT

The legal implications of data sharing, and the most suitable type of agreement, depend on issues such as geographical location, type of institution involved, type of data, level of public health threat, and other contextual factors. It is important that parties understand the legal implications and the legal tools available to facilitate the process. Where guidance exists, the balance between making data accessible, safeguarding privacy and protecting intellectual property is not standardized, which can result in protective policies. Data sharing agreements can help resolve differences or ambiguities in law, and are most successful when the context is well defined, and when relevant laws and regulations are taken into account. In some instances, an agreement that is not legally binding may be more suitable than using legal means.



CASE STUDY

LESSONS FROM THE EUROPEAN SURVEILLANCE SYSTEM

The European Centre for Disease Prevention and Control (ECDC) identifies, assesses and communicates risks to human health from infectious disease. It collects, analyses and disseminates surveillance data on 52 diseases from 31 countries, in a database called The European Surveillance System (TESSy). Data access is regulated, and is made available to third parties only on request and approval.

In 1998 the European Commission formalized disease surveillance networks previously funded as pilot projects. As a result, these informal networks grew and were standardized, and specific reporting meta-datasets were adopted. They were eventually incorporated into ECDC. EU legislation ensures that data are shared in a transparent manner, although reservations have been expressed by member states about making data public before national authorities have first had the opportunity to analyse and use the data for scientific publications. This concern has been partially resolved by allowing for delayed reporting of some datasets, and for anonymization of commercially sensitive data (for example, data on healthcare-associated infections). TESSy data are also restricted by

EU data protection laws that inform ECDC's data access policy. However, member states interpret EU legislation on processing personal data in different ways, with countries transferring different types of data to ECDC. This has made the standardization of surveillance data collection difficult. The solution has been to allow 'mandatory' and 'voluntary' variables to be reported, as well the reporting of aggregate data in some instances.

The EU-wide surveillance system was built gradually, on existing informal networks. Flexibility when harmonizing different pieces of national legislation has been essential, even at the expense of the ability to standardize. It was not possible to engage all countries at the outset. However, legal objections from member states were overcome as the benefits became clear. The legal framework is capable of change as new needs and technologies arise, and agreements can be updated to reflect such developments.



UNDERSTANDING THE LEGAL CONTEXT



KEY POINTS TO CONSIDER

Have the relevant legal frameworks been fully utilized to facilitate data sharing?

The legal issues relevant to data sharing are not necessarily barriers to sharing. Barriers to data sharing are more often political or motivational, but those who oppose sharing the data sometimes cite legal barriers when the obstacles are more political or motivational. Knowledge of the relevant legal frameworks, including national and international law, as well as institutional data sharing policies, can help facilitate sharing when other barriers exist.

How relevant is international law for public health data sharing?

As countries have primary legal authority over activities within their borders, they are responsible for implementing the International Health Regulations (2005), the principal international health law. However, there is no enforcement mechanism to ensure compliance. Another relevant area is international human rights law, which contains a well-established approach

for balancing respect for individual rights with other important interests. For example, the International Covenant on Civil and Political Rights requires states to respect the right to privacy, but recognizes that states can waive the right to privacy for public health reasons. Sharing public health surveillance data across borders has legal implications when the type of data shared is protected by national or international law. One example is disaggregated data containing confidential or personal information. Aggregated and anonymized data are subject to fewer legal restrictions, providing data are anonymized to acceptable standards.

Is the agreement compliant with relevant laws and regulations?

Data sharing agreements should comply with institutional data sharing policies, the national laws of all countries involved, regional and global legislation, and any other legally binding agreements. Both data providers and data recipients must comply with the prerequisites, conditions and limitations established by data



sharing law. However, while data recipients may not be directly affected by law that relates to the data provider, any legal prerequisites, conditions and limitations relating to the data provider should be included in the agreement to achieve compliance.

Has the ownership of resulting outputs been agreed?

Shared data may be used for a variety of analyses and purposes. It is important that the intellectual property status of datasets, for example, as well as authorship for any expected academic outputs, is agreed at the outset.

Have local customs and sensibilities been taken into account?

Adherence to and respect for relevant laws and customs, particularly relating to the country from which the data arise, will help to promote trust and confidence among local stakeholders.

Are there pre-existing agreements in place that need to be taken into consideration?

Pre-existing agreements between organizations, whether focused on data sharing or wider agreements with data sharing components, may restrict sharing the same data with a third party. When data are shared between multiple organizations and multiple agreements are in place, these should be compliant with each other.

Have different types of agreements been considered?

The legal implications of data sharing, and therefore what type of agreement is most suitable, can depend on various factors, including geographical location, type of institution involved, type of data, and level of public health threat. This means that legally binding agreements may not always be necessary.



CREATING DATA SHARING AGREEMENTS

Formal data sharing agreements are not necessary if informal arrangements are sufficient to accomplish the goal of sharing. The rights and interests of stakeholders should be properly taken into account whatever arrangements are made. Agreements can take different forms, from short memoranda of understanding to detailed, legally binding agreements. Depending on the context, an agreement can take place at the local, national or international level. To be successful, it should take into account the needs and expectations of all parties. Failure to address such issues before drafting the agreement can lead to inequitable sharing of benefits and missed opportunities for capacity-building. Parties should take steps to ensure that: the terms of reference are acceptable to all; data providers have the opportunity to take part in data analysis; benefits are shared equitably; and potential harms to individuals and communities are minimized.



CASE STUDY

DIFFERENT MODELS FOR SHARING DATA

The WorldWide Antimalarial Resistance Network (WWARN) facilitates the sharing of data from clinical trials to support large-scale pooled analyses of efficacy trials of antimalarial medicines. This approach provides statistical power to answer key scientific questions that cannot be answered by individual trials.

As a data sharing facilitator, WWARN standardizes and amalgamates datasets from a range of data providers, making the pooled data available to the wider research community and policymakers, as well as to the data providers themselves. WWARN also works to identify knowledge gaps in malaria treatment that might be filled by analysis of platform data, and supports collaborations of data providers to analyse and publish results.

WWARN's success is contingent on the willingness of malaria researchers to share data, based on the collaborative governance framework agreed between all parties. WWARN has adopted an approach that encourages data providers to be fully involved in any analysis, and to have their contribution recognized in accordance with internationally recognized guidance.

The data sharing agreement between WWARN and data providers enables WWARN to curate the data and produce an inventory of available data for potential users to browse. Data recipients wishing to access data are encouraged to involve all data providers in analysis and publication. To date, there are more than 20 individual patient meta-analyses published or ongoing via this type of collaboration; some have resulted in changes to national and international malaria treatment guidelines.

WWARN is increasingly offering more data governance options to data providers. When depositing data to the platform, data providers can choose either to be contacted regarding each application to access their data or to delegate these decisions to a Data Access Committee.



CREATING DATA SHARING AGREEMENTS



KEY POINTS TO CONSIDER

Have all parties to the agreement contributed to it?

It is important to consider the needs and expectations of all parties when drafting an agreement. Addressing any issues at the outset will help create a better first draft document and expedite any discussion regarding the terms. Failure to do so may lead to inequitable sharing of burdens and benefits, and to the interests of some parties being over- or under-represented. A collaborative approach to data sharing agreements helps ensure that all interests are considered, that benefits are shared equitably, and that a trusting environment is created.

Is the agreement appropriate to the specific context in which data sharing is taking place?

Data sharing agreements can take different forms depending on the nature of the data being shared, the type of stakeholders involved, any pre-existing relationship between stakeholders, and whether or not the agreement needs

to be legally binding. Parties should discuss and understand what type of agreement is most appropriate.

Have the right people been involved in creating the agreement?

The context and the objective of the data sharing, as well as the nature of the parties involved, may require different levels of authorization. This could range from an informal agreement at the local, technical level to a formal ministerial authorization. It is important to ensure that signatories to the agreement have the necessary authority. In addition, stakeholders should check that all ethics approval(s) required for the intended use of data have been obtained.

Are the interests of all parties represented?

The data sharing agreement is an opportunity for all parties to identify and articulate the value of sharing data, and to agree on how it can be achieved. This creates trust, and sets a positive



precedent to help public health surveillance data sharing become the norm. It also helps to ensure that benefits are shared equitably and according to expectations.

Have all intended uses of the data been specified?

Intended uses include the public health action resulting from the data, as well as related activities and outputs such as the production of reports and academic publications. Any intellectual property implications should be identified, and provision should be made for the possibility of further discussion in the event that any unanticipated intellectual property issues arise from data reuse. Parties to the agreement should decide what degree of departure from agreed uses requires approval from the data provider.

Is the language used in the agreement understandable to all parties?

This includes the language in which the agreement is written, as well as any specific technical or legal terms used. Access to technical and legal resources may vary. Therefore, the language used should reflect the level of technical and legal expertise available to all parties.

Have existing international agreements relevant to data sharing been consulted?

Existing international agreements should be taken into account where relevant. The Pandemic Influenza Preparedness (PIP) Framework and the Nagoya Protocol on Access and Benefit-sharing are examples of agreements that consider a range of data sharing issues, and these may be useful both in articulating the data sharing agreement and more widely in the data management cycle.



MONITORING AND EVALUATION

Sharing public health surveillance data improves and protects public health when a need is addressed and the proper process is followed. It is therefore important to ensure that the data are shared as planned and are used for the intended purposes, and that the desired impact is achieved. As new sources of surveillance data emerge, and as data are successfully shared, case studies that demonstrate the added value of data sharing should be recorded and disseminated. Sharing success stories can help make data sharing the norm rather than the exception. Equally, cases in which sharing does not have the expected impact, or where lack of data sharing has contributed to negative public health outcomes, should be examined and documented to help make improvements in the future.



CASE STUDY

PROMOTING EQUITABLE ACCESS TO DATA – A FUNDER'S EXPERIENCE

In response to the Health 8 (H8) inter-agency commitment to 'improve the quality, availability, and use of health data to accurately track health progress and increase results-based accountability', the Bill & Melinda Gates Foundation (BMGF) developed a set of Global Health Data Access Principles in 2011. Its data policy aims to increase transparency and information sharing for work supported by BMGF, and to support more equitable access to data and contribute to faster progress on global health challenges. BMGF covers the cost of data curation and storage, provided such costs are noted and requested in grant proposals.

BMGF audited the pilot data policy in order to understand how grantees incorporated this into their work. As a result of the audit, BMGF revisited the policy and incorporated early lessons by modifying the policy as necessary (for example, refining questions in the Data Access Plan template, and adding a glossary of terms so that definitions were explicit), developing a more tailored approach to engaging with partners and grantees, and developing supporting

tools and communications (such as a 'Frequently Asked Questions' document to assist programme officers and grantees). An Open Access Policy was also put in place to ensure that published research resulting from BMGF funding is promptly and broadly disseminated.

BMGF sees the Global Health Data Access Principles and its Open Access Policy as part of a longer-term approach to identifying high-impact methods to promote the sharing of high-value global health data. To date, BMGF has seen evidence that open access data facilitate collaboration and analysis, and strengthen capacity. More broadly, BMGF envisions a framework through which the results of data sharing can be gathered, compared and disseminated to provide further insight into how and when data sharing is most useful.



MONITORING AND EVALUATION



KEY POINTS TO CONSIDER

Have instances of data sharing been documented?

Documenting what is shared with whom, and how the data are used, can help in the dissemination of success stories, and in learning lessons to help improve future data sharing. This documenting process applies to all steps of the data management cycle.

Have expected outcomes been defined?

An important early step in monitoring and evaluating data sharing is defining what constitutes success or failure in a specific context. This should be decided collaboratively to ensure buy-in and mutual understanding of the aims of the data sharing relationship.

Has data sharing produced the desired outcome?

Data sharing agreements should describe the intended uses of the data, and any resulting public health actions should be discussed as part of creating the agreement. Once data have been shared, it is important to evaluate whether the uses and outcomes match expectations. When data are openly accessible, recording data reuse and its public health impact can help improve the perceived value of sharing.

How are outcomes recorded?

Monitoring and evaluation frameworks can help all stakeholders to understand and assess data sharing operations. At the beginning of the data sharing relationship, parties should decide collaboratively what will constitute a positive outcome of the data sharing process and its impact. If this outcome is not achieved, root causes should be analysed. Common reasons for failure are the data not addressing a specific



need, not being visible, not being shared in a timely manner, or not being provided in a usable format.

How have monitoring and evaluation been integrated into the data sharing process?

The most successful monitoring and evaluation systems are those that are fully integrated into the programme of work at the outset, and that factor in the human and material cost. Ensuring that monitoring systems are fully integrated into the data sharing process can increase the accuracy of evaluations, and can help identify and correct errors in the system at an early stage. Failure to plan for monitoring and evaluation can leave individual parties to bear the cost, and can compromise the data sharing activities. Therefore, both parties should discuss who will be responsible for the relevant tasks, and how costs should be covered.

How have data and benefit sharing outcomes and lessons learned been communicated?

Capturing the data and benefit sharing process, and defining and documenting the outcomes and lessons learned, may not be sufficient. In some instances, the outcomes and lessons learned need to be communicated actively to stakeholders, including the general public. This should help to increase transparency and maintain trust between stakeholders. Furthermore, it can be particularly important since surveillance systems are often publicly funded. Disseminating successful outcomes and lessons learned should help to increase the evidence base and encourage further sharing.



ADDITIONAL RESOURCES

The full project website is available at datasharing.chathamhouse.org

The full content includes a model agreement that users can amend according to their specific context, further information and case studies, and links to external guidance that can help users develop a fair and equitable data sharing relationship.

Further information on key resources referenced in this guide is available via the following links:

- Bill & Melinda Gates Foundation Global Health Data Access Principles: <https://docs.gatesfoundation.org/Documents/data-access-principles.pdf>
- Bill & Melinda Gates Foundation Open Access Policy: <http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>
- Health 8: http://www.searo.who.int/entity/partnerships/topics/donors_ghp_h8/en/
- International Association of National Public Health Institutes Call to Share Data: <http://www.ianphi.org/news/2016/datasharing1.html>
- International Health Regulations (2005): http://www.who.int/topics/international_health_regulations/en/
- Mekong Basin Disease Surveillance: <http://www.mbdnet.org/>
- Nagoya Protocol on Access and Benefit-sharing: <https://www.cbd.int/abs>
- Pandemic Influenza Preparedness Framework: http://www.who.int/influenza/resources/pip_framework/en
- Public Health Outcomes Framework: <http://www.phoutcomes.info/>
- Sustainable Development Goal 3: <https://sustainabledevelopment.un.org/sdg3>
- The European Surveillance System: <http://ecdc.europa.eu/en/aboutus/what-we-do/surveillance/Pages/data-access.aspx>
- World Health Organization guidelines on data sharing during emergencies: http://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/
- WorldWide Antimalarial Resistance Network: <http://www.wwarn.org/>

IMAGE CREDITS

Cover image and pp. 16–19: Planning for data sharing

Nurse searching through clinical and health records of patients at the Ngoma health centre, Rwanda, on 5 February 2014.

Photo: Thomas Imo/Photothek via Getty Images

p. 1: Contents

World travel and communications recorded on Twitter.

Data from the Twitter streaming API, 17 May–1 September 2011.

Photo: Eric Fischer, under Creative Commons Attribution licence (see <https://creativecommons.org/licenses/by/2.0/>).

pp. 2–3: Introduction; and pp. 12–15: Articulating the value

Children stand outside an MSF field clinic during a vaccination programme against measles for children living near an IDP camp in Bangui, Central African Republic, on 7 January 2014.

Photo: Miguel Medina/AFP/Getty Images

pp. 4–7: Explaining the guide; and Principles for sharing the data and benefits of public health surveillance

USAID Asia: USAID boosts disease surveillance across borders in Southeast Asia, 22 January 2015.

Photo: Richard Nyberg, USAID, under Creative Commons Attribution-NonCommercial licence (see <https://creativecommons.org/licenses/by-nc/2.0/>).

pp. 8–11: Building trust

A mother and her child waiting in the emergency room of a hospital at Petare slum, a highly suspected Zika virus propagation area, east of Caracas, Venezuela on 4 February 2016.

Photo: Carlos Becerra/Anadolu Agency/Getty Images

pp. 20–23: Achieving quality data

A lab worker performs a procedure during mosquito DNA analysis at the Fiocruz Institute on 2 June 2016 in Recife, Brazil, at the time of the Zika virus outbreak.

Photo: Mario Tama/Getty Images

pp. 24–27: Understanding the legal context

Officials wearing masks and protective suits proceed to cull chickens in Hong Kong on 28 January 2014, as Hong Kong began a mass cull of 20,000 chickens after the H7N9 bird flu virus was discovered in poultry imported from mainland China, authorities said.

Photo: Philippe Lopez/AFP/Getty Images

pp. 28–31: Creating data sharing agreements

Women being screened for cervical cancer at the Kwale rural clinic, Kenya, on 12 June 2014.

Photo: Jonathan Torgovnik for The Hewlett Foundation/Reportage by Getty Images

pp. 32–35: Monitoring and evaluation

Swine flu suspect patients waiting outside the swine flu screening ward at the Ram Manohar Lohia hospital on 23 February 2015 in New Delhi, India.

Photo: Sonu Mehta/Hindustan Times via Getty Images

p. 36: Additional resources

A TB patient holds a medical card from a government hospital as she sits at her home in Sonipat, India, on 3 March 2016.

Photo: Money Sharma/AFP/Getty Images

Independent thinking since 1920

Chatham House, the Royal Institute of International Affairs, is an independent policy institute based in London. Our mission is to help build a sustainably secure, prosperous and just world.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical including photocopying, recording or any information storage or retrieval system, without the prior written permission of the copyright holder. Please direct all enquiries to the publishers.

Chatham House does not express opinions of its own. The opinions expressed in this publication are the responsibility of the author(s).

Copyright © The Royal Institute of International Affairs, 2017

ISBN 978 1 78413 176 0

This publication is printed on recycled paper.