

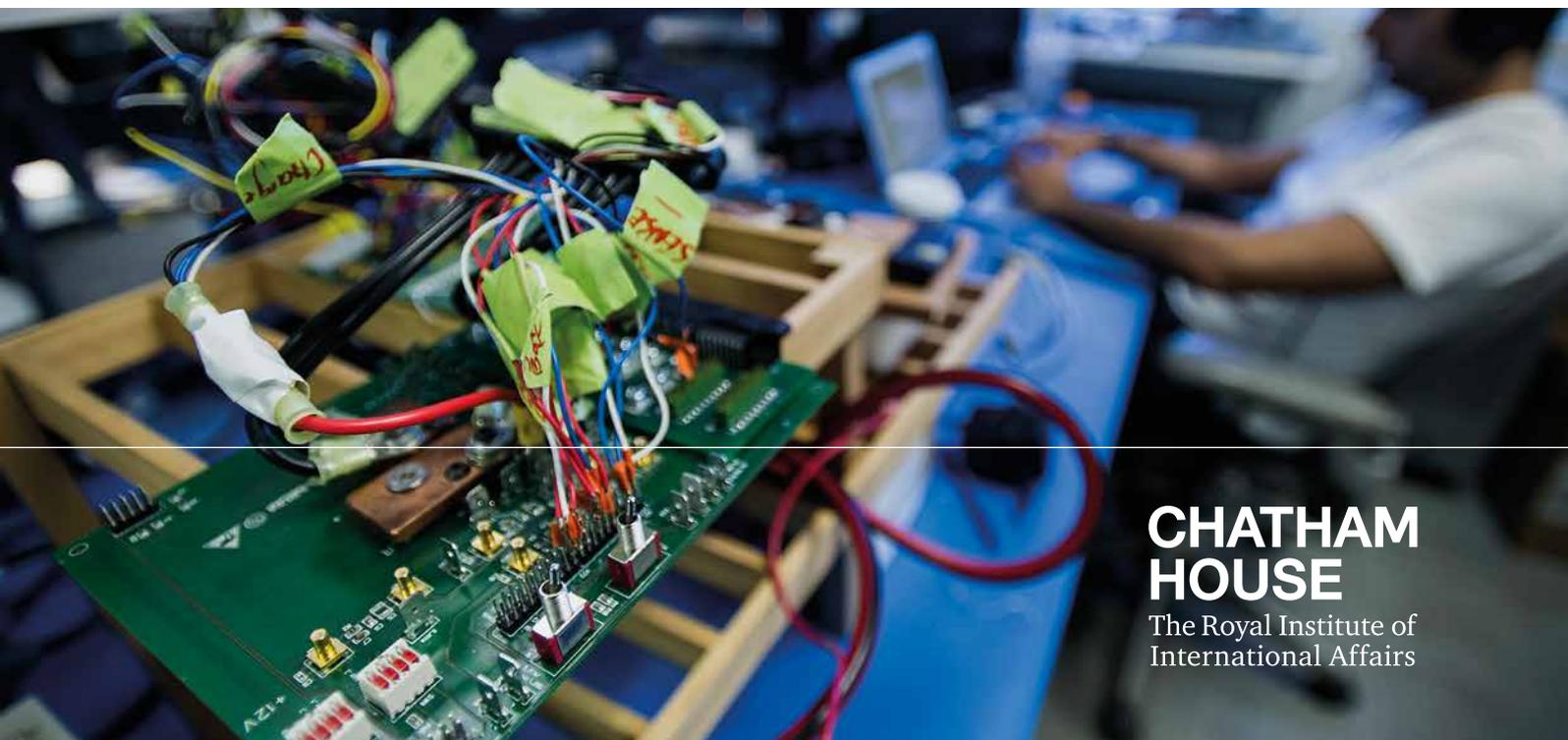
Research Paper

Matthew Brack and Tito Castillo

Centre on Global Health Security | April 2015

Data Sharing for Public Health

Key Lessons from Other Sectors



**CHATHAM
HOUSE**

The Royal Institute of
International Affairs

Contents

Summary	2
Introduction	3
The Individual Perspective in Data Sharing	5
Wider Contexts for Data Sharing	9
Technical Tools	14
Key Lessons for Public Health Data Sharing	18
Acronyms	20
References	21
About the Authors	28

Summary

- Several recent public health emergencies of international concern have demonstrated a worldwide need for improved surveillance data sharing procedures. Lessons learned from other sectors have the potential to inform the future development of data sharing in the public health community.
- At the individual level, the benefits of sharing data have yet to be proven, in any context, and disincentives predominate. Meanwhile, the challenges of achieving truly shared and equitable benefits from data sharing relationships between high-income, and low- and middle-income countries remain largely unrealized.
- The technical solutions to data sharing have already been found and implemented, particularly in the commercial sector, but in many sectors established working models impede the development of a data sharing culture. In contexts where open data is prized, data sharing is no guarantee of transparency and can be highly politicized.
- Sharing public health data is faced with particular problems arising from complex international legislative frameworks that govern intellectual property and data protection. Certain tools, such as open data licences, are allowing data to be shared legally and openly, but the legislative landscape remains poorly understood.
- Global policies for sharing public health data do little to address local concerns, particularly in low- and middle-income countries. Solutions are required that acknowledge the needs of local data collection environments that can extend to supporting public health data sharing on a global scale.

Introduction

When it comes to the challenges involved in public health surveillance data sharing, people are tired of the diagnosis – the barriers have already been well-defined and understood, and the time has now come to direct efforts and resources towards finding and implementing practical solutions. In this spirit, the Centre on Global Health Security at Chatham House – which concerns itself with key global health challenges and their manifestation as problems of foreign policy and international affairs – has support from the Bill & Melinda Gates Foundation to develop guidance on how to create the right environment for public health data sharing and achieve good practice. The aim of this project, Strengthening Data Sharing for Public Health, is to provide guidance that will support pushing the established norms for data sharing towards a model where data are shared as openly as is possible and appropriate.

Background

Since the latter part of the 20th century, increasing globalization has made every nation's public health policy relevant to the wider global health community, as surveillance, pandemic preparedness and response demand coordinated international action.¹ Of these, surveillance is a first line of defence in detecting and identifying emerging health threats,² and a key preventive tool,³ defined within a public health context as a function that ensures that the right information is available at the right time and in the right place to inform public health decisions and actions,⁴ with data sharing required to achieve these aims. Such factors have led to a growing recognition of the need for effective data sharing in the field of public health surveillance, as evidence for its efficacy mounts.⁵ The sustained development and global prevalence of information technology infrastructure and skills continues to raise the potential of global surveillance systems.

In 2005, in the wake of the SARS outbreak, the WHO International Health Regulations (IHR) were revised to lay the foundations of a global surveillance system for public health emergencies of international concern.⁶ The previous IHR, of 1969, had only mandated reports of plague, yellow fever, cholera and smallpox.⁷ Yet the 2005 IHR revision has been only one step forward in a longer journey towards the goal of open and transparent surveillance data sharing. There remains a reluctance to share data at all levels within the public health community, from an individual researcher's data to a health ministry's information regarding the latest outbreak.⁸ A pressing concern is, then, how ready or willing individual organizations or states are to open their data to one another in efforts to counter emerging threats, particularly when realizing equity and reciprocity between high-income countries (HICs) and low- and middle-income countries (LMICs) represents an ongoing challenge.⁹

The costs of not sharing surveillance data are very real. The decision to delay reporting the SARS epidemic in China hindered its containment, with severe socio-political repercussions for the Chinese leadership.¹⁰ More than two years after its discovery, the source of MERS Coronavirus and its transmission mechanism to humans remains unknown, in part because of inadequate data sharing.¹¹ One year into the Ebola outbreak in West Africa, insufficient viral genetic data have been released to the public domain, preventing collaborative efforts to address fundamental questions on transmissibility or on the efficacy of vaccines.¹² In WHO Director-General Margaret Chan's Report to the Special Session of the Executive Board on Ebola, she highlighted the need to improve surveillance capacity.¹³

The Chatham House project on data sharing is situated within a wider climate of opinion on the benefits of data sharing and how it should be conducted. In particular, there has been a protracted debate on data protection, personal privacy and security in the context of the US National Security Agency revelations during 2013 (lending the term ‘surveillance’ distinctly negative connotations), and a push within many countries to share patient information and clinical trial data. While there is broad agreement on the benefits of data sharing, how this balance between privacy and openness is addressed is a key policy question. While the data sharing landscape is diverse and complex, there is a common need across all sectors for high-volume, high-quality data and the capacity for effective analysis.

Approach

This paper has been written to inform roundtables convened to develop guidance through cross-disciplinary dialogue and achieve the aims of the Strengthening Data Sharing for Public Health project. It is published alongside a companion background paper that focuses specifically on the solutions to barriers in public health surveillance data sharing. The paper examines the experiences of data sharing within other sectors as the point of leverage to identify additional solutions for problems confronting the public health sector. By stepping outside that environment and the standard information channels that inform decisions within public health, there arises the potential to examine the problem from fresh perspectives, and to arrive at novel analysis, judgment and decisions on how to solve those challenges.

The paper examines data sharing processes from several perspectives, interspersed with several in-depth case studies representing actual data sharing in practice within different contexts, drawn from interviews with subject experts conducted under the Chatham House Rule.^a Individual perspectives are addressed first, with the understanding that most data sharing processes are ultimately facilitated at the individual level. The study then draws back to explore the wider context for data sharing at the organizational and sector level, acknowledging that it is this wider context that most influences the actions of the individual. Technical tools utilized by other sectors to facilitate data sharing are then examined, including those that address legal barriers. In conclusion, a series of key lessons to be taken from these perspectives and case studies is presented.

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

The Individual Perspective in Data Sharing

The importance of the role of trust between individuals in the sharing and reuse of data cannot be overestimated. Beyond the data producer and secondary user there are also other key individual stakeholders – in some instances data managers, and, frequently, the person whose data is being shared. The prevalence of personal data means that striking the optimal balance between individual privacy and use of data for wider benefit is key for data sharing. Organizations and networks are ultimately composed of individuals – to affect cultural change or strategic shift towards open and transparent data sharing within an organization or community, the significance of individual data sharing skills and personal incentives must be appreciated.

Building skills capacity

In order to produce, prepare and share data of value for sustainable access and reuse, these processes must be planned and resourced in advance. Regardless of the external pressures applied on an individual to share their data, the result of sharing will not be effective without a data management plan, which typically states what data will be created and how, and outlines the terms for sharing and reuse, including any restrictions that may need to be applied.¹⁴ This is particularly pertinent in the scientific research sector, where there is a strong desire to promote open access to discrete data sets that may be of value for future research, thus gaining the maximum benefit from funding. Often these outputs have been publicly funded, and so organizations such as the US National Science Foundation (NSF) demand that all grant proposals must include a data management plan.¹⁵ Open access and open data – data that can be freely used, reused and redistributed by anyone¹⁶ – are becoming more prominent on the global research agenda, and a growing number of funding agencies and publishers around the world are advocating and enforcing data management for open access.¹⁷

Ensuring that this skill is acquired is a joint responsibility shared by policy-makers, researchers and their organizations. Many research institutions provide guidelines to their own researchers to enable them to gain funding, further their research field and mitigate risk. Often facilitated by university libraries and information services, leaders in this area of guidance are in turn referenced by other research institutions globally, spreading best practice. Without the availability of such guidelines and training outreach, it is by no means a given that individual researchers would possess the skills to manage their data properly, even if they wanted to share. By engaging with this planning procedure, however, an individual will necessarily be introduced to the additional technical skills required in data management that will support data sharing.

In many cases, individual and organizational barriers exist to impede the acquisition of data management skills. Traditionally, data management has been the realm of IT departments in many organizations, or no precedent for data management exists at all. In some extreme cases, outright data fraud has led to demands for better oversight and training.¹⁸ Yet this is not a phenomenon confined to the research science sector. Business organizations are increasingly reliant on accurate and timely data, which can only be guaranteed through adequate data management: the role of data analysis is on the increase across all types of organization as the volume of data outputs expands globally from a myriad of sources.¹⁹ Dedicated data managers in many organizations play a significant role in enabling

data sharing, facilitating sharing by colleagues who may not themselves have the time or opportunity to develop such skills. The value of such roles is gaining wider recognition, along with the need to incentivize data managers with stronger career path opportunities.²⁰

Individual incentives to share

The narrative of advocacy for data sharing often appeals to a ‘greater good’, whether that is improved health, the scientific method, or the bottom line. Individual incentives, on the other hand, have not been developed in such detail, and lack a critical mass of high-impact research to prove their efficacy.²¹ Instead – and despite the inherently individual nature of data production – advocates of data sharing are often forced to rely on anecdotal evidence of individual benefits associated with sharing, which tends to be less persuasive. Disincentives are often more apparent: when Chinese scientists sequenced H7N9 influenza A virus in 2013 and shared the data in the Global Initiative on Sharing All Influenza Data (GISAID) database, researchers in the United States initially made moves to publish on the data without crediting their colleagues in China. It took the intervention of GISAID’s president to mitigate the problem, which was blamed on poor ‘etiquette’.²² The protection of individual privacy and consent are also issues of concern for researchers. Furthermore, the overall costs incurred by individual data producers in sharing data, in terms of time, effort and finances, can be considerable.

While some key incentives have been identified, these now require wider endorsement by funding agencies, research institutions and publishers to see their value realized for the individual data producer.²³ The key incentives for individuals to share data can be grouped under issues related to trust, and credit for sharing. Within the wider world of data sharing there is generally perceived to be a ‘data trust deficit’, with the unknown nature and purpose of data reuse driving the distrust.²⁴ This concern over potential misuse also permeates public health data sharing, as data producers may feel they have lost control of their data once it is shared.²⁵ For this reason, a peer-exchange model of data sharing can often manifest as sharing with trusted peers via informal networks. This has been for some time an effective model for sharing pre-publication data within the research domain, providing all the benefits of sharing, albeit within the microcosm of a specific research group or community.

Crucially, data that are considered vulnerable to outside exploitation for publication are often not shared.²⁶ Emerging direct from a general lack of confidence over reuse are specific concerns that others may pre-empt the data producer in publishing with data that have been shared, may publish shared data without crediting its originator, or worse, may misuse the data, leading to the data producer becoming discredited. Consequently, there is a need for community standards to level the playing field and instil confidence – either wielded by external actors with policy influence over a community, or, more successfully, maintained through internal peer pressure arising from established community norms on data sharing. Recent work on rewarding and promoting data sharing has also involved an exploration of data metrics, alongside training and career recognition.²⁷ A growing number of journals, such as Nature Publishing Group’s *Scientific Data*,^b provide open access publication for descriptions of scientifically valuable data sets.

This fear of misuse, or even expropriation, is heightened within research interactions between HICs and LMICs. Within such contexts, alternative policies and norms may be required that are adjusted to the needs of researchers working in LMICs. Touching on both capacity-building and incentives, there

^b <http://www.nature.com/sdata/>.

is an expectation in LMICs that those in HICs wishing to collaborate and access their data should also contribute to building research capacity in the country or region from which data are being sourced. Long-term collaboration facilitates these goals, with twinning between HIC and LMIC institutions promoting training, mentoring and equitable partnerships.²⁸ There have also been some efforts to maintain local ownership of data, such as the Rwandan Ministry of Health's policy requiring local authorship on the publication of any research based on local data.²⁹

Case study 1: Human heredity and health in Africa (H3Africa)

Collaborations between HICs and Africa have suffered from a frequent power imbalance. Emotive accusations of extractive research proliferate,³⁰ with many foreign projects neglecting local research needs.³¹ In the worst cases, foreign research has disrupted existing health services, taking health care providers away from their clinical duties or removing health care and research expertise from a country.³² This case study examines an alternative approach to developing workable data sharing agreements in LMICs.

H3Africa^c is a collaborative programme co-funded by the US National Institutes of Health (NIH) and the Wellcome Trust. It has attempted to remedy the problem of unequal partnership through direct funding of African researchers at African institutions conducting research on the genomic bases of health problems of actual importance to Africa. The programme also includes infrastructure provision, with a bioinformatics network (H3ABioNet) and biorepositories, all located in Africa. Unlike similar repositories, H3ABioNet, based at the University of Cape Town, provides support for researchers in data preparation and analysis. The long-term goal is to enable African scientists to conduct world-class research in their own countries, in compliance with international data sharing standards.³³

While H3Africa researchers are required to make their data openly available, in line with established community standards for genomics data sharing, a specific data sharing policy was drawn up to serve the H3Africa programme, developed by working groups of representatives from each of the 21 funded research projects. Before the policy could be written, a period of training on the basics of data sharing (including, even, the terminology associated with it) was conducted over a period of several months. The oversight required from NIH and the Wellcome Trust was therefore intense. Once the policy was completed, hard copies were signed by members of the H3Africa Consortium to confirm their understanding.

Significantly, this policy differs from the data sharing requirements outlined in NIH and Wellcome Trust policy. Once data are deposited in H3ABioNet, an H3Africa researcher has two months for quality control, after which their data are made available to the H3Africa consortium (but not the public) for nine months. It is then mirrored in the European Genome-phenome Archive and is available via an application process to an independent African Data and Biospecimen Access Committee. This is followed by a 12-month publication embargo, providing a total of 23 months before outside researchers can publish on these data, giving H3Africa researchers almost two years to get their papers out.³⁴ This timeline reflects a compromise by both parties, since there was a desire in Africa for more time to prepare, analyse and publish data.

This approach was based on an understanding that the capacity for quick analysis of genomic data is dependent on having the infrastructure and expertise that are currently still under development in Africa. Another key factor in developing H3Africa policy was the additional sensitivity to the long-standing issue

^c <http://h3africa.org/>.

of power imbalance and unequal partnerships, where African researchers contribute samples and data without receiving collaborative credit. This uncovers a truth behind any 'global' data sharing policy: that its implementation must often incorporate compromise to reflect different realities on the ground, or alternative policies must be generated locally. In the case of H3Africa, a separate policy specific to an African research environment has sought to protect those projects from unfair competition from HICs. Applying the established data sharing policies of the funders involved would have set H3Africa researchers up to fail.

Wider Contexts for Data Sharing

While the data sharing process contains certain core characteristics regardless of data type and provenance, it is the context in which data sharing takes place that will have the most impact on how that sharing process is conducted, and how it will dictate the particular nature of the barriers that present themselves to the development of data sharing norms within any given community. In most cases, the technical barriers to data sharing have been resolved and it simply requires sufficient resources to remedy these. It is the cultural and political hurdles that still await a solution in many organizations and sectors.

Data sharing policy in the research community

Recently, the culture of withholding data in one scientific discipline has publicly backfired: the European Space Agency's (ESA) high-profile landing on Comet 67P in November 2014 saw thousands of public supporters of the programme frustrated by a six-month embargo on pictures from the main camera system on the Rosetta probe tracking the comet. ESA's director-general was himself unable to access these images.³⁵ In a survey of more than 1,000 scientists, the NSF DataONE project concluded that: 'Barriers to effective data sharing and preservation are deeply rooted in the practices and culture of the research process as well as the researchers themselves.'³⁶ In many sectors, traditional practices have not been able to keep pace with wider expectations for data sharing.

A 'crisis of reproducibility' in science research has been brewing for some time now, regularly hitting the mainstream media over the last few years.^{37,38} New models proposed as best practice for research professing alignment with proper scientific method are gaining acceptance. In their manifesto *Improving the Future of Research Communications and e-Scholarship*, the international group of academics, publishers and funders known as Force11⁴ have presented a more inclusive vision of scientific knowledge exchange in a redefinition of the 'research object', where journal articles and research papers are only one form of knowledge exchange, alongside data, software tools and research workflows.³⁹ This definition was subsequently endorsed by the Royal Society.⁴⁰

Aspects of the origins of this cultural shift can be traced to one of the more successful research data sharing communities that arose from some of the most contentious circumstances. Now referred to as the 'human genome wars' that played out during the 1990s, the resulting community norms for genomics data sharing continue to resonate within data sharing initiatives across many sectors, particularly in the natural and social sciences.⁴¹ A key factor in the adoption of the open and collaborative norms for data sharing in genomics was the agreement of the 'Bermuda Principles' in 1996 at a summit of the world's leading gene sequencing groups. This agreement served to bolster the nascent Human Genome Project supported by publicly funded groups led by NIH and the Wellcome Trust, and introduced an obligation for human DNA sequence data to be deposited in an open repository within 24 hours and prior to any journal publication.⁴²

⁴<https://www.force11.org/>.

The outcome of the Bermuda meeting was by no means a foregone conclusion. Going into the discussion, there was no consensus on the best way to proceed, with some organizations, not least NIH, exploring patent applications on key gene sequence markers.⁴³ In the wake of the Human Genome Project, data sharing policies now proliferate throughout research funding bodies, both public and private, with policies at the Wellcome Trust and NIH emerging direct from the genomics data sharing meetings at Bermuda, and subsequently at Fort Lauderdale in 2003.⁴⁴ The genomics sector continues to lead the research community in successful collaborative data sharing through the Global Alliance for Genomics and Health,^e an international coalition established to further genomic medicine through data sharing, and public-private partnerships (PPPs) promoting medical developments through open science.^f

The role of open data

Benefits of open data commonly cited include an increase in the amount of data available to solve complex problems, higher levels of reuse as a result of greater access (producing additional outputs from datasets ‘downstream’) and, consequently, a higher return on investment from funding resulting in open data.⁴⁵ The most widely endorsed definition of open data describes free use, modification and sharing by anyone for any purpose. Open data must also possess an open licence.⁴⁶ Within many data sharing contexts, data outputs will not be able to meet such a standard, often due to data protection or intellectual property concerns dictated by legislation that is beyond the control of the data producer. It is important to note then that data sharing and open data are not synonymous – for example, controlled access to data containing sensitive information via a data access committee is a common approach in scientific research, where access is granted on a case-by-case basis.

If a funding agency wishes to show that it is moving research forward, it must display a robust data sharing policy stipulating that data be made open from its funded research outputs where appropriate, usually accompanied by open access provision for publications, and, increasingly, the sharing of open source software code.⁴⁷ A notable addition to this mix is the recent announcement by the Bill & Melinda Gates Foundation of a new open data and access policy,⁴⁸ described by *Nature* as the world’s strongest policy on open access research.⁴⁹ The strength of the policy itself highlights the problems of implementation within different data sharing contexts, particularly within LMICs, where local journals may not have the open access provision to accommodate compliance with Gates Foundation policy.⁵⁰ If data are made open and theoretically available for everyone, researchers in LMICs may not be able to process, or even access, data as readily as can researchers in HICs. While open data mechanisms promote equal access, this does not necessarily equate to access that is simultaneous and fair.

Nevertheless, the open access movement is gathering pace globally. India’s Ministry of Science and Technology, for instance, now requires researchers who receive even partial funding from its biotechnology, and science and technology departments to deposit copies of their papers in publicly accessible depositories, with data sharing encouraged.⁵¹ Globally, there are now 34 funders who require data deposit and another¹⁶ who encourage it. While data sharing policies have been in place in most funding agencies for many years, enforcement is a resource-intensive exercise that often only begins years after the initial policy is introduced. This can require cultural acceptance of the policy

^e <http://genomicsandhealth.org/>.

^f <http://www.thesgc.org/>.

first within the institution itself, before it can then be effectively implemented to change the culture of the community that it looks to influence.

A key characteristic associated with open data is ‘transparency’ – data that can be accessed and reused to shed light on the underlying activities that have produced that data – but whether data are open or not is no guarantee of their quality for secondary use, and thus transparency is by no means assured. Preparing data for release is costly and time-consuming, and guidelines are needed on the quality of data required for secondary use, which will be dependent on the context of data generation and reuse.⁵² Related to this is the question of whether raw data or prepared data are released: while in many instances some would consider any data release to be welcome, others, such as the recent US Institute of Medicine report *Sharing Clinical Trial Data*, determined that sharing raw data was too burdensome for secondary use and should only be made available on a case-by-case basis.⁵³ In the end, sharing data is necessary, but not in itself sufficient for future reuse.⁵⁴ Data that cannot be reused is at best a waste of resources, and at worst casts doubt on the integrity and competence of the data producer.

The implications of big data: lessons from commercial data sharing

The commercial sector has long grasped the potential of data sharing for business and the need for interoperability, allowing systems and organizations to work together.⁵⁵ As an example, the formation of the Transportation Data Coordinating Committee in 1968 led to a standard for electronic data interchange (EDI) for shipping logistics.⁵⁶ This standard tackled the problem of multiple information formats across different businesses, which needed to be harmonized prior to gaining efficiencies from computer processing. In this instance, a clear business case of cost savings, accuracy and real-time data led to the global adoption of EDI across multiple industries and technology platforms. Widespread adoption has brought with it certain technical barriers, including different EDI standards in various versions which must be utilized in a synchronized way by business parties for data sharing to occur.

More recently, businesses have gained interest in uncovering new data sets for analysis, or in what has become known as ‘big data’. The original definition of big data was offered by Doug Laney of Gartner in 2001, attributing data volume, velocity and variety as key attributes at a time when ‘a rise in merger/acquisition activity, increased collaboration, and the drive for harnessing information as a competitive catalyst’ was pushing businesses to be more aware of how data were managed.⁵⁷ Within the concept of big data is also the notion that such data may previously have been overlooked because of a lack of technology, expertise or motivation to leverage such resources successfully.⁵⁸ Once such data are being mined, big data has the potential to enhance decision-making, insight and work processes, benefiting an organization’s bottom line. It is from this precedent that Apple has launched ResearchKit,⁸ opening up iPhone user health data to further medical research.⁵⁹

The profit motive within business often makes for efficiencies that go unrealized in other sectors, and investment has seen a number of data sharing tools and processes deployed by businesses that facilitate the realization of their products. The fact that a company such as Facebook is so keen to share data across its many applications, is, while at times controversial, indicative of the efficacy of data sharing: Facebook and its affiliates use this data sharing to improve services and increase the value proposition of their business for shareholders. An independent expert advisory group commissioned by UN Secretary-General Ban Ki-Moon to make recommendations on data and

⁸ <https://www.apple.com/researchkit/>.

sustainable development concluded that a ‘data revolution’ to support the post-2015 Sustainable Development Goals would not be possible without the commercial sector.⁶⁰ The World Bank has strongly endorsed this recommendation, stating the need for PPPs for improved data management and sharing, citing Gavi, the Vaccine Alliance^h for equal access to new and underused vaccines as one of the most successful global PPPs.⁶¹

The need for PPPs is also not surprising when one considers the costs associated with data sharing. Data management, done properly, is very resource-intensive: traditional database approaches are unable to scale up or write data fast enough to keep up with data creation, while new solutions introduce more complex management and the need for skilled resources.⁶² The non-profit sector has been exploring big data analytics for several years, and is looking to make savings in cloud services and open source software rather than opting for the more expensive distributed storage and processing frameworks popular for commercial big data analytics.⁶³ Whatever the model, data sharing requires investment: the better the infrastructure and the simpler the process, the more relevant parties will share.

Case study 2: Open government data

The rhetoric of open government presents open data as holding the key to transparency, efficiency and expanded citizen participation in politics and public services. Increasingly, this narrative is switching to one of economic benefit through data reuse in the commercial sector. In 2013 the G8 Open Data Charter acknowledged that the world had reached a ‘tipping point’ at which government data should now be open ‘by default’.⁶⁴ This tipping point – fuelled by the proliferation of information technology, and social media in particular – sees open data as promoting a new collaborative or participatory democracy in which governments and citizens pool expertise to tackle problems, and where ‘citizens are no longer talking about the process: they are the process’.⁶⁵ This case study examines the socio-political implications of dealing with open data.

Open government is a truly international phenomenon. In 2014 the African Development Bank launched its Africa Information Highway to assist African countries in open data management and dissemination through country-specific Open Data Platform tools.⁶⁶ Latin American governments have also embarked on substantial open government reforms in recent years, with many countries working with the Open Government Partnership.⁶⁷ The World Wide Web Foundation recently launched Open Data Lab Jakartaⁱ to explore open data solutions to challenges facing Indonesia and Southeast Asia, modelled on findings from the Foundation’s open government data feasibility studies.⁶⁸ To a greater or lesser extent, such initiatives are being undertaken in many countries, including in Saudi Arabia and elsewhere in the Middle East.

Yet this remains a work in progress. The Open Data Barometer ranks countries on the basis of open government data, and highlights the continuing lack of such data in the performance of key public services.⁶⁹ Indeed, even where open data have been made available, cautionary tales have emerged. As the G8 Open Data Charter was launched, the UK Information Commissioner warned that open data should not be perceived as a substitute for freedom of information, and that a reliance on open data could potentially lead the UK government to focus on publishing information that casts it in a positive light.⁷⁰ Championed by

^h<http://www.opengovpartnership.org/>.

ⁱ<http://labs.webfoundation.org/>.

former prime minister Gordon Brown, and Tim Berners-Lee, the inventor of the World Wide Web, data.gov.uk was launched in 2009 to share non-personal UK government data as open data. A year later it was alleged that poor data quality was hindering the programme.⁷¹

Open government data is characterized by a tension between simply making data open, regardless of quality, via a remote portal for technical users, and sharing quality data that retains its social context through direct interaction between citizens and a government data producer. In the latter instance, the data producer would be expected to interact with the secondary user to enable them to engage with and fully utilize the government data. Currently, the focus is on ‘data dumping’ – i.e. sharing non-sensitive data sets, and relying on technical users to interpret them. Within LMIC settings, this approach may render government data producers susceptible to the type of expropriation already outlined in the H3Africa case study above. An optimal system for engagement with open data has been proposed, where open data should be demand-driven, be placed in context, be supported by comment and conversation around data sets, and provide provision for building capacity, skills and networks, alongside collaboration on data as a common resource.⁷²

Open data is highly politicized and has opened the door to a new debate on how governments collect data, giving rise to the term ‘data discrimination’.⁷³ As an example, in 2010 the Canadian government made elements of its census voluntary, a result of which was the increased isolation and invisibility of already marginalized communities in terms of public services provision.⁷⁴ Even with high quality data available, a ‘data divide’ has been observed between those technically able to work with available data and those without such skills, who often don’t even know that the data are there in the first place. When the Indian state of Karnataka digitized around 20 million land titles in an effort to make them more accessible, the main beneficiaries were the wealthy, who used the newly shared data to challenge deeds, exploit errors in documentation and snap up property.⁷⁵ In LMICs information technology has the potential to reinforce inequality as much as to counter it.⁷⁶

For effective open government data sharing, two components are critical: a strong civil society with adequate technological infrastructure, and a technical capacity within a willing government to launch such an initiative. In instances where only the latter exists, a well-intentioned programme of open government data may simply find no audience because of lack of infrastructure, skills or political engagement.⁷⁷ In some cases, where a government has followed the form of making data open, an audience is lacking because of the absence of a functional civil society, resulting from government-imposed restrictions and censorship. The next major hurdle for open government data involves negotiating the balance between openness and privacy – data of most interest frequently entails privacy issues – in the context of expanding data protection legislation around the world and the recognized failure of anonymization techniques.⁷⁸

Technical Tools

The technical tools associated with data sharing here include not only products of information technology, such as standards to facilitate sharing, but also methods to address the legislative framework that governs it. Both require interoperability and international collaboration to function effectively.

Standards for data sharing

The ability to reuse shared data relies on good documentation, or metadata (best understood as ‘data about data’), to specify the location and meaning of variables contained within a data set. This must also describe the sample and sampling procedures, as well as other contextual information.⁷⁹ The Data Documentation Initiative (DDI) metadata standard has been adopted for describing public health and demographic surveys, and is utilized by the World Bank, UNICEF and the UK Medical Research Council.⁸⁰ The establishment of DDI was a response to the need to harmonize non-standard documentation between data archives, so as simultaneously to improve data access and assist data producers in packaging and disseminating data for sharing.⁸¹ DDI has been recommended by a report to the Wellcome Trust’s Public Health Research Data Forum as the preferred basis for a centralized domain portal for public health and epidemiology research.⁸²

Work on DDI began in 1995, as an international effort led by the US Inter-university Consortium for Political and Social Research, and was first published in 2000. Standards are an ongoing process requiring continuous updates, particularly as new technologies emerge. DDI’s evolution has been overseen by the DDI Alliance^l, a self-sustaining collaborative organization whose members have a voice in the development of the DDI specification. To promote wider interoperability, DDI also maps to other standards and is compliant with the ISO/IEC 11179 international standard for representing metadata. Establishing interoperable standards like DDI also generates new open source software tools, to enable implementation by data producers. For example, DdiEditor is a data processing tool developed by the Danish Data Archive for use with DDI that is freely available from the Google Code open software repository.^k

Where large quantities of data are being aggregated from disparate sources, data harmonization is often required to create a coherent, reusable dataset. As an example, the Integrated Public Use Microdata Series (IPUMS) is the world’s largest individual-level population database, comprising data samples from international census records collected since 1960. The records are converted into a consistent format and made available to researchers through a web-based data dissemination system. This format is ASCII (American Standard Code for Information Interchange), first introduced in 1963. Such a simple encoding scheme allows researchers to load data with ease into statistical computing environments for analysis. Command files provided by IPUMS read ASCII data into popular proprietary statistical software, while scripts are also available to load ASCII into open source software available from open source repositories. Many examples of non-proprietary data analysis ecosystems proliferate: the tools are available for data sharing and collaborative research, with many well-established standards for data and open source software for data interpretation and analysis.

^l<http://www.ddialliance.org/alliance>.

^k<https://code.google.com/p/ddieditor/>.

Open licensing: overcoming intellectual property barriers

In most jurisdictions there are intellectual property rights in data that prevent third parties from using, reusing and redistributing without explicit permission.⁸³ In some cases, recent legislation has actually made sharing harder, as with the European Database Directive of 1996 relating to the legal protection of databases. Open data licensing appears to provide a potential solution to the problem of legal clarity and clear signage for reuse by third party users, particularly across international borders. Because of the automatic nature of copyright, and with the intellectual property position of databases and data sets being particularly complex, a licence or waiver provides a simple and effective way to make data truly open, and clarifies the exact manner of its reuse. Advocacy for open data licensing is therefore considered a high priority for multinational data sharing initiatives in particular.⁸⁴ There is a growing precedent for the use of open data licences within the research community, particularly in science publishing and data repositories.⁸⁵

In order to make data open, according to the Open Knowledge Foundation, the first step is to apply an open licence. This ‘legal openness’ must come before the ‘technical openness’ of actually making the data available.⁸⁶ Such legal openness is the principal advantage of licensing open data, followed by the clarity for reuse that such a licence provides. According to the Open Data Institute: ‘the key is how the data is licensed. Data that doesn’t explicitly have a license is not open data.’⁸⁷ It is important to note that no open licence is ever able to provide complete clarity or be fully open unless it dedicates data to the public domain. Even the popular Creative Commons Attribution Only (CC-BY) licence suffers from the problem of ‘attribution stacking’, where compiling a data set from multiple other data sets may be problematic because of the administrative burden of crediting each individual contributor to the superset. For data to be truly open, data producers must waive all of their rights, including attribution, which some data producers will be unwilling to accept. Supporters of open licensing argue that the right to attribution is based on professional ethics, irrespective of any waiver, and that such licences are the best way to ensure the future research potential of data sets.

Navigating the data protection landscape

Data protection creates a particular barrier for the sharing of public health surveillance data, since almost all such data are collected without any knowledge or consent. While the processing of personal data concerning health has to date enjoyed exception status in US and European Union (EU) data protection legislation, this legislative landscape is changing worldwide: in January 2012, the European Commission introduced a draft European General Data Protection Regulation that will supersede the 1996 Data Protection Directive, now perceived to be out of date. The existing Directive has allowed member states to determine their own data protection legislation, providing for a varied and complex overall framework. The new regulation will be directly binding for member states, harmonizing EU data protection. However, with debates in the European Parliament and Council ongoing, the outcome for public health surveillance data is as yet unknown.

Meanwhile, the African Union has proposed its own data sharing legislation in line with the EU model, the Convention on Cyber Security and Personal Data Protection, paving the way for the creation of new governmental offices enforcing data protection within participating states. Since most countries have their own unique data protection legislation, differences in legislation from multiple jurisdictions need to be considered when transferring data across international borders. If data are being brought from one jurisdiction to another, such as from outside Europe into the United Kingdom, data

collection and processing must comply with the laws of the source country. It is possible that a UK recipient will not be obtaining the data fairly and lawfully under UK legislation if they are aware that processes have breached overseas legislation.⁸⁸

A recent examination of the use of mobile phone data for humanitarian goals has found that an international cross-sector approach is needed to agree best practice guidelines for privacy-conscientious data sharing, while accepting that data can never be fully de-identified.⁸⁹ Despite a widespread belief in the effectiveness of de-identification, or anonymization, research has shown that re-identification is often possible, resulting in mistaken assumptions regarding data security achieved through de-identification that pervade privacy law, regulation and debate.⁹⁰ Given that data can never be completely de-identified, a constant balance needs to be found on a case-by-case basis between the risk of re-identification and the potential for greater social good.⁹¹ This reality, combined with increasing prevalence of data protection legislation around the world, makes for a challenging legal and policy environment for sharing health data, and in particular for conducting research with those data.

Given this situation, other tools are sometimes necessary to navigate data protection legislation. One approach has been the provision of access to sensitive data sets within safe settings (a controlled physical location), or data enclaves (which may be distributed but secure networks). In both these cases, access is supervised by the data custodians. For example, the UK Medical Research Council's National Survey of Health and Development has allowed access to particularly sensitive information only on its own premises.⁹² Beyond these tangible measures, there has yet to be any large-scale effort to attempt to analyse the current data protection landscape and to create some form of unified guidance. This is a problem in individual states (in the United Kingdom, for example, there is a common law duty of confidence for medical professionals, alongside the 1998 Data Protection Act, with no harmonized guidance on how these legal instruments fit together), never mind in terms of the EU or other supranational scenarios.

Case study 3: NHS England's care.data programme

The aims of NHS England's care.data programme were to improve efficiency, patient care and further medical research. The programme enjoyed broad support from the medical community, many charities and some of the most influential patient groups in the United Kingdom. Yet this case study will present care.data as an example of failure in data sharing negotiation and implementation. As it stands now, care.data will only proceed with a pilot stage consisting of a small-scale extraction of primary care data deposited in a secure data store with limited access allowed to the Health and Social Care Information Centre (HSCIC). There is a concern that this pilot is not scalable, and that any wider access to data will be limited in the short term. If the pilot fails, it seems unlikely that care.data has any future in its current form.⁹³ Previous broad support for the programme has significantly weakened, and has contributed to declining public trust in the confidential nature of the NHS itself.⁹⁴

Two principal causes underlie this failure. First, the Care.data Inquiry, organized by the Patients Association and chaired by an All-Party Parliamentary Group, concluded that 'the public had been inadequately consulted in the early stages of the Care.data programme and that it was therefore correct to halt the programme to allow further public consultation'.⁹⁵ This included open consultation on the content of the data sets themselves to verify that the data would actually meet the programme's intended aims. The Independent

Information Governance Oversight Panel, which advises the NHS and the UK government on patient data, also indicated that the original process was rushed.⁹⁶ An expert panel convened by the Nuffield Council on Bioethics recommended that the technical limitations of data privacy should be explained to patients.⁹⁷

Second, no delays or pilot stage would have been necessary if proper safeguards had been in place at the time of the original programme announcement. In a Review of Data Releases by the NHS Information Centre (HSCIC's predecessor organization, which handled NHS data releases until 31 March 2013), a number of lapses were clearly identified. Among them, commercial data management partners had released data without consent, access to data by researchers was being provided without due process, and in two instances a data recipient organization could not be identified.⁹⁸ While HSCIC has worked to implement the recommendations of the review, enhanced safeguards may come too late to save the current care.data initiative.⁹⁹

It is likely that the ultimate cause of this lack of consultation and procedural oversight was an overriding focus on the benefits of the scheme that distracted planners from the existing failures within their system, and from how the public would react. This may also explain why the 'opt-out' mechanism for patients to remove their data from the scheme was confusing. With the benefits seeming so apparent, there was probably no realistic consideration that a sizeable number of patients might wish to opt out. Patients were assured that they could opt out of data sharing without it affecting their care, when in fact data relating to critical services such as cancer screening were included in the scope of the programme without alternative provision being made for non-participating patients. As such, it transpired that objections raised by patients were disregarded by the NHS.¹⁰⁰

While secondary data, such as Hospital Episode Statistics (HES), had already been shared for years, the sharing of primary health care data from general practitioners proved to be a more sensitive public issue, raising strong concerns that this might fall into the hands of commercial entities. Chronically poor communication meant that the majority of patients were probably never aware of prior HES data sharing. Going into the launch of care.data, there was very little understanding by the public of how their data were being shared, if indeed most people were aware of the programme at all. Thus, after inadequate communication and consultation, the launch acted to lift the lid on a Pandora's box for the NHS, albeit one that has led to much-needed, if costly, reform.

Key Lessons for Public Health Data Sharing

Data sharing is happening, and on an increasing scale. Technically, effective data sharing is possible within any sector, and has been for some time – the question being whether resources are available to implement those technical solutions. This resource limitation extends to infrastructure and skills capacity, as well as financial constraints, and exists within both HICs and LMICs, although the fundamental digital literacy required to engage with data sharing in policy and practice may be entirely absent in certain LMIC settings.

Data sharing policy and guidance generated in HICs consequently may have little relevance on the ground in LMICs. Furthermore, the wide enthusiasm for the benefits of data sharing has not been matched by a clear value proposition that accommodates the concerns of individuals, organizations and countries, and motivates them to share their data. The resource reallocation and cultural change required to establish data sharing is a long-term challenge, entailing sustained engagement and proven benefits for all stakeholders involved. Beyond the commercial sector, determining exactly who is responsible for managing data so that it can be shared, as well as the actual process of sharing itself, merits consideration.

Lessons are presented following the order of enquiry within this paper.

Lesson 1: Harness the political momentum for action

The Ebola outbreak in West Africa has heightened popular awareness of global disease surveillance and response, presenting a key opportunity to push the public health surveillance agenda forward. Progress often results from crises. Outside public health, the UK's Open Government data platform came hot on the heels of the parliamentary expenses scandal; the Bermuda Principles were agreed after gene sequence patenting caused deep concern within the genomics community.

Lesson 2: Push the wider benefits

The case for data sharing has repeatedly been won based on strong ethical arguments for wider societal benefits. It is for this reason that public health surveillance data sharing enjoys exceptions within personal data legislation, while use of personal data for research remains contentious. Pursuing this ethical argument goes hand in hand with mitigating the negative consequences of sharing for both individuals and organizations, which has received less attention.

Lesson 3: Clarify the value proposition

Many initiatives fail, regardless of sector, as a result of never truly articulating a clear or compelling value proposition. Establishing a substantive value proposition is critical if the journey from initial concept to successful enterprise is to be completed. This is the basis for 'buy-in' from stakeholders – both individuals and organizations – providing them with something tangible to align with once their motivations are understood.

Lesson 4: Consider local policy for LMICs versus global policy

Where data sharing is mandated by policy, implementation of that policy and enforcement present another layer of activity and resource commitment. In most cases, data sharing policies take a number of years to 'bed in' before they are actually implemented and enforced – and this is in HICs where resources are not lacking. The prospect for implementing the same policy in LMICs is one of outright failure; thus, compromise and alternative policy should be investigated.

Lesson 5: Maintain collaborative engagement

Successful data sharing communities are highly collaborative. Collaboration holds the key to producing and abiding by community standards, and building and maintaining productive networks, and is by definition the essence of data sharing itself. Time should be invested in establishing and sustaining collaboration with all stakeholders concerned with public health surveillance data sharing.

Lesson 6: Enhance data sharing literacy in LMICs

Before data sharing collaborations can occur between HICs and LMICs, the basic foundations of data sharing literacy need to be in place. This comes before any data sharing policy or infrastructure development, and paves the way for an equitable collaboration in the establishment of both. This will also enable LMIC data producers to engage with the value proposition of data sharing.

Lesson 7: Make data sharing easy and equitable

Data producers anywhere are more likely to share when simple and effective infrastructure is available and they have the training to use it. This can reduce impact on their time and on the associated costs of data sharing. From an LMIC perspective, data should be retained locally, even if they are mirrored in an HIC. Long-term collaborations between HIC and LMIC institutions increase the likelihood of equitable data sharing.

Lesson 8: Data sharing is not sufficient for reuse

Sharing in itself is not necessarily of any value. In this sense, quality is more important than quantity. The ability to prepare data suitable for reuse presents its own set of problems, but these can be mitigated if data are conscientiously shared in adherence to a standard data management plan. Successful sharing for reuse is also facilitated by established standards.

Lesson 9: Standards and norms need to be actively maintained

Once established, technical standards and community norms need to be actively maintained and developed. The peer pressure of an established community data sharing norm is one of the strongest drivers to share data in any given environment. Incentives will also manifest when a community expects to see data shared, increasing the likelihood of appropriate credit and prestige being assigned to the data producer.

Lesson 10: Data protection is a significant barrier, yet is poorly understood

Data sharing across international borders remains a significant legal problem because of the presence of different legislation in each state. Even when broader legislation is present, lack of clarity in its application – such as legal exceptions for public health data – can impede sharing. The data protection landscape requires guidelines for public health data sharing, and potential lobbying for clarity and harmonization as new legislation is developed.

Acronyms

ASCII	American Standard Code for Information Interchange
CC-BY	Creative Commons Attribution Only
DDI	Data Documentation Initiative
EDI	electronic data interchange
ESA	European Space Agency
EU	European Union
GISAID	Global Initiative on Sharing All Influenza Data
HES	Hospital Episode Statistics
HICs	high-income countries
HSCIC	Health and Social Care Information Centre
IHR	International Health Regulations
IPUMS	Integrated Public Use Microdata Series
ISO/IEC	International Organization for Standardization/International Electrotechnical Commission
IT	information technology
LMIC	low- and middle-income countries
MRC	Medical Research Council
NHS	National Health Service
NIH	National Institutes of Health
NSF	National Science Foundation
PPPs	public-private partnerships
UNICEF	United Nations Children's Fund
WHO	World Health Organization

References

1. St. Louis M. Global Health Surveillance. *CDC's Vision for Public Health Surveillance in the 21st Century – CDC Morbidity and Mortality Weekly Report*. [Online] 2012; 61: 15–19. Available from: <http://www.cdc.gov/mmwr/pdf/other/su6103.pdf> [Accessed 24 February 2015].
2. Steinhardt B. Emerging Infectious Diseases: Consensus on Needed Laboratory Capacity Could Strengthen Surveillance. In: Davis J, Lederberg J (eds.) *Public Health Systems and Emerging Infections: Assessing the Capabilities of the Public and Private Sectors: Workshop Summary*. [Online] Washington, DC: National Academies Press; 2000: 90–96. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK100254/> [Accessed 24 February 2015].
3. Heymann D. Control, elimination, eradication and re-emergence of infectious diseases: getting the message right. *Bulletin of the World Health Organization*. [Online] 2006; 84(2): 81–160. Available from: <http://www.who.int/bulletin/volumes/84/2/editorial10206html/en/> [Accessed 24 February 2015].
4. Department of Health and Public Health England Transition Team. *Public Health Surveillance: Towards a Public Health Surveillance Strategy for England*. [Online] London: Department of Health; 2012. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213339/Towards-a-Public-Health-Surveillance-Strategy.pdf [Accessed 14 March 2015].
5. Ross E. *Perspectives on Data Sharing in Disease Surveillance*. Chatham House Centre on Global Health Security Research Paper; [Online] London: The Royal Institute of International Affairs. 2014. Available from: <http://www.chathamhouse.org/publications/papers/view/199274> [Accessed 24 February 2015].
6. Baker M, Fidler D. Global Public Health Surveillance under New International Health Regulations. *Emerging Infectious Diseases*. [Online] 2006; 12(7): 1058–1065. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3291053/> [Accessed 24 February 2015].
7. World Health Organization. *International Health Regulations (1969)*. 3rd edition. [Online] Geneva: WHO; 1983. Available from: <http://whqlibdoc.who.int/publications/1983/9241580070.pdf?ua=1> [Accessed 24 February 2015].
8. Sane J, Edelstein M. *Overcoming Barriers to Data Sharing in Public Health: A Global Perspective*. Chatham House Centre on Global Health Security Research Paper. [Online] London: The Royal Institute of International Affairs. London. April 2015. Available from <http://www.chathamhouse.org/publication/overcoming-barriers-data-sharing-public-health-global-perspective>.
9. Fidler D, Gostin L. WHO's Pandemic Influenza Preparedness Framework: A Milestone in Global Governance for Health. *Journal of the American Medical Association*. [Online] 2011; 306 (2): 200. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1890912 [Accessed 27 February 2015].
10. Huang Y. The SARS Epidemic and its Aftermath in China: A Political Perspective. In: Knobler S, Mahmoud A, Lemon S, Mack A, Sivitz L, Oberholtzer K (eds.) *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary*. [Online] Washington, DC: National Academies Press; 2004: 116–136. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK92462/pdf/TOC.pdf> [Accessed 24 February 2015].

11. McNabb S, Shaikh A, Nuzzo J, Zumla A, Heymann D. Triumphs, Trials, and Tribulations of the Global Response to MERS Coronavirus. *The Lancet Respiratory Medicine*. [Online] 2014; 2(6). Available from: <http://www.upmchealthsecurity.org/our-work/publications/triumphs-trials-and-tribulations-of-the-global-response-to-mers-coronavirus> [Accessed 24 February 2015].
12. Briggs H. *Ebola: Call for more sharing of scientific data*. [Online] Available from: <http://www.bbc.co.uk/news/science-environment-31091816> [Accessed 25 February 2015].
13. Chan M. *Report by the Director-General to the Special Session of the Executive Board on Ebola*. [Online] World Health Organization; 2015 [Online] Available from: <http://www.who.int/dg/speeches/2015/executive-board-ebola/en/> [Accessed 25 February 2015].
14. Digital Curation Centre. *Data Management Plans*. [Online] Available from: <http://www.dcc.ac.uk/resources/data-management-plans> [Accessed 14 March 2015].
15. National Science Foundation. *Grant Proposal Guide*. [Online] Washington, DC: NSF; 2011. Available from: http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp [Accessed 25 February 2015].
16. Open Knowledge Foundation. *The Open Definition*. [Online] Available from: <http://opendefinition.org/> [Accessed 14 March 2015].
17. Hahnel M. 2015 – The Year of Open Data Mandates. *Figshare*. Weblog. [Online] Available from: http://figshare.com/blog/2015_The_year_of_open_data_mandates/143 [Accessed 25 February 2015].
18. Stapel Investigation. *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel*. [Online] Tilburg University, University of Amsterdam, University of Groningen; 2012. Available from: https://www.commissielevelt.nl/wp-content/uploads_per_blog/commissielevelt/2013/01/finalreportLevelt1.pdf [Accessed 25 February 2015].
19. Cukier K. Data, data everywhere. *The Economist*. Special Report: Managing Information. [Online] Available from: <http://www.economist.com/node/15557443> [Accessed 25 February 2015].
20. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bulletin of the World Health Organization*. [Online] 2010; 88(6): 462–466. Available from: <http://www.who.int/bulletin/volumes/88/6/09-074393/en/>
21. Noorman M, Kalaitzi V, Angelaki M, Tsoukala V, Linde P, Sveinsdottir T, et al. *Institutional barriers and good practice solutions*. [Online] RECODE Project; 2014. Available from: <http://recodeproject.eu/wp-content/uploads/2014/09/RECODE-D4.1-Institutional-barriers-FINAL.pdf> [Accessed 25 February 2015].
22. Butler D, Cyranoski D. Flu papers spark row over credit for data. *Nature*. [Online] 2013; 497: 14–15. Available from: <http://www.nature.com/news/flu-papers-spark-row-over-credit-for-data-1.12901> [Accessed 25 February 2015].
23. Expert Advisory Group on Data Access. *Establishing Incentives and Changing Cultures to Support Data Access*. [Online] EAGDA; 2014. Available from: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtp056495.pdf [Accessed 25 February 2015].
24. Ipsos MORI. *New research finds data trust deficit with lessons for policymakers*. [Online] Available from: <https://www.ipsos-mori.com/researchpublications/researcharchive/3422/New-research-finds-data-trust-deficit-with-lessons-for-policymakers.aspx> [Accessed 25 February 2015].
25. Rudolph R, Davis R. Administrative Data and Disease Surveillance: An Integration Toolkit. In: *Public Health Data Dissemination Guidelines: NAHDO Working Technical Paper Series*. Shah G (ed.) Salt Lake City: The National Association of Health Data Organizations; 2005: 18.

26. Van den Eynden V, Bishop L. *Incentives and motivations for sharing research data, a researcher's perspective*. [Online] Knowledge Exchange; 2014. Available from: http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf [Accessed 25 February 2015].
27. Costas R, Meijer I, Zahedi Z, Wouters P. *The Value of Research Data Metrics for datasets from a cultural and technical point of view*. [Online] Knowledge Exchange; 2013. Available from: <http://www.knowledge-exchange.info/datametrics> [Accessed 25 February 2015].
28. Kaaya E, Macfarlane S, Mkony C, Lyamuya E, Loeser H, Freeman P, et al. Educating Enough Competent Health Professionals: Advancing Educational Innovation at Muhimbili University of Health and Allied Sciences, Tanzania. *PLoS Medicine*. [Online] 2012; 9(8). Available from: <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001284> [Accessed 25 February 2015].
29. Republic of Rwanda Ministry of Health. *Guidelines for Researchers Intending to Do Health Research in Rwanda*. [Online] Republic of Rwanda Ministry of Health; 2012. Available from: <http://www.moh.gov.rw/fileadmin/templates/Docs/Researchers-Guidelines.pdf> [Accessed 25 February 2015].
30. Wolffers I, Adjei S, van der Drift R. Health Research in the Tropics. *Lancet*. [Online] 1998; 351 (9116): 1652–1654. Available from: <http://www.sciencedirect.com/science/article/pii/S0140673697102379> [Accessed 25 February 2015].
31. Binka F. Editorial: north-south research collaborations: a move towards a true partnership? *Tropical Medicine & International Health*. [Online] 2005; 10: 207–209. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3156.2004.01373.x/pdf> [Accessed 25 February 2015].
32. Edejer T. North-South research partnerships: the ethics of carrying out research in developing countries. *British Medical Journal*. [Online] 1999; 319: 438–441. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1127046/> [Accessed 25 February 2015].
33. Collins F, Ferrar J. Opinion: On “Funding Research in Africa”. *The Scientist*. [Online] Available from: <http://www.the-scientist.com/?articles.view/articleNo/41549/title/Opinion--On--Funding-Research-in-Africa/> [Accessed 25 February 2015].
34. de Vries J, Tindana P, Littler K, Ramsay M, Rotimi C, Abayomi A. The H3Africa policy framework: negotiating fairness in genomics. *Trends in Genetics*. [Online] Available from: [http://www.cell.com/trends/genetics/fulltext/S0168-9525\(14\)00196-6#/trends/genetics/fulltext/S0168-9525\(14\)00196-6](http://www.cell.com/trends/genetics/fulltext/S0168-9525(14)00196-6#/trends/genetics/fulltext/S0168-9525(14)00196-6) [Accessed 25 February 2015].
35. Amos J. *Rosetta will prompt science images rethink*. [Online] Available from: <http://www.bbc.co.uk/news/science-environment-30859411> [Accessed 25 February].
36. Tenopir C, Allard S, Douglass K, Aydinoglu A, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. *PLoS ONE*. [Online] 2011; 6 (6). Available from: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101> [Accessed 25 February 2015].
37. The Economist. Unreliable Research: Trouble at the lab. *The Economist*. Briefing. [Online] Available from: <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble> [Accessed 25 February 2015].
38. Puniewska M. Scientists have a sharing problem. *The Atlantic*. [Outline] Available from: <http://www.theatlantic.com/health/archive/2014/12/scientists-have-a-sharing-problem/383061/> [Accessed 25 February 2015].
39. Allen B, Birukou A, Blake J, Bourne P, Shum S, Burns G, et al. *Force11 White Paper: Improving The Future of Research Communications and e-Scholarship*. [Online] Force11; 2012. Available from: http://www.force11.org/sites/default/files/book_attachments/Force11Manifesto20120219.pdf [Accessed 25 February 2015].

-
40. The Royal Society. *Science as an open enterprise*. [Online] The Royal Society Science Policy Centre; 2012: 31. Available from: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf [Accessed 25 February 2015].
 41. Jasny B. Realities of data sharing using the genome wars as case study - an historical perspective and commentary. *EPJ Data Science*. [Online] 2013; 2 (1). Available from: <http://link.springer.com/article/10.1140/epjds13/fulltext.html> [Accessed 25 February 2015].
 42. Wellcome Trust. *Statement on Genome Data Release*. Policy statement. [Online] Available from: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002751.htm> [Accessed 25 February 2015].
 43. Contreras J. Bermuda's Legacy: Policy, Patents and the Design of the Genome Commons. *Minnesota Journal of Law, Science & Technology*. [Online] 2011; 12: 61. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1667659 [Accessed 25 February 2015].
 44. Wellcome Trust. *Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility*. Report of a meeting organized by the Wellcome Trust and held on 14–15 January 2003 at Fort Lauderdale, USA. [Online] Available from: <http://www.genome.gov/pages/research/wellcomereport0303.pdf> [Accessed 25 February 2015].
 45. Janssen M, Charalabidis Y, Zuiderwijk A. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*. [Online] 2012; 29 (4): 258–268. Available from: http://repository.tudelft.nl/assets/uuid:952c8b0e-c55e-4443-9b41-14c38806a840/Author_version_-_Benefits_Adoption_Barriers_and_Myths_of_Open_Data_and_Open_Government.pdf [Accessed 19th March 2015].
 46. Lord P, MacDonald A, Sinnott R, Ecklund D, Westhead M, Jones A. *Large-scale data sharing in the life sciences: Data standards, incentives, barriers and funding models (The "Joint Data Standards Study")*. UK e-Science Technical Report Series. [Online] National e-Science Centre; 2005. Available from: http://www.nesc.ac.uk/technical_papers/UKeS-2006-02.pdf [Accessed 25 February 2015].
 47. Stodden V. Reproducible Research for Scientific Computing: Tools and Strategies for Changing the Culture. *Computing in Science and Engineering*. [Online] 2012; 14 (13). Available from: <https://web.stanford.edu/~vcs/papers/CiSE2012-LMS.pdf> [Accessed 25 February 2015].
 48. Bill & Melinda Gates Foundation. *Bill & Melinda Gates Foundation Open Access Policy*. Policy statement. [Online] Available from: <http://www.gatesfoundation.org/how-we-work/general-information/open-access-policy> [Accessed 25 February 2015].
 49. Van Noorden R. Gates Foundation announces world's strongest policy on open access research. *Nature News Blog*. [Online] Available from: <http://blogs.nature.com/news/2014/11/gates-foundation-announces-worlds-strongest-policy-on-open-access-research.html> [Accessed 25 February 2015].
 50. Rabesandratana T. Gates' strict open access policy may have domino effect. *SciDevNet*. [Online] Available from: <http://www.scidev.net/global/publishing/news/gates-open-access-policy.html> [Accessed 25 February 2015].
 51. Chaa D. India's major science funders join open-access push. *Science*. [Online] Available from: <http://news.sciencemag.org/asiapacific/2014/12/india-s-major-science-funders-join-open-access-push> [Accessed 26 February 2015].
 52. Mladovsky P, Mossialos E, McKee M. Access to research data in Europe. *Eurohealth*. [Online] 2008; 14 (1): 15–17. Available from: <http://www.lse.ac.uk/LSEHealthAndSocialCare/pdf/eurohealth/VOL14No1/MladovskyMossialosMcKee.pdf> [Accessed 26 February 2015].

-
53. Krumholz H. Why data sharing should be the expected norm. *British Medical Journal*. [Online] 2015; 350. Available from: <http://www.bmj.com/content/350/bmj.h599> [Accessed 26 February 2015].
 54. Peer L. Mind the gap in data reuse: Sharing data is necessary but not sufficient for future reuse. *LSE Impact Blog*. [Online] Available from: <http://blogs.lse.ac.uk/impactofsocialsciences/2014/03/28/mind-the-gap-in-data-reuse/> [Accessed 26 February 2015].
 55. Gasser U, Palfrey J. *Breaking Down Digital Barriers: When and How ICT Interoperability Drives Innovation*. [Online] Berkman Publication Series; 2007. Available from: <http://cyber.law.harvard.edu/interop/pdfs/interop-breaking-barriers.pdf> [Accessed 16th March 2015].
 56. Clarke R. *Electronic Data Interchange (EDI): An Introduction*. [Online] Available from: <http://www.rogerclarke.com/EC/EDIIntro.html> [Accessed 26 February 2015].
 57. Laney D. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. [Online] META Group; 2001. Available from: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [Accessed 26 February 2015].
 58. Ward J, Barker A. *Undefined By Data: A Survey of Big Data Definitions*. [Online] Available from: <http://arxiv.org/pdf/1309.5821v1.pdf> [Accessed 26 February 2015].
 59. Cortez M, Chen C. Thousands Have Already Signed Up for Apple's ResearchKit. *Bloomberg*. [Online] Available from: <http://www.bloomberg.com/news/articles/2015-03-11/apple-researchkit-sees-thousands-sign-up-amid-bias-criticism> [Accessed 16 March 2015].
 60. UN Data Revolution Group. *A World that counts: Mobilising the data revolution for sustainable development*. [Online] United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development; 2014. Available from: <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf> [Accessed 26 February 2015].
 61. Ballivian A, Hoffman W. *Public-Private Partnerships for Data: Issues Paper for Data Revolution Consultation*. [Online] World Bank; 2015. Available from: http://data.worldbank.org/sites/default/files/issue-paper-financing-the-data-revolution-ppps_0.pdf [Accessed 26 February 2015].
 62. Bantleman J. The Big Cost of Big Data. *Forbes*. [Online] Available from: <http://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data/> [Accessed 26 February 2015].
 63. Saran C. How to slash the cost of big data analytics. *ComputerWeekly*. [Online] Available from: <http://www.computerweekly.com/news/2240215936/Interview-how-to-slash-the-cost-of-big-data-analytics> [Accessed 26 February 2015].
 64. UK Cabinet Office. *G8 Open Data Charter and Technical Annex*. Policy Paper. [Online] Available from: <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex> [Accessed 26 February 2015].
 65. Noveck B. The Single Point of Failure. In: Lathrop D, Ruma L. (eds) *Open Government*. Sebastopol, CA; O'Reilly: 2010.
 66. Africa Development Bank Group. *AfDB's Africa Information Highway: The continent's first "one-stop centre" for development data*. [Online] Available from: <http://www.afdb.org/en/news-and-events/article/afdb-africa-information-highway-the-continent-first-one-stop-centre-for-development-data-12841/> [Accessed 27 February 2015].
 67. OECD. *Open Government in Latin America*. OECD Public Governance Reviews. [Online] OECD; 2014. Available from: http://www.oecd-ilibrary.org/governance/open-government-policies-in-latin-america_9789264223639-en [Accessed 27 February 2015].

-
68. World Wide Web Foundation. *Project: Open Government Data Feasibility Studies*. [Online] Available from: <http://webfoundation.org/projects/open-government-data-feasibility-studies/> [Accessed 27 February 2015].
 69. World Wide Web Foundation. *Open Data Barometer Global Report*. 2nd edition. [Online] Available from: <http://www.opendatabarometer.org/> [Accessed 27 February 2015].
 70. Curtis S. Information Commissioner: 'Open data is no substitute for freedom of information'. *The Telegraph*. [Online] Available from: <http://www.telegraph.co.uk/technology/news/10412374/Information-Commissioner-Open-data-is-no-substitute-for-freedom-of-information.html> [Accessed 27 February 2015].
 71. Ballard M. Poor data quality hindering government open data programme. *ComputerWeekly*. [Online] Available from: <http://www.computerweekly.com/news/2240227682/Poor-data-quality-hindering-government-open-data-transparency-programme> [Accessed 27 February 2015].
 72. Open Data Engagement. *Five Stars*. [Online] Available from: <http://www.opendataimpacts.net/engagement/> [Accessed 27 February 2015].
 73. Gangadharan S, Eubanks V, Barocas S (eds.) *Data and Discrimination: Collected Essays*. [Online] New America Open Technology Institute; 2014. Available from: <http://www.newamerica.org/downloads/OTI-Data-an-Discrimination-FINAL-small.pdf> [Accessed 27 February 2015].
 74. Eaves D. Lies, Damned Lies, and Open Data. *Slate*. [Online] Available from: http://www.slate.com/articles/technology/future_tense/2012/09/open_data_movement_how_to_keep_information_from_being_politicized_.html [Accessed 27 February 2015].
 75. Lichtenstein J. Why Open Data Alone Is Not Enough. *Wired*. [Online] Available from: http://www.wired.com/2011/06/st_essay_datafireworks/ [Accessed 27 February 2015].
 76. Arun S, Heeks R, Morgan S. *ICT Initiatives, Women and Work in Developing Countries: Reinforcing or Changing Gender Inequalities in South India?* Development Informatics Working Paper Series. [Online] Institute for Development Policy and Management; 2014: 20. Available from: http://www.seed.manchester.ac.uk/medialibrary/IDPM/working_papers/di/di_wp20.pdf [Accessed 27 February 2015].
 77. Heller N. The Case Against Open Data in Yemen. *Global Integrity Blog*. [Online] Available from: <https://www.globalintegrity.org/posts/case-against-open-data-in-yemen/> [Accessed 27 February 2015].
 78. Deffor S. *Why privacy considerations matter in the current open data environment*. [Online] Open Knowledge Personal Data and Privacy Working Group; 2014. Available from: <http://personal-data.okfn.org/2014/04/29/why-privacy-considerations-matter-in-the-current-open-data-environment/> [Accessed 27 February 2015].
 79. Blank G, Rasmussen K. The Data Documentation Initiative: The Value and Significance of a Worldwide Standard. *Social Science Computer Review*. [Online] 2004; 22 (3): 307–318. Available from: http://www.researchgate.net/profile/Grant_Blank/publication/234092249_The_Data_Documentation_Initiative_The_Value_and_Significance_of_a_World-Wide_Standard/links/09e4150fecec3b7b5a000000.pdf [Accessed 27 February 2015].
 80. Data Documentation Initiative. *Projects Using the DDI*. [Online] Available from: <http://www.ddialliance.org/ddi-at-work/projects> [Accessed 27 February 2015].
 81. Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, Laaksonen L, et al. Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal*. [Online] 2004; 3: 135–152. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.9016&rep=rep1&type=pdf> [Accessed 27 February 2015].

-
82. Wellcome Trust. *Enhancing Discoverability of Public Health and Epidemiology Research Data: Summary*. [Online] Wellcome Trust; 2014. Available from: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp056925.pdf [Accessed 27 February 2015].
 83. Pollock R, Gray J. *Guide to open data licensing*. Version 1.0. [Online] Open Definition. Available from: <http://opendefinition.org/guide/data> [Accessed 27 February 2015].
 84. Davies T. *Open Data in Developing Countries: Emerging Insights from Phase I*. [Online] Open Data in Developing Countries Project; 2014. Available from: <http://www.opendataresearch.org/sites/default/files/publications/Phase%201%20-%20Synthesis%20-%20Full%20Report-print.pdf> [Accessed 27 February 2015].
 85. Hrynaszkiewicz I, Cockerill M. Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. *BMC Research Notes*. [Online] 2012; 5: 494. Available from: <http://www.biomedcentral.com/1756-0500/5/494> [Accessed 27 February 2015].
 86. Open Knowledge Foundation. *Open Data Handbook Documentation*. Release 1.0.0. [Online] Open Knowledge Foundation; 2012. Available from: <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf> [Accessed 27 February 2015].
 87. Open Data Institute. *Publisher's Guide to Open Data Licensing*. [Online] Available from: <http://theodi.org/guides/publishers-guide-open-data-licensing> [Accessed 27 February 2015].
 88. Pinsent Masons. *Introduction to overseas transfers of personal data*. [Online] Available from: <http://www.out-law.com/page-8170> [Accessed 16 March 2015].
 89. de Montjoye Y, Kendall J, Kerry C. *Enabling Humanitarian Use of Mobile Phone Data*. Series: Issues in Technology Innovation. [Online] The Brookings Institution; 2014. Available from: <http://www.brookings.edu/research/papers/2014/11/12-enabling-humanitarian-use-mobile-phone-data> [Accessed 27 February 2015].
 90. Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*. [Online] 2010; 57: 1701. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 [Accessed 27 February 2015].
 91. Narayanan A. No silver bullet: De-identification still doesn't work. *Freedom to Tinker*. [Online] Princeton Center for Information Technology Policy; 2014. Available from: <https://freedom-to-tinker.com/blog/randomwalker/no-silver-bullet-de-identification-still-doesnt-work/> [Accessed 27 February 2015].
 92. Lowrance W. *Access to Collections for Data and Materials for Health Research: A Report to the Medical Research Council and the Wellcome Trust*. [Online] Wellcome Trust; 2006. Available from: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_grants/documents/web_document/wtx030842.pdf [Accessed 27 February 2015].
 93. O'Dowd A. GP pilots are "last chance saloon" for care.data scheme. *British Medical Journal*. [Online] 2015; 350. Available from: <http://www.bmj.com/content/350/bmj.h419> [Accessed 27 February 2015].
 94. NHS England. *Privacy Impact Assessment: care.data*. [Online] NHS England; 2014. Available from: <http://www.england.nhs.uk/wp-content/uploads/2014/01/pia-care-data.pdf> [Accessed 27 February 2015].
 95. The Patients Association. *Care.data Inquiry*. [Online] The Patients Association; 2014. Available from: <http://www.patients-association.com/wp-content/uploads/2014/06/APPG-Report-on-Care-data.pdf> [Accessed 27 February 2015].

96. Independent Information Governance Oversight Panel. *The Independent Information Governance Oversight Panel's report to the care.data Programme Board on the care.data Pathfinder stage*. [Online] Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389219/IIGOP_care.data.pdf [Accessed 27 February 2015].
97. Sample I. 'Public should be consulted on NHS medical data sharing scheme'. *The Guardian*. [Online] Available from: <http://www.theguardian.com/society/2015/feb/03/medical-data-sharing-scheme-nuffield> [Accessed 27 February 2015].
98. Partridge N. *Review of data releases by the NHS Information Centre*. [Online] Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/367788/Sir_Nick_Partridge_s_summary_of_the_review.pdf [Accessed 27 February 2015].
99. Health and Social Care Information Centre. *HSCIC learns lessons of the past with immediate programme for change*. [Online] Available from: <http://www.hscic.gov.uk/article/4780/HSCIC-learns-lessons-of-the-past-with-immediate-programme-for-change> [Accessed 27 February 2015].
100. Ramesh R. NHS disregards patient requests to opt out of sharing medical records. *The Guardian*. [Online] Available from: <http://www.theguardian.com/society/2015/jan/22/nhs-disregards-patients-requests-sharing-medical-records> [Accessed 27 February 2015].

About the Authors

Matthew Brack is seconded to the Centre on Global Health Security at Chatham House as project manager for the Strengthening Data Sharing for Public Health project. Prior to this secondment, he was working as a project manager and policy officer at the Wellcome Trust, directing the development and implementation of the Trust's revised data and software sharing policy, and delivering digital content for the Trust's Wellcome Collection online resource through large-scale digitization projects. He is certified with the Project Management Institute and has an MA in digital asset management from King's College London.

Dr Tito Castillo is an independent consultant and the current managing director of Xperimint Ltd, a specialist data management company for biomedical data. Previously he established the epiLab-SS secure research data service at the University College London Institute of Child Health, and led the adoption of the Data Documentation Initiative metadata standard for epidemiology and public health data sets for the Secure Epidemiology Research Platform (SERPent) Project. As director of the National Institute for Health Research BioResource at the University of Cambridge, he led research commissioned on behalf of the Wellcome Trust Public Health Research Data Forum on enhancing discoverability of public health and epidemiology research data.

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.

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

© The Royal Institute of International Affairs, 2015

Cover image © David Paul Morris/Bloomberg via Getty Images

ISBN 978 1 78413 049 7

All Chatham House publications are printed on recycled paper.

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700 F +44 (0)20 7957 5710
contact@chathamhouse.org www.chathamhouse.org

Charity Registration Number: 208223