

THE OPEN DATA ERA IN HEALTH AND SOCIAL CARE



A blueprint for the National Health Service (NHS England) to develop a research and learning programme for the open data era in health and social care

Stefaan Verhulst
Beth Simone Noveck
Robyn Caplan
Kristy Brown
Claudia Paz

Table of Contents

FOREWORD	4
EXECUTIVE SUMMARY	6
INTRODUCTION	9
PART I: THE OPEN DATA ERA	11
I.1. Health Data and the Open Data Revolution	12
I.2. The NHS and Open Data: Where We Stand	14
PART II: POTENTIAL AND LIMITATIONS OF OPEN DATA	18
II.1. Value propositions for using Open Health Data	20
II.1.1. Accountability	22
II.1.2. Choice	24
II.1.3. Efficiency	27
II.1.4. Outcomes	29
II.1.5. Customer Service and Patient satisfaction	31
II.1.6. Innovation and Economic Growth	33
II.2. Potential Challenges and Barriers of Open Data	36
II.2.1. Cultural and Institutional Barriers	36
II.2.2. Privacy	38
II.2.3. Standards and Interoperability	40
II.2.4. Good Analysis	41
PART III: TOWARDS A CONCEPTUAL FRAMEWORK FOR OPEN DATA IN HEALTHCARE	42
III.1. What is a conceptual framework and why do we need one?	43
III.2. Inputs: What open data set was used?	46
III.3. Users: Who is using the data?	48
III.4. Activities: How is data used?	49
III.5. Outputs: What was the value produced by the open data initiative (activity)?	49
III.6. Indicators: To what end?	50
III.7. Methodologies to Measure Impact: According to what measure?	50
III.8. Impact: To add what value to the NHS and the UK as a whole?	54
PART IV: BUILDING AN OPEN DATA LEARNING ENVIRONMENT	57
PART V: CONCLUSION	63
APPENDICES	66
Appendix I: The GovLab Open Data Index	66
Appendix II: Open Data Definitions	69
Appendix III: Open Data and Big Data	72
Appendix IV: Interviews	75
Appendix V: Open Data Sites in the NHS	76
Appendix VI: Activities	78
Appendix VII: Outputs	79
Appendix VIII: Indicators	80

THE OPEN DATA ERA IN HEALTH AND SOCIAL CARE

A blueprint for the National Health Service (NHS England) to develop a research and learning programme for the open data era in health and social care



Stefaan Verhulst
Beth Simone Noveck
Robyn Caplan
Kristy Brown
Claudia Paz

The GOVLAB (NYU)
May 2014

Foreword

We are in the midst of a data revolution. More data is useful, but opening data – in other words making it accessible to a wider array of appropriate audiences can yield even more important results. Evidence suggests that giving patients access to their own healthcare records can increase healthy behavior and improve decision-making. Giving providers access to comparative performance indicators across hospitals and physicians increases cost-efficiency in the healthcare system. Giving researchers access to clinical data improves medical outcomes. Giving the public-at-large access to nationwide prescription data and hospital performance increases choice, empowerment and accountability. These are just some examples of the value behind making health data more open and accessible, into what we may collectively call shared data. Through making data more accessible, while also ensuring privacy and confidentiality, NHS England has committed itself to spur a revolution in health care, through increasing patient power, saving lives, and greatly improving quality of treatment.

Three key objectives lie at the core of all our work:

- ▶ *Transparency*: The safe sharing of data and information between clinicians, patients, and the public.
- ▶ *Participation*: Supporting patients and citizens to take more control of their health and care and fully engage in the design of local services.
- ▶ *Interoperability*: The development of seamless digital records across all care settings, based on open standards.

Since April 2013, NHS England has undertaken a number of initiatives to achieve all three objectives. These include: The scheduled launch of care.data in 2014, an initiative designed to address gaps in the information currently held by the NHS and to promote seamless data flows across NHS departments; the publication of comparative provider outcomes; and the introduction of the “Friends and Family Test,” a measure of patient satisfaction based on the number of patients who would recommend hospital inpatient and A&E services to friends and family, and associated free text feedback.

All these initiatives have gone a substantial way towards helping us understand the potential impact of making data accessible. Through these investigations, we have gained a better sense of what works, and what does not. To take our efforts to the next level, we are adopting a more systematic approach to the way we use and evaluate the effects of open data. To do so, we need methods to measure the impact of open data and to analyze the specific circumstances under which it is most (or least) effective. In other words, we require a roadmap that will guide us in our efforts to build on existing successes and applications.

This report, prepared by the GovLab, a global action research network, goes a considerable way towards establishing such a roadmap. It lays the foundations for a substantive research agenda that can help health and care services advance their objectives and, at the same time, deepen interaction with open data, as well as enhance patients’ and providers’ experiences. I am particularly pleased with the report’s emphasis on establishing *metrics* of success and *methodologies* to collect the data required to support these metrics. Metrics are critical in evaluating success and failure. They are essential to helping us understand what we do right and where our work needs refinement or improvement.

The report's broader emphasis on establishing an organisation-wide culture of learning (what the authors call an Open Data Learning Environment) is also invaluable. Ultimately, a successful application of open data can come only from a transformation of the way we use information across departments, hospitals, providers, trusts, and other agencies and institutions. Data transparency requires coordination and commonality of purpose, and it obliges us to cut through existing silos. The proposed Open Data Learning Environment is an important step towards achieving this goal of systematic and organisation-wide transformation.

It is my belief that making data more accessible, in an appropriate manner, will ultimately revolutionize the delivery of health and social care in England and around the world. I see this transformation occurring in years, not decades. This report represents yet another milestone in our efforts to ensure that NHS England will be at forefront of this global transformation.

A handwritten signature in black ink, consisting of a large, stylized initial 'T' followed by 'Kelsey' and a long horizontal stroke extending to the right.

Tim Kelsey
National Director for Patients and Information

EXECUTIVE SUMMARY

The United Kingdom has been a leader in the open data movement – a new movement by governments around the world to open up the vast repositories of data they hold across agencies and departments, and to collect new kinds of data for public use. Open data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. It is changing the way governments, nonprofits, and the private sector use data to understand public issues and solve problems in areas as diverse as financial regulation, energy, education, and more.

Open data holds particular potential in the health sector. By releasing health data to patients and, when appropriate, on an anonymized basis to researchers and the public, governments and healthcare organizations are betting on the power of greater openness of data to improve the quality of care, lower healthcare costs, and facilitate patient choice. The NHS has made and continues to make significant investments in opening data. Over the past several years, it has launched a series of initiatives that have already had a positive impact on patient education, healthcare choice, healthcare costs, and patient outcomes. Now the NHS is planning a broader, more ambitious programme that has the potential to serve as a worldwide model for the opening of data in healthcare. The purpose of this report is to help design this programme, establishing priorities and ways of measuring impact to guide this and future efforts at data transparency.

This report examines the current literature, drawing on case studies and published research to highlight the following value propositions for using more open data in healthcare:

- ▶ **Accountability:** The use of data to hold healthcare organizations and providers accountable for treatment outcomes.
- ▶ **Choice:** Providing open data to help patients make informed choices from among the healthcare options available to them.
- ▶ **Efficiency:** Improving the efficiency and cost-effectiveness of healthcare delivery.
- ▶ **Outcomes:** Improving treatment outcomes by using open data to make the results of different treatments, healthcare organizations, and providers' work more transparent.
- ▶ **Patient satisfaction and customer service:** Using open data to educate patients and their families and make healthcare institutions more responsive.
- ▶ **Economic growth and innovation:** Using open data to fuel new healthcare companies and initiatives, and to spur innovation in the broader economy.

Despite widespread recognition of these potential value propositions, there is little understanding of whether and how open data initiatives definitively lead to these outcomes. One of the central goals of this paper is to establish a conceptual framework, or logic model, that can be used by researchers and programme managers to design their open data initiatives and then measure their impact. The paper outlines the components of this framework, linking specific inputs and activities with indicators and their potential impact, to be used in measurement. The conceptual framework also discusses a variety of methodologies that can be used to facilitate measurement and impact analysis.

Overall, this draft blueprint from the GovLab describes the need for careful assessment of healthcare open data initiatives and outlines an approach that can help in such an assessment. The approach helps both in assessing the opportunities of open data, and also in considering potential risks and challenges. Though there has been significant discussion about the concerns

and risks emerging from open data, there has been little research into its potential economic and social benefits. In today's budgetary climate, we need to enhance our ability to marshal evidence on the efficacy of innovative and potentially important new programmes, while at the same time taking into account the societal and other risks (e.g., to privacy) that may accompany such programmes. By laying out a research agenda to accompany the NHS's open data strategy, our hope is to ensure that public investments in open data are supported by concrete evidence of its value, and that this, in turn, can be used to guide and develop the NHS's ambitious plans to shift an entire nation's bureaucracy to more evidence-based decision-making.

In conclusion, this report recommends that the NHS create an **Open Data Learning Environment (ODLE)** that combines the use of innovative technology platforms, robust consultation with a wide array of stakeholders, close partnerships with the research community, and fellowships to bring innovators into the NHS. We make ten specific recommendations for the creation of this ODLE, which would facilitate ongoing assessment of progress against goals, rather than simply offering a one-time snapshot of results. Taken together, these recommendations would help the NHS improve and refine how it opens data, and to whom, in a nimble, real-time fashion.

To create the proposed ODLE, we recommend that the NHS:

- ▶ **Build an Open Data Learning Capacity and Culture** within the NHS. This can include a number of steps to institutionalize action and learning about open data, such as creating a corps of volunteers with technical knowledge, setting up an Open Health Data Academy, soliciting calls for questions from the academic community, and posing challenges and offering prizes for the best uses of open health data.
- ▶ **Engage the public in defining metrics** of success with regard to open data. In addition to ensuring that direct stakeholders (e.g., patients and providers) are included in decisions about open data, the NHS may also want to consider allowing citizens to participate – for example, through a citizens' open health data panel.
- ▶ **Develop a common assessment framework** with partners in other sectors and countries to measure the impact of open health data, using meetings, online communication, and additional tools.
- ▶ **Stay flexible and adaptive in measuring impact.** As the use of open data evolves and becomes more dispersed, the NHS will need to find a balance between a centralized measurement function (which sets measurement standards, aggregates data across programmes, and monitors data quality) and a more diffused structure that empowers users at different levels of the organization.
- ▶ **Share what is learned (including failures) with everyone.** By sharing what is learned (including mistakes and failures) on a regular basis, the NHS can enable distributed oversight and peer review mechanisms to ensure the quality of its research and enhance the way open data is used.
- ▶ **Build a research network and expert network** to build capacity quickly and broker debate around methodology and findings. Outside expertise, gathered through meetings, online input, or an advisory board, will be critical in determining the best ways to assess impact.
- ▶ **Develop an open health data ecology map**, possibly using crowdsourcing, with a dictionary of all open health data sets used along with the variety of uses and users.
- ▶ **Publish, integrate and fine-tune the open data conceptual framework** for feedback and improvement, using expert advisors and input from stakeholders.
- ▶ **Engage citizens and practitioners in shaping the open health data programme.** The NHS should engage experts in the field on a regular basis to determine which datasets have the highest priority for them; what new datasets should be released as open data; and which open data collections are particularly easy or difficult to use.

THE OPEN DATA ERA IN HEALTH AND SOCIAL CARE

A blueprint for the National Health Service (NHS England) to develop a research and learning programme for the open data era in health and social care



Stefaan Verhulst
Beth Simone Noveck
Robyn Caplan
Kristy Brown
Claudia Paz¹

The GOVLAB (NYU)

May 2014

¹ With important input and support from Akash Kapur, Joel Gurin, Jon Billings and Wagner Capstone students: Shammara Wright, Brenden Desmond, Christian Laurence, Christina Rogawski, and Jillian Dimaano

Introduction

- ▶ In a recent study, a company called Mastodon C, based at the Open Data Institute in Shoreditch and working in conjunction with Ben Goldacre and Open Healthcare, analyzed massive amounts of data on prescription patterns released by the National Health Service. The analysis found significant geographic variations across England, with expensive prescription drugs being prescribed in some areas when generics would work just as well. If the findings of the study were applied to change prescription patterns, the savings could amount to more than 1 billion GBP per year.²
- ▶ Nine years ago, the Society for Cardiothoracic Surgeons began publishing the results of surgeries done by individual physicians across the UK. Surgical outcomes improved rapidly, and a report by the Society stated that “we believe that the marked, sustained, incremental improvement in the quality of care the surgical teams have achieved is directly associated with the process of recording, reporting and publishing outcomes at the level of the individual clinician.”³
- ▶ The National Health Service has begun doing a “friends and family” study, using well-established survey methods to ask patients and their families whether they would recommend a particular hospital or care center. By publishing these “friends and family” ratings, the NHS hopes to give consumers more information and choice, and in the process improve the quality of care they receive.⁴

These examples, and many others we discuss in this report, are beginning to show the power of “open data” in healthcare. In recent years, a growing number of governments --from cities to states to regions -- have committed to opening up the vast repositories of data they hold across agencies and departments, and in many cases to collecting new kinds of data for public use. The United Kingdom has been at the forefront of this open data revolution. According to two third-party measures, the Open Data Index⁵ and the Open Data Barometer,⁶ the UK is ahead of other developed countries (including the US, Australia and France) when it comes to the “reach” of its open data (measured by visits to and incoming links at the main open data websites).⁷ In the words of one leading UK official, the government has committed to “unleash a tsunami of data.”⁸

Healthcare is one of the sectors with the potential to be most radically transformed by greater data accessibility. Indeed, many of the most important steps taken by the UK government with regard to open data involve open health data. In recent years, the NHS has invested substantial resources (financial and intellectual) into developing a comprehensive open data strategy. Information concerning GP prescribing, mortality rates, waiting times, and a wide variety of other data has been released to the public and researchers. An incipient ecology of innovation, analysis and research has begun to emerge as a result of such initiatives. Overall, the NHS’s efforts with regard to open data have the potential to transform the delivery and experience of healthcare, spurring greater efficiencies, more innovation, and, perhaps most importantly, better health outcomes for patients.

The philosophical (or ethical) argument for open data can be explained by two principles. First, information whose collection is paid

2 James C Moon, Andrew S Flett, Brian B Goodman, Anthony M Grosso, and Anthony S Weirzbicki, “Getting better value from NHS drug budget,” *British Medical Journal* 341(2010).

3 The Society for Cardiothoracic Surgery in Great Britain & Ireland, “Maintaining patients’ trust: modern medical professionalism,” SCTS.org, 2011, http://www.scts.org/_userfiles/resources/634420268996790965_SCTS_Professionalism_FINAL.pdf.

4 NHS Choices, “The NHS friends and family test,” NHS Choices, accessed February 23, 2014, <http://www.nhs.uk/NHSEngland/AboutNHSservices/Pages/nhs-friends-and-family-test.aspx>.

5 “Open Data Index,” Open Knowledge Foundation, accessed February 23, 2014, <https://index.okfn.org/country/>.

6 Tim Davies, “Open Data Barometer: 2013 Global Report,” 2013, <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>.

7 Deloitte, “Open growth: Stimulating demand for open data in the UK,” 2012, accessed February 23, 2014 http://www.deloitte.com/view/en_GB/uk/market-insights/deloitte-analytics/bfb570a79416b310VgnVCM1000003256f70aRCRD.htm.

8 Simon Rogers, “Meet the man who turned David Cameron onto open data,” *The Guardian*, April 30, 2013, accessed February 23, 2014, <http://www.theguardian.com/news/datablog/2013/apr/30/rohan-silva-interview-david-cameron-open-data>.

for by taxpayers should be available to those taxpayers; second, personal information about individuals (e.g., medical data) should likewise be accessible to those individuals. There is also an instrumental rationale for open data, which is that raw datasets lend themselves to computation and algorithmic analysis. What makes open data potentially transformative is the fact that it facilitates third parties to use it in ways that create new analyses, visualizations, and mash ups. This kind of innovation can transform raw data into knowledge. Solving complex challenges requires many people with diverse skills and talents to work together.⁹ When experts of all kinds have access to open data, it becomes a catalyst for creative problem solving and community innovation. “The real value [of open data] comes from interpretation, analysis, linking-up and reflection – in short, from being used,” says one database expert. “Keeping hold of data that you don’t use much is wasting potential. By making data open, you enable others to bring fresh perspectives, insights, and additional resources to your data, and that’s when it can become really valuable.”¹⁰

Today, the potential of open health data is widely recognized. Researchers, journalists and of course healthcare providers celebrate an opportunity to improve efficiency of service delivery, catalyze provider accountability, and help patients make more informed choices about their care. Yet much of this promise is still anecdotal and unproven. There is a dearth of systematic analysis and solid, empirical data. It remains very difficult to measure the value of open data interventions, and by extension to decide which initiatives are worth pursuing and which not.

The central premise of this paper is that as the NHS moves to release data systematically, it needs to put in place a strategy for measuring the value of open data for the various stakeholders involved in the nation’s healthcare system—and, indeed, for citizens in general. In today’s budgetary climate, it is not enough to assess the value of expensive and complicated government programmes after the fact. We need to enhance our ability to marshal an arsenal of evidence in order to protect investments in innovative and potentially important new programmes. By becoming more agile in how we measure innovations in governance like open data, we can make government more efficient, and more effective.

To aid in this goal, this draft whitepaper articulates recommendations for the NHS to follow as it seeks to measure the impact of open data empirically. By laying out a research agenda to accompany the NHS’s open data strategy, our hope is to ensure that public investment in open data is supported by concrete evidence of its value, which, in turn, can be used to guide and evolve the ambitious plan to shift an entire nation’s bureaucracy to more evidence-based decision-making. We are releasing this as a draft in order to encourage discussion and additional insights from interested readers.

The paper is divided into four parts:

Part I explains open data as a driver of innovation. We summarize the open data plans of the NHS, including the data the NHS holds, what it is planning to release and when, and the challenges to implementing a nationwide open data plan.

Part II lays out the arguments in favor of using open data in a healthcare setting (six value propositions), such as improving patient choice and strengthening provider accountability, and outlines the empirical evidence we currently have in support of each.

Part III presents a series of metrics that can help the NHS measure its performance and improve its use of open data. It establishes a conceptual framework to use for continuously evolving and accelerating the ability to measure the impact of open data in healthcare.

Finally **Part IV** concludes with specific principles and recommendations to establish an Open Data Learning Environment (ODLE)--the practices and platforms by which to operationalize agile assessment and enable programme evolution.

9 Beth Simone Noveck, “Open Data: The Democratic Imperative,” *Crooked Timber*, July 7, 2012, accessed February 21, 2014, <http://crookedtimber.org/2012/07/05/open-data-the-democratic-imperative/>.

10 Matt Parker, “Lamplight Database Systems, in Kate Hodge, *Best Bits: Open Data for Charities*,” *The Guardian*, May 25, 2012, accessed February 21, 2014, <http://www.guardian.co.uk/voluntary-sector-network/2012/may/25/open-data-charities-advice>.



Part I

The Open Data Era

1.1 Health Data and the Open Data Revolution

The last decade or so has seen a data explosion around the world. In late 2012, it was estimated that there existed some 2.8 zettabytes (2.8 trillion GB) of data; this was expected to grow to 40ZB by 2020.¹¹ Until recently, much of this data was closed to public access--hidden behind firewalls and passwords, protected by copyright or other legal barriers. In addition, some of the data was kept from the public due to legitimate concerns over privacy or national security.

Recent years have witnessed something of a sea change in the way data is treated, particularly by governments and (to a lesser extent) private companies. Across the world, there has been a gradual—and sometimes not so gradual--transition underway from what McKinsey calls a culture of “protect” to one of “share, with protections.”¹² With this transition comes huge opportunity. It is estimated that open data could help unlock \$3-\$5 trillion globally.¹³ In the EU27 economy, the value of open data was calculated to be €32 billion in 2010, with a 7% annual growth rate (see Appendix I for additional statistics).¹⁴

It is this move towards greater data transparency that marks what some observers have called an open data revolution. But just what is open data? Open data is often confused with another information science trend that has received significant media attention: Big Data. The two trends are related, but not identical.

There is no single, universally accepted definition of open data. In order to develop a working definition, the GovLab assembled nine current definitions (Appendix II) and placed them within a framework for comparative analysis (Appendix II: Table 1). The resulting definition, a synthesis that guides our discussion in this paper, is as follows:

Open data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. Open data is released in ways that protects private, personal or proprietary information. It is structured for usability and computability.

Ultimately, the *openness* of open data depends on the development of standards and best practices to guide the way data is presented and organized. The goal in opening data is to make it easier to find the data, present it in usable formats, and streamline terminology when possible. This will ensure that data is not only accessible, but also that it is readable, comprehensible and usable by a wide range of individuals. To achieve these goals, data should be: (1) published in a standard format; (2) published without proprietary conditions; and (3) available online in a downloadable format.¹⁵

Today, some forty countries have developed open data policies.¹⁶ These policies vary in the extent to which they meet these three conditions or match our working definition. Almost all, however, are motivated by efforts to break down existing silos and bureaucratic barriers across a wide range of information categories. Open data policies today apply to a variety of different kinds of information, including information related to land records, welfare schemes, and political operations and fundraising that can be accessed and used, is non-personally identifiable (through anonymization or pseudonymization), and is structured for usability and release, without restrictions of copyright or other mechanisms of control. There is reason

11 John Murdoch-Burn, “Study: less than 1% of the world’s data is analysed, over 80% is unprotected,” December 19, 2012, accessed February 23, 2014, <http://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume>.

12 Basel Kayyali, David Knott, and Steve Van Kuiken, “The big-data revolution in US health care: Accelerating value and innovation,” 2013, accessed February 23, 2014, http://www.mckinsey.com/insights/health_systems_and_services/the_big_data_revolution_in_us_health_care.

13 James Manyika, Michael Chui, Diana Farrell, Steve Van Kuiken, Peter Groves, and Elizabeth Almasi Doshi, “Open data: Unlocking innovation and performance with liquid information,” October 2013, accessed February 23, 2014, http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.

14 Dinand Tinholt, “The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data,” 2013, accessed February 23, 2014, <http://www.capgemini-consulting.com/the-open-data-economy-0>.

15 Beth Simone Noveck, and Daniel L. Groff, “Information for Impact: Liberating Nonprofit Sector Data,” 2013, http://www.aspeninstitute.org/sites/default/files/content/docs/psi/psi_Information-for-Impact.pdf.

16 Manyika et al., “Open Data: Unlocking Innovation.”

to believe, however, that data transparency might prove to be a particularly revolutionary tool when it is applied to health information. As the six value propositions we discuss below suggest, health data might in some respects prove to be a unique category of information. Around the world, providers, researchers, and policy makers are articulating a new vision that some have begun to call “Health 2.0” or “Medicine 2.0”—a vision of coordinated, patient-centered healthcare that leverages data and technology to engage patients, integrate providers, and improve clinical outcomes. At the same time, health data poses certain challenges—particularly concerning privacy and identity—that require treatment with particular sensitivity and, at times, even restrictions. We also discuss these challenges below.

In truth, health data—like all data—exists on a continuum from open to closed. In this paper, we suggest three categories on this continuum of “accessibility:”

1 - Open Health Data: Currently open data is published under the Open Government License, and is available through the *Data.gov.uk* site, and through the Health & Social Care Information Centre at *HSCIC.gov.uk*, Public Health England,¹⁷ and other sites (see Appendix V for a complete list of data types and locations). Open data is made publicly available, in a user-friendly, interactive, or downloadable format.

2 - Restricted Health Data: Data that potentially carries personal or proprietary information, which is anonymized through the removal of personally identifying information. Researchers or charities can then request to access this data, though they may have to fulfill additional criteria determined by the NHS.

3 - Published Health Data: Data that has been published in a particular format by the NHS or other governmental entities. This data is generally prepared for presentation in an interactive, .doc, or PDF format and the original data files are not included in the release. An example of published data would be NHS patient satisfaction surveys.

These categories have been developed through an analysis of health data as it is currently released through the NHS. What both the NHS and the HSCIC currently consider under the umbrella of “open health data” is limited, but eventually can come to include a greater variety and quantity of datasets throughout the NHS open data initiative.

¹⁷ “Public Health England,” *Gov.UK*, accessed April 23, 2014, <https://www.gov.uk/government/organisations/public-health-england>.

1.2 The NHS and Open Data: Where We Stand

The United Kingdom has been at the forefront of global efforts to open up the data in governance, and especially in healthcare. The government's 2012 *White Paper on Open Data* begins with these ambitious words: "Data is the 21st century's new raw material. Its value is in holding governments to account; in driving choice and improvements in public services; and in inspiring innovation and enterprise that spurs social and economic growth."¹⁸ By the end of 2012, it was estimated that there were some 37,500 datasets publicly available on the UK government's main data sites, including *Data.gov.uk*, *Ons.gov.uk*, and a variety of other local data stores including *Data.London.gov.uk*. In total, there had been some 2.7 million downloads from these sites.¹⁹

From the beginning, the health sector and the NHS have been identified as priority areas for data initiatives. In fact, systematic collection of data in the NHS can be traced to the 1980s, long before current interest in open data or big data. It wasn't until the 1990s, however, that a more robust data collection framework was established. Even then, the decision to upgrade and refine the way data was collected was driven not so much by a desire for increased transparency as by efforts to achieve greater economic efficiencies and introduce "internal markets" within the NHS. Such efforts were part of the government's New Public Management (NPM) policies, and were designed to foster competition inside the NHS, separating out "providers" from "purchasers." These efforts grew during the early 2000s, when NHS organizations were required to collect many more kinds of data, including information on waiting times, performance evaluation of general practitioners, names of patients with diabetes and strokes, and adverse events, which were reported to the National Patient Safety Agency.²⁰

In the last decade or so, the NHS has expanded its commitment to collecting and publishing open health data. **The table "NHS Open Data Milestones"** includes some key initiatives undertaken by the NHS and its partners. Two broad points stand out in particular. The first is a renewed commitment to data transparency following the election of the coalition government in 2010, a commitment that marks continuity with the previous government's initiatives and thus a cross-party consensus on the value of open data. The current government's commitment has been evident in a range of research and publications, notably the *White Paper on Open Data* and *The Power of Information* (both published in 2012), and the passage of the Health and Social Care Act of 2012, which placed open data and data transparency at the core of the government's healthcare strategy.

In addition, a hallmark of the UK's data transparency policies (and arguably one of the key reasons for their success) has been third-party innovation. Current government and NHS policies call for de-identified data to be made available to researchers and industry in order to encourage such innovation.²¹ As a result, many new data-driven businesses have been developed with the support and collaboration of government agencies. For example, the Open Data Institute (a non-profit company seeded with £10 million of public money) has incubated several new companies. Likewise, Dr Foster Intelligence, which provides health-related information and research to the public, such as hospital and consultant guides, is now partially owned by the Department of Health.

18 "Open Data White Paper: Unleashing the Potential," 2012, http://data.gov.uk/sites/default/files/Open_data_White_Paper_10.pdf.

19 Deloitte, "Open growth: Stimulating demand for open data in the UK," 2012, accessed February 23, 2014, http://www.deloitte.com/view/en_GB/uk/market-insights/deloitte-analytics/bfb570a79416b310VgnVCM1000003256f70aRCRD.htm.

20 Justin Keen, Radu Calinescu, Richard Paige and John Rooksby, "Big data + politics = open data: The case of health care data in England," *Policy & Internet* 5(2) (2013), doi: 10.1002/1944-2866.POI330.

21 Deloitte, "Open growth."

NHS OPEN DATA MILESTONES

- ▶ In 2005, the NHS Information Centre was established to be the authoritative repository of NHS health information, making a wide range of data available to providers, researchers and patients.²²
- ▶ In 2007, NHS Choices was launched, empowering patients to make decisions about their healthcare based on comparative data about hospitals and GPs while offering information about conditions, treatments and healthy living. Users were also able to share their experiences of using NHS services. Currently, the site attracts 35 million monthly visits and studies have indicated its important (and largely positive) role in changing patient behavior.²³
- ▶ After the 2010 elections, the newly elected coalition government also gave priority to data transparency--including open health data--in a series of policy statements and choices, notably the White Paper on Open Data of 2012. In 2011, the Prime Minister David Cameron made his commitment to open data clear in a speech he delivered to a pharmaceutical and biotechnology conference:

“Now there’s something else that we’re doing... and that is opening up the vast amounts of data generated in our health service. From this month huge amounts of new data are going to be released online. This is the real world evidence that scientists have been crying out for and we’re determined to deliver it. We’ve seen how powerful the release of data can be... It is simply a waste to have a health system like the NHS and not to do this kind of thing. So we don’t want to stop there... Because believe me, we’re determined to convert these words and policy papers into action.”²⁴

- ▶ In 2012, the new Health and Social Care Act placed open data and data transparency at the core of its healthcare strategy. Among other things, it required the Health and Social Care Information Centre to release virtually all of the data it possessed (in anonymized form). Eighty-three new datasets were subsequently released as part of this requirement.
- ▶ Also in 2012, the NHS published its new IT strategy, a ten-year plan called The Power of Information. The document promised, among other things, that “all health and care records held by hospitals and other service providers will be made securely available to patients;” and made a commitment that “by 2015, anyone in England will be able to access their GP health record online as well as book appointments with their GP or request repeat prescriptions online.”²⁵
- ▶ In September 2013, the NHS launched the CareConnect Pilot, a phone and web service that allows citizens to interact with the NHS, get information, and provide real-time data through feedback and other mechanisms. The NHS has also announced its intention to launch a multi-channel customer service platform, the Health and Social Care Digital Service. Locally, the NHS developed a series of websites geared towards localities or regional healthcare needs (for example addressing dementia in South West England), such as *NHSLocal.nhs.uk* (now closed) and *MyHealthLondon.nhs.uk*.
- ▶ The UK is benefitting from other open data initiatives around the world. In January, 2014, the Secretary of State for Health, NHS England, signed a Memorandum of Understanding with the US Department of Health and Human Services, recognizing the importance of establishing a collaboration to allow both countries to learn from the experiences of the other, and to align their approaches.

22 Recently reborn as the Health and Social Care Information Centre, or HSCIC.

23 Nicolaus Henke, Tim Kelsey, and Helen Whately, “Transparency the most powerful driver of health care improvement?” 2011, accessed February 23, 2014, https://www.mckinsey.com/~media/mckinsey/dotcom/client_service/Healthcare%20Systems%20and%20Services/Health%20International/Issue%2011%20new%20PDFs/HI11_64%20Transparency_noprint.ashx.

24 Prime Minister David Cameron, “Speech on life sciences and opening up the NHS,” 2011, accessed February 23, 2014, <https://www.gov.uk/government/speeches/pm-speech-on-life-sciences-and-opening-up-the-nhs>.

25 “The Power of Information,” 2012, accessed February 23, 2014, <http://www.england.nhs.uk/2012/05/21/the-power-of-information/>.

LOOKING FORWARD

In particular, the policies and data architecture implemented over the past decade have laid the foundations for three main objectives: releasing more data; expanding datasets to measure health outcomes; and making data actionable for healthcare consumers. These priorities, already evident in recent initiatives, are likely to define the NHS' open data efforts going forward.

Releasing more open data: Motivated in part by the success of publishing comparable data for cardiac surgical outcomes, England began in July 2013 publishing surgery outcome data for 10 major specialties.²⁶ In February 2014, Secretary of State for Health, Jeremy Hunt, noted that these efforts would be taken even further with hospitals being asked to publish monthly staffing ratios, on a ward-by-ward basis.²⁷ These initiatives reflect the NHS' ongoing commitment to the release of more open data that can be used by patients, physicians and commissioners to compare the quality of care delivered across the NHS.

Expanding datasets to measure health outcomes: The 2014/15 Outcomes Framework released by the Department of Health reiterates the commitment established in the 2010 White Paper, *Liberating the NHS*, to shift the focus from measuring process targets to health outcomes.²⁸ In addition to re-focusing data collections on health indicators, NHS plans to transform the Hospital Episode Statistics (HES) into Care Episode Statistics (CES) by linking hospital datasets to a wide range of other datasets including primary care data, community health services data, clinical audit data, cancer registry data, social care data, mental health data, and, for those who wish to donate it, genomic data.²⁹ This effort, known as Care Data, will enrich data sources by collecting more granular data on all aspects of care received within hospitals, and will link clinical data collected throughout the NHS.

Making data actionable for healthcare consumers: As part of an effort to move towards patient-centered and customer-focused care, the NHS is redesigning its publication systems to better meet the needs of individual patients. For example, the NHS is revamping its NHS Choices website to add feedback and transactional functionality. In addition, the new Health and Social Care Digital Service will be designed to offer enhanced functionality to the public through several mechanisms, including online appointment booking and prescription refills, secure connections to personal health records, and better social media integration as well as greater use of patient social networks.^{30,31}

Together, these recent initiatives act in concert to help the NHS achieve greater transparency and improve health outcomes. They define the parameters and contours of the NHS's open data thinking, and help give direction to a data transparency strategy to guide the coming years.

26 Jeremy Hunt, "Why the NHS is crossing the Rubicon," *The Telegraph*, February 4, 2014, <http://www.telegraph.co.uk/health/nhs/10617243/Jeremy-Hunt-Why-the-NHS-is-crossing-the-Rubicon.html>.

27 *Ibid* 25.

28 NHS, "The NHS Outcomes Framework 2014/15," *Gov.Uk*, November 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/256456/NHS_outcomes.pdf.

29 Geraint Lewis, "Introducing the care.data programme," July 2013.

30 John Coulthard, e-mail message to *The GovLab*, July 18, 2013.

31 Informatics Services Commissioning Group, "Health and Social Care Digital Service," NHS, accessed February 23, 2014, <http://www.england.nhs.uk/iscg/wp-content/uploads/sites/4/2013/10/ISCG-007-006.pdf>.

The US Context

**A MEMORANDUM OF UNDERSTANDING BETWEEN
THE UNITED STATES' DEPARTMENT OF HEALTH
AND HUMAN SERVICES (HHS) AND NHS ENGLAND**

*“Openness will strengthen our democracy
and promote efficiency and effectiveness in Government.”*

- President Barack Obama

In January, 2014, the United States' Department of Health and Human Services (HHS) and NHS England announced they had signed a bi-lateral agreement for the use and sharing of health IT information, tools, and services. This Memorandum of Understanding (MoU) is just part of a commitment that the United States and President Obama has made towards increasing transparency and openness in governance through declaring data maintained by the Federal government to be a national asset, calling for its rapid disclosure. In this MoU, the parties express a commitment to collaborate on open data initiatives, the adoption of digital records systems, and on priming the health IT market for additional collaborative practices.

As a result of the MoU, NHS England will be able to benefit from not only its own open health data, but also that of the HHS. The Department of Health and Human Services was among the first agencies to begin sharing public health. Since then, the Health Data Initiative has posted nearly 1,000 datasets from the Centers for Disease Control and Prevention (CDC), Centers for Medicare & Medicaid Services (CMS), the National Institutes of Health (NIH), the Administration for Children and Families, and other agencies on a dedicated health data website called HealthData.gov. These datasets include, among others, data that has been provided from publicly sponsored organizations such as Medicare and Medicaid, and other organizational data at the hospital, regional and state levels.

To continue to be effective, the Department of Health and Human Services (HHS) has acknowledged that data must not only be open, but should be accessible to as many individuals and organizations as possible. To achieve this objective, the HHS is revising their HealthData.gov platform to make it more efficient, flexible, and user-friendly.

As part of the MoU, the HHS, NHS, and HSCIC are working with international organizations on initiatives regarding public and private interoperability, and international and inter-agency standards for health data. The MoU thus recognizes that ensuring interoperability is essential for promoting health information exchange, and that interoperability must take into account not only information technology, but also the social, political and organizational factors that may affect information exchange.



Part II

Potential and Limitations of Open Data

A number of governments, including in the UK, have argued that open data is the key to more transparent and accountable governance. Open data also has a range of other benefits.³² Through a review of existing evidence and literature, we have identified six domains that can be positively affected through open data initiatives in the health sector.³³ These six domains, or value propositions, include: accountability, choice, efficiency, outcomes, customer service, and innovation and economic growth.

Figure 1: Six Value Propositions

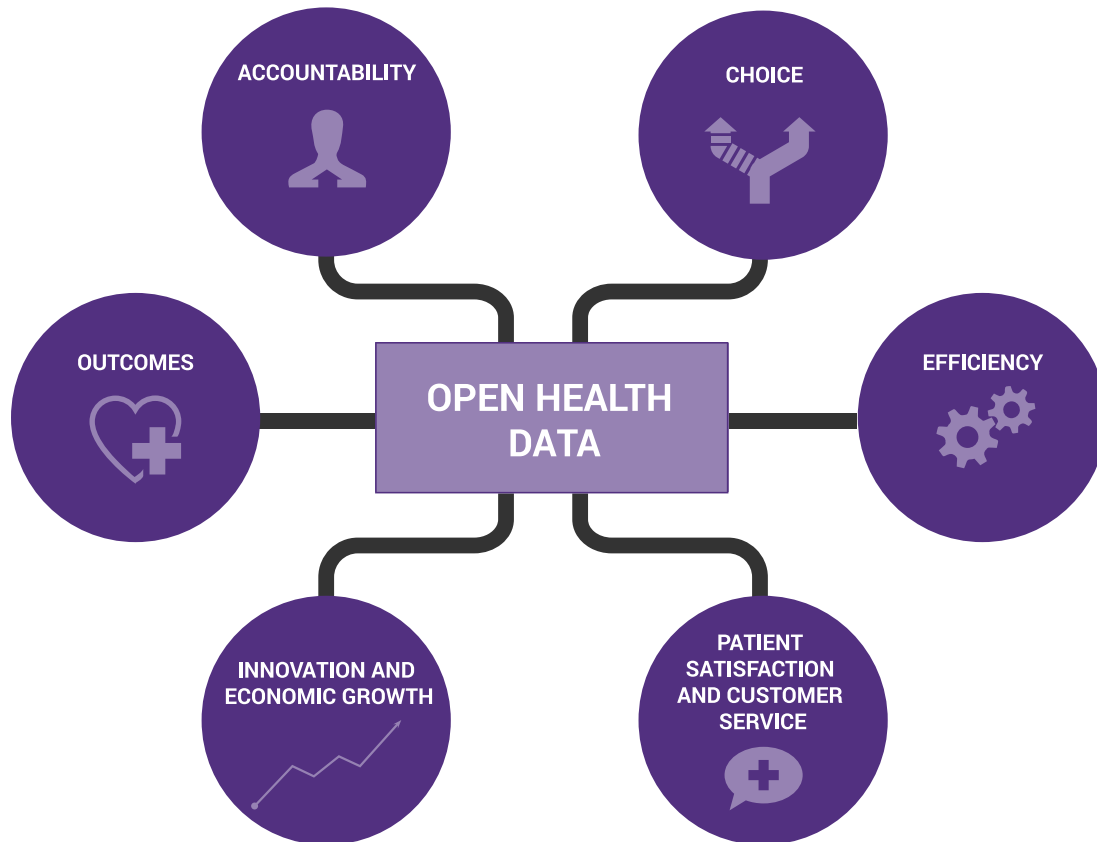


Table II.1 summarizes these value propositions, and the following discussion explains them more fully. Taken together, they offer a powerful argument in favor of greater transparency—for healthcare in particular, but also more generally. At the same time, we have identified a number of challenges that, if left unaddressed, may limit the potential of open data. We discuss these challenges and barriers below.

³² Henke, Kelsey and Whately, "Transparency," 2011.

³³ The research team reviewed in a systematic manner a variety of sources to establish the current framework and to develop a scan of the available evidence till date with regard to open data. The universe of sources included publications released by the UK government on open data, publications released by related and non-governmental agencies, publications released by foreign governments such as the US and international bodies, peer-reviewed research on healthcare and bioethics, and academic research on open data more generally. Through this collection of resources, the research team took a discursive approach, identifying keywords in the relevant literature, using these keywords to further define – or in some cases expand – their areas of study. Through this analysis the research team identified six broad areas of interest in the use of open data more generally, and its potential application in the healthcare industry: accountability and transparency, efficiency, choice and behavioral science, consumer advocacy and customer service, and economic growth and entrepreneurial innovation. These categories were then further refined in concert with the goals for the open data program, expressed by key NHS stakeholders.

II.1 Value Propositions for Using Open Health Data

POTENTIAL IMPACT	ADDITIONAL CATEGORY	DESCRIPTION
Accountability	Political	Healthcare professionals and patients interact as citizens within a community, and are accountable to/affected by governing bodies.
	Economic	Healthcare exists as part of an economic system, in which funding is allocated, and accountability is mediated through transparency in financial data.
	Participatory and Citizen Involvement	Open data can be used to facilitate participatory and citizen involvement.
	Organizational/Clinical	Healthcare professionals are engaged in a set of shared decision making practices both internal to the NHS (with their professional colleagues) and with their patients. Healthcare professionals should be held accountable in both sets of relationships.
Choice	Variety	The availability of comparative information about public services enables patients to make more informed decisions in choosing treatment.
	Quality of information	An increase in the quality, completeness, and timeliness of information provided to patients can enable better decision-making in treatment plans or choice of services.

POTENTIAL IMPACT	ADDITIONAL CATEGORY	DESCRIPTION
Efficiency	Operational	The ratio of valued inputs to valued outputs in the healthcare system, not limited to cash flow, but people (labour) and time and effort.
	Economic	The relationship between increased operational efficiency to the cost savings. Enables assessment of which services are ineffective or inappropriate.
	Communication and Technology	Assessing which technologies facilitate improved communication between healthcare professionals, physicians and patients, and healthcare and the general population.
	Treatment	Assessing efficiency of treatments, diagnostic tests and medical examinations for a variety of medical issues.
	Resource allocation	Assessing efficiency of medical treatments/techniques can help in developing priorities for allocation of resources.
	Outcome	Assessing healthcare outcomes and the costs associated with achieving those outcomes.
Outcomes	Health	Comparing indicators with desired outcomes for general health measures as outlined by the NHS.
	Quality of care	Improving quality of care while achieving health outcomes.
	Efficiency	Comparing indicators with desired outcomes with regards to measures of efficiency.
	Resource allocation	Comparing indicators with desired outcomes with regards to the allocation of resources and the distribution of healthcare throughout the general population.
Customer Service & Patient Satisfaction	Quality	Assessing whether patients received quality care while in the NHS system.
	Quantity of patients served	Ensuring an increase in patients served by the NHS in a timely manner.
Innovation and Economic Growth	Quantity	Number of new businesses, products, or services emerging out of open data.
	Variety	Variety of services and products offered through using open data.
	Labour	Number of individuals employed through the creation of new businesses, products or services using open data.
	Economic	Amount of economic value added to market through creation of new businesses, products and services.
	Research	Number of research programmes emerging from or using open data.

II.1.1. ACCOUNTABILITY

Proposition: Opening data can enable accountability at the political and economic levels, helping to ensure that health regulators are meeting standards, and that physicians and hospitals are held accountable for unwanted variation and poor practices (for example, in the number of individuals dying from preventable diseases).³⁴

Evidence from the Literature: Open data has a powerful role to play in ensuring that healthcare providers and regulators are held accountable for the quality of care they provide. Public access to performance data can provide important indications of variability in cost or quality of healthcare, allowing patients to make more informed decisions, and regulators to take necessary action. One well-known example concerns the publication by the New York State Cardiac Surgery Reporting System of hospital mortality rates following coronary artery bypass graft surgery. A number of subsequent studies found that the publication of this data increased market share for hospitals with lower mortality rates. In addition, there is evidence to suggest that the availability of the data served as something of a “shaming mechanism,” encouraging hospitals with higher mortality rates to change their practices and achieve improved outcomes.³⁵ Similarly, an analysis of forty-five peer-reviewed articles, published in the *Annals of Internal Medicine*, found evidence linking the public release of data to quality improvements in hospitals.³⁶

Increasing data transparency appears to have a particularly powerful role on the behavior of regulators. In the United Kingdom, a series of news stories based on public data brought to light the fact that the number of elderly patients dying from dehydration had doubled, and that the number killed by superbugs had risen by seven times between 1997 and 2010. Also in the United Kingdom, a doctor (unaffiliated with the NHS) used public data in 2009 to find that a hospital in Stafford had unusually high mortality rates. In each case, the revelations led to enquiries and subsequent actions by NHS regulators.³⁷

Concern does exist over the quality of data and the criteria used to measure quality. Data can be particularly problematic (and misleading) when it is based on subjective criteria or patient self-reporting. In addition, an over-reliance on quality indicators and rankings can skew incentives and behaviors among providers in ways that are not necessarily beneficial to patients.³⁸ Such concerns are, without a doubt, valid. However, they argue not so much against the use of data for accountability, but rather make a case for more thought being put into how data is collected, designed and disseminated. We discuss issues related to data quality below.

Organizations affiliated with the NHS have already instituted a set of practices geared towards increasing accountability at the organizational and clinical levels through the use of public data, specifically audits. Using a clinical auditing process (including a critical analysis of national audit reports and findings from an online survey of NHS Trusts England), Christopher Loughlan and the Cambridge Institute for Research, Education and Management (CiREM) showed how audits can lead to changes in practice and greater adherence to published recommendations or guidelines.³⁹ Loughlan found that these audits resulted in a number of positive outcomes, including increased prescription quality, patient management, reductions in patient harms, and resource allocation.

34 Henke, Kelsey and Whately, “Transparency,” 2011.

35 The data is somewhat mixed, but on balance, the evidence suggests that transparency did in fact lead to greater accountability. See <http://www.health.org.uk/public/cms/75/76/313/554/Public%20release%20of%20performance%20result.pdf?realName=UWXIXp.pdf>.

36 Constance Fung, Yee-Wei Lim, Soeren Mattke, Cheryl Damberg, Paul Shekelle, “Systematic Review: The Evidence That Publishing Patient Care Performance Data Improves Quality of Care,” 2008, accessed February 23, 2014, <http://annals.org/article.aspx?articleid=738899>.

37 Henke, Kelsey and Whately, “Transparency,” 2011.

38 For a discussion in the Canadian context, see <http://www.openmedicine.ca/article/view/497/465>.

39 Christopher Loughlan, “Clinical audit, impact and quality improvement: a discursive review,” in *New Principles of Best Practice in Clinical Audit*, ed. Robin Burgess (Radcliffe Medical PR: 2011), 128-149.

Taxonomy: As Table II.1 illustrates, within accountability, there are a number of distinctions that can be made, which should be considered in any discussion of the impact of open data.⁴⁰

- ▶ **Political:** Improved transparency and accountability in terms of reaching desired outcomes in public health.⁴¹
- ▶ **Economic:** Improved accountability and transparency in spending, cost, waste, and fraud.
- ▶ **Participatory/Citizen Involvement:** Increased public participation in healthcare services.
- ▶ **Organizational/Clinical:** Higher success rates with regard to internal NHS objectives on patient outcomes.

Examples:



Dr Foster Intelligence, a joint venture with the UK Department of Health, publishes a variety of open health data products, benchmarking hospitals and healthcare commissioners, to make the “NHS a safer, more efficient and more transparent service.” It also funds the Dr Foster Unit at Imperial College London to enable identification of potential problems in clinical performance. In the past Dr Foster Intelligence has, for instance, used mortality datasets to improve accountability at the hospital level, which ultimately identified hospitals in the UK with abnormally high mortality rates and poor clinical practices as the cause.



Following the release of the report entitled “**Better Procurement, Better Value, Better Care**,” the UK Department of Health is now requiring all NHS Trusts to start publishing their spending data openly and online to improve procurement efficiency and accountability.



IMS Health, a large US technology corporation with investments in England via IMS Ardentia, uses open health and other data to provide a set of commercial health intelligence tools that allow their customers to assess healthcare performance and impact.⁴²

40 Ezekiel Emanuel, and Linda Emanuel, “What is Accountability in Health Care?” *Annals of Internal Medicine*, 124, no. 2 (1996), accessed February 23, 2014, <http://annals.org/article.aspx?articleid=709376>.

41 Noor Hujiboom and Tijs Van den Broek, “Open data: an international comparison of strategies,” 2011, http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf.

42 IMS Health, “About IMS Health,” *IMSHealth.com*, accessed February 23, 2014, <http://www.imshealth.com/portal/site/imshealth>.



Dollars for Docs is a tool developed by **Pro Publica** in the United States to increase accountability by providing citizens with a tool to determine whether their healthcare professional has received drug company money. Citizens can either search by institution or organization, or they can view state data to see the total number of payments that were made to healthcare practitioners and institutions. They can also view data for each drug company respectively.

II.1.2. CHOICE

Proposition: Open data on patient experience and health outcomes will allow patients to seek out the services that best meet their individual healthcare needs. Furthermore, increased access to health data will allow patients to choose among many providers and treatments, in turn driving patient satisfaction, quality improvements and the enhancement of standards in many areas within the NHS.

Evidence from the Literature: Greater public access to data and information has increased consumer choice in a range of sectors, ranging from travel to finance to shopping. A similar enhancement of choice can be expected with regard to healthcare. Indeed, patient choice is an essential driver of quality and efficiency in healthcare. In the UK, this fact was recognized as far back as 2002, when the then Secretary of State for Health Alan Milburn stated that “choice is the primary means by which you can drive the NHS to focus on the needs of its individual members.”⁴³

There now exists widespread evidence that data does indeed influence the way patients seek healthcare, expanding their options and the range of sources from which they seek information. It has long been established, for instance, that medical information on the Internet plays a role in influencing patients’ decisions, in some cases an even bigger role than that played by their doctors.⁴⁴ A recent survey of 6,000 patients in the US found that nearly half (48.6%) went to the Internet first for information on cancer (compared to 10.9% who went to their doctor).⁴⁵ Likewise, a Pew Internet and American Life Project found that of the people who turned to the Internet for healthcare information, more than half reported that it influenced their treatment decision.⁴⁶ A recent survey on the impact of NHS Choices (nhs.uk) similarly finds a range of evidence to suggest that the availability of medical data plays a significant role in influencing how patients seek care.⁴⁷ Jeremy Hunt, the UK Secretary of State for Health has emphasized the importance of using data to provide patients with choices, arguing that “there is the business of healthcare but underneath all of this there is something much more important, which is putting people in control of the data that affects the single thing that matters more to them and their families...their healthcare.”⁴⁸

Open data can be particularly helpful in aiding patients in the access of information that will allow them to evaluate different treatment programmes. Open health data in this area should support the guidelines outlined in the NHS

43 Catherine Bochel, Nick Ellison, and Martin Powell, editors, *Social Policy Review 15: UK and international perspectives* (Bristol: The Policy Press, 2003).

44 Elizabeth Murray et al., “The Impact of Health Information on the Internet on Health Care and the Physician-Patient Relationship: National U.S. Survey among 1,050 US Physicians,” *Journal of Medical Internet Research* 5(3) (2003).

45 *Medical news Today*, “Patients Turn to Internet For Medical Advice Even Though They Trust Their Doctors,” July 12, 2012, accessed February 23, 2014, <http://www.medicalnewstoday.com/articles/247786.php>.

46 Pew Research Project, “Health Fact Sheet, 2013, accessed February 24, 2014, <http://www.pewinternet.org/fact-sheets/health-fact-sheet/>.

47 Paul Neslon, Joanna Murray, and Muhammed Saleem Kahn, “NHS Choices Primary Care Consultation Final Report, 2010, accessed February 23, 2014, <http://www.nhs.uk/aboutNHSChoices/professionals/developments/Documents/annual-report/primary-care-consultation-report.pdf>.

48 Jeremy Hunt, “Datapalooza IV Keynote Address,” (presented at Datapalooza IV, Washington D.C., June 3, 2013), accessed March 17, 2014, <http://www.youtube.com/watch?v=Om1BORltPTI>.

2013/14 Choice Framework.⁴⁹ In fact, the policy of offering patients a choice regarding where and from whom they receive hospital treatment was implemented by the NHS already in 2006, in order to drive up competition between providers and encourage efficiency and responsiveness.⁵⁰ Results from a study assessing the effectiveness of choice programmes showed that quality assessments will become more important in influencing patient decisions and choices as quality-related information is made more readily available. Recognizing this, NHS England has made a commitment to facilitating shared decision making between patients and their physicians, through embedding it as an objective within the NHS Constitution and the NHS Mandate.⁵¹

The link between data-influenced choices and positive outcomes is somewhat more complicated. Evidence does exist to suggest that data can have a positive influence. An NHS survey concluded that 37% of patients who used the NHS Choices website witnessed an appropriate decrease in GP consultations, and that this translated into savings of approximately £44 million a year. Based on an analysis of data conducted during the Swine Flu panic of 2009, the study also concluded that availability of data could be particularly helpful during epidemic and public services by limiting demand for medical services.⁵²

It is however important to note that a growing body of evidence also suggests the information on its own may not always lead patients to make better (or more informed) medical decisions.⁵³ This may be because, as Neil Manson and Onora O'Neill point out, patients are not always "rational" decision makers. In many cases, it may also have to do with limited numeracy skills, or limited ability to analyze and understand vast amounts of medical data.⁵⁴ Thus while recent years have witnessed a move towards more "patient-centered" practices, there is also a growing appreciation of the need for "shared decision-making" approaches.⁵⁵ Such approaches combine data-driven and information-armed patients with the clinical judgment of doctors. In such approaches, the presentation of data and its integration with structured "decision aids" can be critical.⁵⁶

Taxonomy: Choice has at least two dimensions:

- ▶ **Variety:** Increasing the amount of information available to patients for their decision-making.
- ▶ **Quality:** Increasing the quality, completeness, and timeliness of information provided to patients and the broader public to determine whether treatment is needed and/or available.

Examples:

NHS Choices provides information on NHS healthcare, social care and healthy lifestyles. It uses a combination of decision tools, open data on service providers and, increasingly, patient feedback to enable patients to compare and choose treatments and services. Content is co-produced with "trusted affiliates" such as Patientopinion.org.uk.⁵⁷

49 Department of Health, "NHS Outcomes Framework 2013 to 2014," Gov.UK, November 13, 2002, accessed February 23, 2014, <http://www.nhs.uk/choicein-theNHS/Rightsandpledges/NHSConstitution/Documents/2013/choice-framework-2013-14.pdf>.

50 Anna Dixon, John Appleby, Ruth Robertson Peter Burge, Nancy Devlin, and Helen Magee, "Patient Choice: How patients choose and how providers respond," 2010, accessed February 23, 2014, <http://www.kingsfund.org.uk/publications/patient-choice>.

51 "NHS England Commitment to Shared Decision Making," NHS England, accessed April 23, 2014, <http://www.england.nhs.uk/ourwork/pe/sdm/commitment/>.

52 Paul Nelson, Joanna Murray, and Muhammad Saleem Kahn, "NHS Primary Care Consultation Final Report," 2010, accessed February 23, 2014, <http://www.nhs.uk/aboutNHSChoices/professionals/developments/Documents/annual-report/primary-care-consultation-report.pdf>.

53 Judith Hibbard, "Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches that Facilitate the Use of Information in Choice," *Annual Review of Public Health* 24 (2003).

54 Neil Manson and Onora O'Neill, *Rethinking Informed Consent in Bioethics* (Cambridge: Cambridge University Press, 2007).

55 Glyn Elwyn, Adrian Edwards, Paul Kinnersley and Richard Grol, "Shared Decision Making and the Concept of Equipoise: The Competences of involving patients in healthcare choices," *British Journal of General Practice* 50 (2000), 892-897.

56 Annette O'Connor, Alaa Rostom, Valeria Fiset, Jacqueline Tetroe, Vikki Entwistle, Hilary Llewellyn-Thomas, Margaret Holmes-Rovener, Michael Barry and Jean Jones, "Decision aids for patients facing health treatment or screening decisions: systematic review," *British Medical Journal* 319 (1999).

57 National Health Service, "Your health, your choices," NHS, accessed February 23, 2014, <http://www.nhs.uk/Pages/HomePage.aspx>.



US company **WiserTogether** gathers open data through its Wiser Health Platform from consumers and doctors who have encountered similar medical issues and creates a list of best options based on clinical efficacy, financial considerations and treatment preferences. The aim of its products, using crowdsourced information, is to help patients “choose the right care at the right time” and make “evidence-based, cost-effective treatment decisions that are aligned to their personal preferences and financial constraints.”⁵⁸



US company **WiserTogether** gathers open data through its Wiser Health Platform from consumers and doctors who have encountered similar medical issues and creates a list of best options based on clinical efficacy, financial considerations and treatment preferences. The aim of its products, using crowdsourced information, is to help patients “choose the right care at the right time” and make “evidence-based, cost-effective treatment decisions that are aligned to their personal preferences and financial constraints.”⁵⁹



iTriage is a U.S. based company that provides patients with the information they need to make decisions about their symptoms, possible treatment methods, nearby care options, and medical providers.⁶⁰ Content is produced by an in-house clinical team of physicians, as well as through a survey of published literature.⁶¹ Drug information is licensed from the American Society of Health-System Pharmacists. iTriage’s database of symptoms, treatments, and other information is updated weekly.



A U.S. company, **MedWatcher**, is having an important impact on public health and accountability. Using data from the Food and Drug Administration, MedWatcher maintains a database of adverse side effects reported from different drugs. Though the company is still building its database, it will eventually serve as an important tool for patients in choosing between drug treatment programs.



Healthgrades is a U.S. service comparing healthcare providers, including hospitals, nursing homes, physicians and dentists based upon open patient safety data. Its ratings aim, for instance, to inform patients about where they are more likely to survive their hospitalization or encounter the least risk of major complications.⁶²

58 WiserTogether, “Home,” Wisertogether.com, accessed February 23, 2014, <http://wisertogether.com/>.

59 WiserTogether, “Home,” Wisertogether.com, accessed February 23, 2014, <http://wisertogether.com/>.

60 iTriage, “Take Charge of Your Health,” iTriageHealth.com, accessed March 17, 2014, itriage.com/.

61 iTriage, “About Our Medical Content,” iTriageHealth.com, accessed March 17, 2014, <http://about.itriagehealth.com/company-info/medical-content/>.

62 Healthgrades, “Find a Doctor | Doctor Reviews | Hospital Ratings,” Healthgrades.com, accessed February, 23, 2014, <http://www.healthgrades.com/>.

II.1.3. EFFICIENCY:

Proposition: The availability of open data will prompt NHS organizations to be more cost-effective, will accelerate data-sharing practices and communication strategies, and will improve care quality. Comparative data regarding efficiency across hospitals and administrations can be used to develop best practices and better standards. The impact of open data on efficiency places a special importance on the organizational setting of healthcare, including purchasing, labour, and professional practices.⁶³

Evidence from the Literature: Open data has a vital role to play in lowering costs, reducing fraud and increasing productivity—all vital steps towards creating a more efficient healthcare system. With healthcare today consuming ever-larger shares of national GDP (9.3% in the UK, 17.9% in the US),⁶⁴ finding a way to achieve greater efficiencies without negatively affecting outcomes is a major public policy priority. Whereas accountability focuses on broader, macro-level indicators of spending and costs, efficiency examines data between hospitals, clinics, and regions of the NHS, to identify areas of concern. These measures go beyond cost and distribution of resources, to include rates of treatment (including drug prescriptions, surgical operations, etc.), and the use of systems that enable better communication between healthcare practitioners and across agencies. Efficiency also encompasses other areas, such as wait-times and management of health records.

Numerous examples exist in which public access to data has identified or could potentially identify significant cost savings. In one widely publicized instance (mentioned in the introduction), Mastodon C and Open Healthcare UK (two startups) worked in collaboration with Ben Goldacre, a journalist and doctor, to examine data sets containing GP prescriptions for generic and patented statins in England. By examining regional variations, they were able to identify potential savings of £200 million.⁶⁵ Similarly, a study conducted by McKinsey concluded that using open data to assist patients with chronic conditions could help reduce the costs associated with poor drug adherence, currently estimated at \$100-\$289 billion annually in the United States.⁶⁶

Greater use of open data can also help control fraud—a major drain on health budgets, estimated at around £3.35 billion annually in the United Kingdom, and some \$80-\$98 billion in the United States.⁶⁷ For example, regulators and public watchdogs could combine demographic data and claim information for given regions to find cases of over-billing or fraudulent billing. In the United States, the Centers for Medicare & Medicaid Services (CMS) use public records and a variety of other data to assign risk scores to patients and providers, thereby allowing the CMS to better target its compliance resources.⁶⁸

Efficiency is a complicated (and sometimes loaded) concept in the context of healthcare. A narrow focus on economic efficiency can in fact be deleterious to patient welfare. More than fifty years after the publication of Kenneth Arrow's seminal essay on health economics, "Uncertainty and the welfare economics of medical care," markets and healthcare continue to co-exist uneasily.⁶⁹ Nonetheless, there is reason to think that better and more public data can help reduce some of the inefficiencies inherent in healthcare markets—and, importantly, do so without harming patient interests.

63 Rowena Jacobs, Peter C. Smith, and Andrew Street, *Measuring Efficiency in Health Care: Techniques and Health Policy* (Cambridge: Cambridge University Press, 2006).

64 World Bank, "Health Expenditure, total (% of GDP)," *Worldbank.org*, accessed February 21, 2014, <http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS>.

65 *The Economist*, "Beggars thy Neighbor: How scrutiny of freely available data might save the NHS money," *TheEconomist.com*, accessed February 23, 2014, <http://www.economist.com/news/britain/21567890-how-scrutiny-freely-available-data-might-save-nhs-money-beggar-thy-neighbour> (accessed February 23, 2014).

66 Manyika et al., "Open Data: Unlocking Innovation."

67 Mark Button and Colin Leys, "Healthcare Fraud in the New NHS Market – a Threat to Patient Care," *Centre for Health and the Public Interest*, 2013, <http://chpi.org.uk/wp-content/uploads/2012/06/CHPI-Healthcare-Fraud-a-threat-to-patient-care1.pdf>.

68 eHealth Initiative, "Leveraging Data and Analytics to Detect Fraud & Abuse in Healthcare," *ehidc.org*.

69 Kenneth J. Arrow, "Uncertainty and the Welfare Economics of Medical Care," *the American Economic Review* LIII(5) (1963)..

Taxonomy: There are a number of dimensions involved in any discussion of efficiency within the NHS system.

- ▶ **Operational:** Improving the speed at which patients are processed, treated, and discharged.
- ▶ **Economic:** Lowering the costs associated with each hospital or service, for example by comparing open data across regional and local contexts.
- ▶ **Communication and Technology:** Identifying spots in the healthcare system where communication needs to be improved both internally and externally (i.e., by reaching the broader public).
- ▶ **Treatment:** Determining which treatments work effectively, and for which populations.
- ▶ **Resource allocation:** Determining where and when services are needed most.

Examples:



The Camden Health Metrics Explorer, created by the Camden Coalition of Healthcare Providers (CCHP) (based in Camden, NJ, USA) and BlueLabs, is leveraging open datasets linked to health exchange information and claims data, providing actionable metrics in real time. They place a focus on “hotspotting”—the identification of patients who are heavy users of the healthcare system--providing targeted interventions to address their needs and change their utilization patterns. Based upon a metrics dashboard, the Coalition seeks to reach out to high risk patients to coordinate the critical period following discharge and the transition back to primary care. The Camden Coalition also works with community groups to provide preventive care and health education.⁷⁰



Aidin is a New York based startup that lists and assesses specialized-care providers to simplify the patient-discharge process, and is aimed at preventing overspending. Its mission is “to bring transparency to the healthcare system and empower patients with data to make more informed decisions on where they receive care”. Aidin integrates into the discharge planning workflow to free social workers from administrative tasks and re-center their time around patients.⁷¹



The CDEC Open Health Data Platform was developed in the United Kingdom by the Connected Digital Economy Catapult (CDEC) and Mastadon C, and is intended to facilitate innovation in the field of data analysis, data visualization, service design, and other web and app development enabling innovation using linked health data.⁷²

70 Ereik Dykstrant, “Camden Health Explorer,” 2013, accessed February 23, 2014, <https://www.newschallenge.org/challenge/healthdata/evaluation/camden-health-metrics-explorer>.

71 MyAidin.com, accessed February 23, 2014, <http://www.myaidin.com/>.

72 “Summary,” [OpenHealthData.cdehub.org/summary](http://openhealthdata.cdehub.org/summary), accessed April 25, 2014, <http://openhealthdata.cdehub.org/summary/>.

II.1.4. OUTCOMES:

Proposition: Publishing data on general outcomes across the NHS system drives competition between healthcare professionals, which can in turn spur quality improvements and innovation. Opening health data to analysts and researchers improves both the quality and consistency of evaluation.

Evidence from the Literature: Perhaps one of the most significant value propositions for open data is its potential to positively affect the mission that is in many ways at the heart of any healthcare system: improved patient outcomes. Early evidence linking open data with improved outcomes was somewhat mixed. Today, however, there exists considerable evidence suggesting that increased data transparency can lead to improved outcomes. For instance, several studies have found reduced mortality rates following the implementation of the New York State Cardiac Surgery Reporting System mentioned earlier (“Accountability”).⁷³ The NHS has also found that publishing comparable data on clinical outcomes, which it began some nine years ago, has reduced the number of deaths in heart surgeries by up to 1,000 annually.⁷⁴ Similar results have been seen in Germany and Sweden, where public reporting of health data has significantly improved a variety of health outcomes, including recovery following aortic aneurysm repair, and cardiology-related ailments.⁷⁵

Open data can also improve outcomes by playing a role in preventive healthcare. The rise of the “quantified self” movement allows individuals to compare their health status with demographic information contained in public databases. Applications like 100plus.com, for instance, benchmark individual data against public records maintained by the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC). Data can also be used to avoid adverse drug reactions by matching patient profiles to patterns of adverse reactions contained in public data sets. Conversely, public databases can be used to match patients, based on demographic or other features, to treatments that are likely to be effective. In another example, suggestive of the range of possibilities for data in preventive healthcare, the Veterans Health Administration in the United States uses data (including public registers of deaths) to monitor and prevent suicides among veterans.⁷⁶

Taxonomy: Though it is beneficial to think of outcomes primarily in terms of the health outcomes prioritized by the NHS, there are a number of other types of outcomes that can be identified.

- ▶ **Health outcomes:** Assessing whether the entire NHS, or its individual hospitals and regions, have achieved NHS health outcome measures.
- ▶ **Quality of care:** Increasing the quality of treatment while achieving health outcomes.
- ▶ **Efficiency:** Enables hospitals and regions to compare their services and outcomes, to identify weak spots within their individual systems.
- ▶ **Resource allocation:** Identifying high-risk patients, directing resources to high-need communities, all in service of improving overall outcomes.

73 Paul G. Shakelle, Yee-Wei Lim, Soren Mattke and Cheryl Damberg, “Does public release of performance results improve quality of care?,” *Health.org.uk*, 2008, accessed February 23, 2014, <http://www.health.org.uk/public/cms/75/76/313/554/Public%20release%20of%20performance%20result.pdf?real-Name=UWXIXp.pdf>.

74 Tim Kelsey, “Transparency in the NHS not only saves lives – it is a fundamental human right,” *The Guardian*, 2013, accessed February 23, 2014, <http://www.theguardian.com/society/2013/mar/12/nhs-transparency-open-data-initiative>.

75 Henke, Kelsey and Whately, “Transparency.”

76 Erin Bagalman, “Suicide Prevention Efforts of the Veterans Health Administration,” *Congressional Research Service*, January 10, 2013, <https://www.fas.org/sgp/crs/misc/R42340.pdf>.

Examples:



Kaiser Permanente, a nonprofit healthcare system in the US, used their paperless practice management system to discover that mortality rates were higher among stroke patients not offered statins as part of their care plan or those who discontinued use post-stroke. Based on their findings, Kaiser embedded a prompt within their electronic medical record system, Health Connect, to alert doctors at the point of care as to which ischemic stroke patients should receive statins. This is an example of how data can influence the so-called “choice architecture” of statin prescriptions and use to improve overall outcomes among stroke patients.⁷⁷



In the US, **Propeller Health** provides a small cap that fits on a standard inhaler for patients with asthma problems. When the inhaler is used, the cap records the time and location, using GPS circuitry. Those healthcare professionals that are providing management of the patient’s respiratory problems can subsequently retrieve the data. By comingling the open data taken over long periods of time, or by aggregating anonymized data across multiple patients, a picture emerges of hot spots – times and places where breathing is difficult-- that can help other patients improve their condition.⁷⁸



The Society for Cardiothoracic Surgery (SCTS) and the NHS have been collecting and publishing open data about all heart surgery carried out in the UK (through the Adult Cardiac Surgery, which is managed by the National Institute for Cardiovascular Outcomes Research (NICOR)) for the past nine years. Their iData app, for instance, allows users to place filters on different types of surgery data to generate a report showing analysis of all heart surgery in the UK over the last five years. By using the comparative data, there have been 1,000 fewer deaths in heart surgery units each year. Due to the success of this initiative, and in coordination with open data policies, this practice has now been expanded to include other types of surgeries, resulting in lower mortality rates and more cost savings system-wide.

77 See Kaiser Permanente, “Celebrating Our Stroke Research Saving Lives,” Kaiser Permanente, accessed February 23, 2014, <http://share.kaiserpermanente.org/article/celebrating-our-stroke-research-saving-lives/>; Douglas McCarthy, Kimberly Mueller, and Jennifer Wrenn, “Kaiser Permanente: Briding the Quality Divide with Integrated Practice, Group Accountability, and Health Information Technology,” *Commonwealth Fund*, June 2009, http://www.commonwealthfund.org/~media/Files/Publications/Case%20Study/2009/Jun/1278_McCarthy_Kaiser_case_study_624_update.pdf; Grant Okawa, “Integrating Evidence into KP HealthConnect: Making the Right Thing Easier to do,” *The Permanente Journal* 9(2) (2005).

78 See Tim Maly, “Asthmapping: Smart Inhalers Track Usage to Save Lives,” *Wired*, November 1, 2012, accessed February 23, 2014, <http://www.wired.com/2012/11/asthmapolis/>; David Van Kannon, “Open Data Standards and the Breath of Little Children,” October 16, 2013, accessed February 23, 2014, <http://davidvunkannon.wordpress.com/2013/10/16/open-data-standards-and-the-breath-of-little-children/>.



Archimedes and **Evidera**, two US-based companies that recently merged, use real-world open data alongside simulation data to provide decision makers with quantitative information about the outcomes they could expect from different clinical and administrative policies and programmes.⁷⁹



Within the US, the Office of the Assistant Secretary for Planning and Evaluation at the Department of Health and Human Services created an open data portal called **the Health System Measurement Project**. This portal tracks government data on key U.S. health system indicators, such as estimated numbers of young adults with health insurance coverage, percentage of employers offering health insurance by firm size, and other open data.⁸⁰



The **Scottish Longitudinal Study (SLS)** is a large-scale linkage study using statistical and administrative data, including census data, vital data such as births, deaths, and marriages, NHS Central Registrar Data, and education data.⁸¹ The SLS was created in response to a lack of longitudinal datasets in Scotland compared to other areas in the United Kingdom, and is being used to gain insight into the health and social status of Scotland over time.

II.1.5. CUSTOMER SERVICE AND PATIENT SATISFACTION:

Proposition: Improving patient access to data and providing convenient customer-oriented services, such as online appointment booking and prescription refills, will enable the NHS to achieve world-class customer service and improve efficiency.

Evidence from the Literature: Patient satisfaction is determined by multiple variables. Outcomes are of course among the most important. But customer service—and the overall patient experience—also plays an important role not only in determining satisfaction, but also in affecting medical decisions to seek or not seek treatment in the first place.

Measuring patient satisfaction is not straightforward. Self-reported satisfaction surveys can be unreliable, biased by unrealistic patient expectations. Nonetheless, there is growing evidence that better use of data (and information technology more generally) can help address common patient complaints such as waiting times or impersonal service. In a report on

79 Evidera, "About Us," Evidera.com, accessed February 23, 2014, <https://archimedesmodel.com/>.

80 Health System Measurement Project, "About the Project," ASPe.hhs.gov, accessed February 23, 2014, <https://healthmeasures.aspe.hhs.gov/> (accessed February 23, 2014).

81 Scottish Longitudinal Study, "Why was the SLS set up?" accessed April 26, 2014, <http://sls.lscs.ac.uk/about/why-was-the-sls-set-up/>.

“the wait list problem,” for instance, the British Columbia Medical Association found an important role played by public awareness and patient access to information about waiting times.⁸² In Ontario, a new programme uses clinical data on waiting periods for cancer patients to help doctors adjust staffing patterns and better serve patient needs.⁸³ Electronic health records that make use of data and cutting edge technology can also help: one study by the Arizona State University found that patient waiting times in emergency rooms were reduced by 22 percent at hospitals that deployed such EHRs.⁸⁴

Using data to improve customer service is still in an exploratory phase. A range of innovations are being deployed across healthcare systems. In the NHS, there have been efforts to use real-time data and qualitative data to better understand and enhance the patient experience.⁸⁵ Similarly in the United States, the Cleveland Clinic has been at the forefront of many customer service innovations. These include “open access scheduling,” “patient reported outcomes,” and an “open medical records policy” that allows patients to access their own records online. All these initiatives are part of an ongoing effort to use greater transparency as a tool not only to improve outcomes efficiencies, but also to change the way patients experience the process of seeking medical care.⁸⁶

Taxonomy: In the relationship between the NHS and their patients as consumers, two important dimensions can help to measure the effectiveness of open data in enhancing customer service.

- ▶ **Quality of care:** Ensuring that patients have a positive experiences of care, both with regard to their physician and the general healthcare system. Also, increasing the amount of information the patient has in their treatment programme.
- ▶ **Quantity of patients served:** Decreasing wait times, and increasing the number of patients served through emergency services, general practitioners and specialists. Shorter wait-times for GPs and specialists are especially important.

Examples:



The Friends and Family Test (FFT),⁸⁷ the biggest ever collection of patient feedback, has been used by the NHS since 2013 to improve the patient experience. It provides a simple headline metric which, when combined with follow-up questions, can be used as a tool to ensure transparency and seek improvements. While initially focused on all NHS Inpatient, A&E departments and providers of NHS funded maternity services, it is scheduled to be introduced to GP practice and the community and mental health services by the end of December 2014, and to the rest of NHS-funded services by the end of March 2015.

82 BCMA Council on Health Economics Policy, “Waiting Too Long: Reducing and Better Managing Wait Times in BC,” British Columbia Medical Association, June 2006, https://www.bcma.org/files/waiting_too_long.pdf.

83 Advizor Solutions, Inc., “MedAdvisor: Sudbury Regional Hospital Leverages Clinical Data to Reduce Patient Wait Times,” Advizor Solutions, Inc., accessed February 23, 2014, <http://www.advizorsolutions.com/medadvisor/sudbury/>.

84 Diana Manos, “ER wait times reduced 22 percent with advanced EHRs,” Healthcare ITNews, 2010, accessed February 23, 2014, <http://www.healthcareitnews.com/news/er-wait-times-reduced-22-percent-advanced-ehrs>.

85 Patient Experience, “From Data to Action: The Challenges of Driving Service Improvement,” Patient Experience, accessed February 23, 2014, <http://patientexperienceportal.org/article/from-data-to-action-the-challenges-of-driving-service-improvement>.

86 Jeff Rowe, “5 Ways Cleveland Clinic Improved its Patient Engagement Strategies,” Healthcare IT News, October 1, 2013, accessed February 23, 2014, <http://www.healthcareitnews.com/news/5-ways-cleveland-clinic-improved-its-patient-engagement-strategies>.

87 NHS England, “Introduction to the Friends and Family Test,” England.nhs.uk, accessed February 23, 2014, <http://www.england.nhs.uk/ourwork/pe/fft/>.



Patient Opinion is an independent non-profit feedback platform for health services in the United Kingdom, facilitating conversations between patients and healthcare services. The NHS uses the stories and comments left by patients on this site to respond directly to patients, occasionally publishing comments on the NHS Choices website, or integrating feedback to change and improve services.⁸⁸



OpenNotes started a movement in 2010 to enable patients to easily read notes written about their care, and to bring more transparency to medical records. More than 100 primary care doctors from three diverse medical institutions across the United States began sharing notes online with their patients. Each site was part of a 12-month study to explore how sharing doctors' notes could affect healthcare.⁸⁹



CakeHealth, a U.S. based company, organizes medical billing plans online, so patients can know what they are paying for, and find plans that are suitable for their needs.⁹⁰ This site is suited for the U.S. context, where private insurance companies require deductibles, and make it difficult for patients to track the different services and costs that come out of their health insurance plan.

II.1.6. INNOVATION AND ECONOMIC GROWTH

Proposition: Currently, health data is difficult to request and often expensive. If more data is collected and distributed, the UK will build on its standing as a center for research and innovation. Having all the data in one place will mean analyses can be done at the population level, making the UK an attractive destination for health research as well as health innovation.

Evidence from the Literature: Considerable evidence exists to suggest that open data can spur economic growth and innovation at both a national and regional level. The figures are striking. McKinsey estimates that the United States stands to create more than \$300 billion in new value through greater data transparency.⁹¹ The Centre for Economics and Business Research (CEBR) estimates that the EU's move towards more openness in data will create 58,000 jobs in the UK by 2017 and add £216 billion to the nation's GDP.⁹² Such figures are not limited to developed nations. As the World Bank (and others) have found, developing countries stand to be among the biggest beneficiaries of greater transparency.⁹³

88 "About Patient Opinion," Patient Opinion, accessed April 23, 2014, <https://www.patientopinion.org.uk/info/about>.

89 OpenNotes, accessed February 23, 2014, <http://www.myopennotes.org/>; Andrew Clark, "Could bedside TVs be used to give patients access to medical records," The Guardian, February 18, 2014, accessed February 23, 2014, <http://www.theguardian.com/healthcare-network/2014/feb/18/hospital-televi-sions-patients-medical-records>.

90 "About." CakeHealth.com. <https://cakehealth.com/about> (accessed March 17, 2014).

91 Henke, Kelsey and Whately, "Transparency."

92 Graham Jarvis, "EU open data promotion could benefit UK economy, says CEBR," accessed February 23, <http://www.computerweekly.com/feature/EU-open-data-promotion-could-benefit-UK-economy-says-CEBR>.

93 The World Bank, "Open Data Essentials," Data.Worldbank.org, accessed February 23, 2014, <http://data.worldbank.org/about/open-government-data-tool-kit/knowledge-repository#benefits>.

The link between open data and growth runs through many pathways. One of the most important is the opportunity offered by open data to entrepreneurs and innovators. Pharmaceutical companies, for instance, can use open data to run more efficient clinical drug trials.⁹⁴ We have already seen some of the potential efficiencies offered by data to providers and hospitals. Importantly, the potential for innovation is not restricted to existing players. In recent years, a changing healthcare data landscape has led to the creation of entirely new industries and new types of companies. Data mining operations, automated imaging and image analysis businesses, a plethora of new consumer websites and mobile apps: these are just some of the entities emerging in the new ecosystem. Tim Kelsey, the National Director for Patients and Information at the NHS, has emphasized that open data comes first, and innovation will follow. “Our job is to simply make the data available,” he said at the Strata Conference in London 2013. “The benefit comes only when [entrepreneurs] become properly engaged and start turning that data into accessible tools and services.”⁹⁵

Taxonomy: Indicators to determine the impact of open data on innovation and economic growth can be split into several categories. These include:

- ▶ **Quantity:** Increasing the number of services provided, and increasing competition between different businesses and innovators.
- ▶ **Variety:** Increasing the variability of new services and applications.
- ▶ **Labour:** Increasing the number of individuals employed through the development of new services and applications.
- ▶ **Economic:** Increasing the amount of financing given to third party initiatives that can benefit the healthcare system.
- ▶ **Research:** Ensuring that researchers are made aware of available data, and that it can contribute to their research programmes effectively.

Examples:



Sanofi’s Inaugural Data Design Diabetes Challenge is an ongoing challenge that encourages private sector and public innovation to improve patient/ user experience or change patient behavior by creating solutions from open datasets.⁹⁶ Ginger.io, the inaugural winner, is a Boston-based company that offers analytics for both patients and providers.⁹⁷ The latest winner, PHRQ’s **Connect & Coach**, links supermarket-based dietitians with customers who want to improve their eating habits, particularly those living with diabetes.⁹⁸

94 Institute of Medicine of the National Academies, “Discussion Framework for Clinical Trial Data Sharing: Specific Topics for Public Feedback,” *Questions for the Public*, January 2014, <http://www.iom.edu/Reports/2014/Discussion-Framework-for-Clinical-Trial-Data-Sharing.aspx>.

95 Tim Kelsey, “Demonstrating the Actual Economic Value of Data,” (presented at the Strata Conference, London, UK, 2013), <http://www.youtube.com/watch?v=s8HCbXsC4z8>.

96 “Data Design Diabetes: Three years of entrepreneurs taking on the challenge to innovate in diabetes,” Sanofi, accessed February 23, 2014, <http://www.data-designdiabetes.com/the-challenge/>.

97 “About,” Ginger.io, accessed February 23, 2014, <http://ginger.io/about/>.

98 “Connect & Coach,” PHRQL, accessed February 23, 2014, <http://phrql.com/connect-coach/>.



Kaiser Permanente, a nonprofit healthcare system in the US, announced in 2013 the launch of its first application programming interface (API) programme, InterchangeSM. Interchange makes Kaiser's public facility and location information available for third-party developers to innovate and integrate into apps. The open data made available include locations, hours of operation and specialty information for Kaiser Permanente's hospitals and medical offices, encompassing over 600 facilities. Kaiser also announced its plan to increase the number of securely accessible public data sets to include more health and wellness data, health plan data, brand-related data, and research publications.⁹⁹



Across the US, the **Blue Button** has become the default symbol indicating that patients can download their own health information. Depending on the particular implementation, users can download a variety of information in multiple formats, including text and PDF. The Blue Button Co-design Challenge asks developers to create new patient applications able to leverage open health data. Several ideas gathered from patients through crowdsourcing were published online so developers could create apps to address them. GenieMD was voted as the winning app. It provides users with a variety of patient services, including symptom checking, provider and pharmacy search, and generating alerts through an intuitive interface.¹⁰⁰



Nesta used open data to publish a report on GP practices, to examine which GPs are early adopters of medical innovations.¹⁰¹ By using open data they discovered that GPs rely on a range of different resources to identify and learn about new innovations.

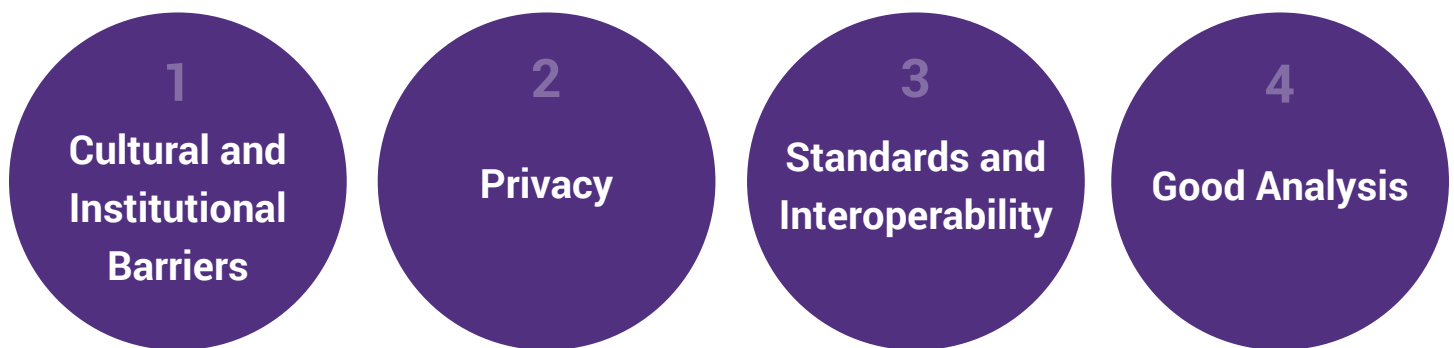
99 "Interchange by Kaiser Permanente Unleashes the Power of Digital health Through Open Apl, PR newswire, June 3, 2013, accessed February 23, 2014, <http://share.kaiserpermanente.org/article/interchangesm-by-kaiser-permanente-unleashes-the-power-of-digital-health-through-open-api/>.

100 Health 2.0 "Blue Button Co-Design Challenge," Health2con, accessed February 23, 2014, <http://www.health2con.com/devchallenge/blue-button-co-design-challenge/>.

101 Nesta, "Which doctors take up promising new ideas? New insights from open data," January 28, 2014, accessed February 23, 2014, <http://www.nesta.org.uk/publications/which-doctors-take-promising-ideas-new-insights-open-data>.

II.2 Potential Challenges and Barriers to Open Data

The preceding discussion makes evident the opportunities offered by open data. Along with the opportunities, however, come certain risks. Open data is in many ways a “disruptive” phenomenon.¹⁰² Like all disruptive technologies, its implementation is likely to throw up several challenges and obstacles. This section discusses four obstacles to the implementation of open data in a healthcare setting:



The list is not exhaustive. Moreover, because the field of open data is rapidly evolving, a moving target, the list is likely to change over time. Nonetheless, this section represents a preliminary attempt to consider some of the key challenges that are likely to affect the potential of open data.¹⁰³

II.2.1. CULTURAL AND INSTITUTIONAL BARRIERS

Institutional, bureaucratic and cultural resistance pose a significant challenge to data transparency. A recent study of barriers to open data conducted in the Netherlands placed “closed government culture” at the top of a ten-item list.¹⁰⁴ Such resistance can stem from vested interests (e.g., fear of losing authority, or of being exposed) or simply from an inherent fear of the unknown. In either case, achieving the potential of open data, in healthcare as well as other sectors, often requires what *The Guardian* calls “cultural change in the public sector.”¹⁰⁵

Not all resistance is bureaucratic in nature, nor is it limited to the public sector. Resistance to open data can be more subtle, stemming from engrained habits, or personal judgments and biases established over decades. Patients, providers, and administrators must all adapt to a new way of collecting and using data; as a recent McKinsey report puts it, “this is a personal revolution as much as an analytical one.” The report cites the example of doctors who may be reluctant to replace

102 Nigel Shadbolt, “The Disruptive Power of Open Data in Health and Care,” *Open Data Institute*, April 2013, Netherlands Organisation for Applied Scientific Research, accessed February 23, 2014 <http://www.england.nhs.uk/iscg/wp-content/uploads/sites/4/2013/07/ISCG-003-slides-NS.pdf>.

103 Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni, Ronald Meijer and Roexsana Sheikh Alibaks, “Socio-technical Impediments of Open Data,” *Electronic Journal of e-Government* 10 (2) (2012): 156-172.

104 Tijs van den Broek, Bas Kotterink, Noor Huijboom, Wout Hofman, and Stef van Grieken, “Open Data Need a Vision of Smart Government,” <http://share.psi.eu/papers/TNO.pdf> (accessed February 23, 2014).

105 Jonathan Gray, “The best way to get value from data is to give it away,” *The Guardian*, 2011, accessed February 23, 2014, <http://www.theguardian.com/world/datablog/2011/dec/13/eu-open-government-data>.

or augment their clinical judgment, built over years of experience, with data.^{106,107}

It is also worth emphasizing that resistance is not limited to the adoption of open data, or indeed to technology. In a 1998 *BMJ* paper on “Change and Resistance to Change in the NHS,” Diane Plamping wrote about a history of opposition to newness, and the difficulty of altering behavior in complex organizations.¹⁰⁸ In this sense, the challenges confronting open data within organizations like the NHS are part of the more general challenge of “change management.” The literature on this topic is vast, but emphasizes collaboration, clear explanation of the benefits, and sincere efforts to address the concerns of those most affected by change.¹⁰⁹ In addition, in the context of open data, creating an ecosystem (including technical documentation, APIs, and applications) that clearly demonstrates the benefits and viability of data transparency can be helpful in overcoming resistance.¹¹⁰

Relevant Legislation

The legislative landscape for open health data should be seen both in light of laws governing the confidentiality and treatment of medical health records, and the legal framework of open data itself. Medical data in the United Kingdom is governed by four legislative acts:¹¹¹ (1) Public Records 1958; (2) Data Protection Act 1998; (3) Access to Medical Reports Act 1988; and (4) Access to Health Records Act 1990.

Open data should also consider laws governing the use of identifiable data in research, such as the Data Protection Act (1998) and the Caldicott Standards (1997) and is based on the European Data Protection Directive (95/46/EC).¹¹² Existing regulation concerns the use of personal data across a wide range of sectors, including patient and medical data. A fundamental principle of the Data Protection Act (1998) is to use the minimal amount of personal data necessary to satisfy a purpose, to ensure that individual service users cannot be identified from data used to support purposes other than primary or direct care.¹¹³ To satisfy the Data Protection Act, the NHS supports the anonymisation or de-identification of data in some cases, and the pseudonymization of data in others. When anonymized, identifiable information is obscured within the person’s records sufficiently so as to garner minimal risk. This is done through the removal of personal identifiers, the use of identifier ranges (such as using value ranges instead of age), and through the use of pseudonyms.

Since not all open health data concerns patient records (e.g., many NHS bodies also generate financial information), these legislative acts are not always relevant to the protection of NHS data. When these acts do not apply, the Freedom of Information Act (2000) should be viewed as the relevant legislative force. This legislation was amended in 2012 to encourage public sector organizations to increase the number of datasets they make available, in a re-usable format.¹¹⁴

106 Manyika, James et al. “Open data: Unlocking innovation,” 86.

107 National- or regional-level cultural factors may also play a role. Certain Asian countries, for instance, may not place the same priority on openness and transparency as Western nations. It is important, of course, not to succumb to cultural essentialism in such discussions. See Robin Hicks, “Cultural Resistance to Open Government Data in China?” *FutureGov*, April 27, 2010, accessed February 23, 2014, <http://www.futuregov.asia/articles/2010/apr/27/cultural-resistance-open-govt-data-asia/>.

108 Diane Plamping, “Change and Resistance to change in the NHS,” *British Medical Journal* 317 (69) (1998).

109 “Leadership and Engagement for Improvement in the NHS,” *The King’s Fund Leadership Review, 2012*, http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/leadership-for-engagement-improvement-nhs-final-review2012.pdf.; and “Successfully Implementing Change,” *Victorian Quality Council, 2006*, http://www.health.vic.gov.au/qualitycouncil/downloads/successfully_implementing_change.pdf.

110 “Sharing ODDC Updates in Montreal,” *Open Data Research*, June 28, 2014, accessed February 23, 2014, <http://www.opendataresearch.org/content/2014/582/sharing-oddc-updates-montreal>.

111 “Relevant Legislation,” *Medical Protection*, accessed March 14, 2014, <http://www.medicalprotection.org/uk/booklets/medical-records/relevant-legislation/>

112 “Data Protection Legislation,” *Wellcome Trust*, accessed March 14, 2014, <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Personal-information/Data-protection-legislation/index.htm>.

113 “Pseudonymisation and Anonymisation of Data Policy,” *NHS Business Service Authority*, accessed March 14, 2014, http://www.nhsbsa.nhs.uk/Documents/NHSBSACorporatePoliciesandProcedures/Pseudonymisation_and_Anonymisation_of_Data_Policy.pdf.

114 Roslie Marshall, “Freedom of Information Act Ameded to Improve government’s Open Data Agenda,” *V3.co.uk*, May 2, 2012, accessed March 14, 2014, www.v3.co.uk/v3-uk/news/2172077/freedom-infomration-act-amended-improve-goernment-s-agenda.

II.2.2. PRIVACY

Sometimes two core societal values can come into tension. Even as governments increase their commitment to data transparency, there is concern that privacy could be threatened. On the one hand, there is little doubt that existing privacy legislation can sometimes represent a barrier to open data.¹¹⁵ The Dutch study mentioned above cites privacy concerns and legislation as the second most important barrier to open data implementation (after cultural barriers).¹¹⁶

On the other hand, open data offers new challenges to privacy, pushing the barriers of existing protections and legislation in ways that must be addressed by organizations and authorities. Concerns over privacy include worries that personally identifiable (PII) data may be misused or inappropriately shared; that data made available in the name of the public interest may be used for unauthorized commercial purposes; and that data may be inaccurate or contain wrong information potentially harmful to individuals.

Efforts to de-personalize and anonymize data form an important part of any attempt to address privacy concerns. Yet anonymization represents only a partial response. It is now well established that even apparently anonymized data can be aggregated or otherwise manipulated so that it is de-anonymized and effectively tied to an individual. For example, researchers have conducted studies using crowdsourced metadata in the form of phone numbers, and found that, when combined with other free and public data sources such as Yelp, Google Places, and Facebook, it is possible to identify ostensibly anonymized individuals.¹¹⁷ Similar instances have been documented in a number of cases by researchers and real-life events, calling into question the usefulness of anonymization as a weapon against privacy.¹¹⁸

The difficulties of protecting privacy should not be taken as an argument against greater transparency. They do, however, call for a more measured approach to open data, one that takes into account the need for public trust and that builds privacy protections into the architecture of open data policies. For example, policies should include informed consent provisions (to whatever extent practical: it is not always possible to seek consent, nor are all individuals always fully informed).¹¹⁹ In addition, to bolster the role of the individual in privacy-protection, data collection policies can integrate a more flexible opt-out clause, allowing users to revoke consent at any time.¹²⁰

It must also be recognized that, although fully informed consent is not always possible, it is not desirable to return to a paternalistic model that takes away a patient's ability or right to consent. Building oversight into the consent and data collection and use process is therefore an important part of ensuring privacy. This can be done through third-party audits and reviews of anonymization techniques, or by building greater awareness of privacy rights, for example in the design and wording of consent forms.¹²¹ Finally, any efforts to open data must be accompanied by strong disclosure and notification mechanisms to inform individuals and the public when data violations do occur. Such oversight is an essential part of ensuring accountability and discouraging practices that may lead to privacy or security breaches.

115 Stefan Kulk and Bastiaan van Loenen, "Brave New Open Data World?" *International Journal of Spatial Data Infrastructures Research* 7 (2012):196-206; and Association of Academic Health Centers, "HIPAA Creating Barriers to Research and Discovery," Association of Academic Health Centers, http://www.aahc-dc.org/Portals/0/pdf/AAHC_HIPAA_Creating_Barriers.pdf.

116 Tijs van den Broek et al., "Open Data -- Smart Government."

117 Jonathan Mayer, "MetaPhone: The NSA's Got Your Number," *Web Policy*, December 23, 2013, accessed February 23, 2014, <http://webpolicy.org/2013/12/23/metaphone-the-nsas-got-your-number/>.

118 Stefan Kulk and Bastiaan van Loenen, "Brave New Open Data World?" 201.

119 Indeed, the EU already requires an explicit consent requirement before data can be collected and used. See Paul M. Schwartz, "The EU-US Privacy Collision: A Turn to Institutions and Procedures", 126 *Harv. L. Rev.* (1966).

120 S.K. Ketler, "The Rebirth of Informed Consent: A Cultural Analysis of the Informed Consent Doctrine After *Schreiber v. Physicians Insurance Co. of Wisconsin*," *Northwestern University of Law*, 95(3) (2001): 1029-1056.

121 Daniel Solove. "Privacy Self-Management and the Consent Dilemma," 126 *Harv. L. Rev.* 1880 (2013).

Safeguards for data sharing in the NHS: the care.data Programme

care.data - NHS England's initiative to evolve existing data flows from hospitals to other NHS funded care contexts in order to enable analysis of patient outcomes along the pathway of care and planning for more integrated services, as well as improve resources for clinical research - was postponed for six months after concerns were raised by doctors and patient groups about the safeguards for the uses of the data and the quality of public information. The postponement was welcomed by critics and civil liberties groups who argued that most people did not understand the procedures under the new system, and that there would be little way for the public to determine who was using their medical records.

In light of these concerns, the British government has introduced new legislation to provide for independent scrutiny of data sharing in health care, and also to give patients the right to opt out.

care.data is now being re-developed to ensure that patients and professionals are properly informed of the safeguards around confidentiality and in relation to uses of the data - in ways that enable the clinical and research benefits to be realised.

II.2.3. STANDARDS AND INTEROPERABILITY

Many of the value propositions outlined for open data depend to a significant degree on “data liquidity”—the ability of data to flow easily among providers, patients and other groups involved in healthcare. Data liquidity has been described as occurring “when health [data] flows faster and more freely”; or when health IT can “bring ... information immediately to clinicians, patients, and others when and where they need it.”¹²² When a complex, multi-leveled agency (such as the NHS, but this also applies to many other government agencies) decides to make data open, this does not necessarily mean that data becomes widely accessible. For one thing, much government data exists in what Bill Schrier, writing for the O’Reilly Labs, calls “ancient media” and “medieval databases”— non-digital, archived formats.¹²³ Even when data is electronic, it often exists in proprietary formats, effectively trapped in information silos.

It is therefore essential that the release of open data is accompanied by openness and transparency at the level of the technical standards that define that data. The more widespread a standard, and the more transparent its definition, the more the spirit of openness inherent in open data can actually be fulfilled. Every effort should be made to define data using standards like XML, and through the use of APIs that enhance accessibility. As a recent report commissioned by the UK Cabinet Office makes clear, the use of open standards doesn’t simply promote access and openness: it also prevents lock-in to proprietary IT systems, promotes innovation and competition, and can lower the overall cost of large IT projects.¹²⁴

In addition to technical interoperability, there is a need to promote semantic interoperability through the establishment and enforcement of data standards across datasets. The NHS has established a set of data standards and a set of guidelines to ensure consistent data formats and coding systems for electronic records within the NHS.¹²⁵

Finally, it is worth mentioning that interoperability itself can mean many things and has multiple levels in a healthcare setting. The box below (Box I) describes three different types of interoperability, as included in a 2008 report by the UK Information Standards Board for Health and Social Care.

1) Technical Interoperability: “the ability of two or more systems to exchange data accurately, effectively, securely, and consistently such that the relationship between the data items is preserved between the sending and receiving system.”

2) Semantic Interoperability: “the ability to communicate and exchange information accurately, effectively, securely, and consistently between different information technology systems, software applications, and networks in various settings, and exchange data such that clinical or operational purpose, context and meaning of the information are preserved and unaltered. The system will be capable of utilising this information to provide simple and complex alerts and interface with decision support applications. “

3) Process Interoperability: “the ability to communicate and exchange information accurately, effectively, securely, and consistently between different information technology systems, software applications, and networks in various settings, and exchange data such that clinical or operational purpose, context and meaning of the information are preserved and unaltered. The system will be capable of utilizing this information to provide simple and complex alerts and interface with decision support applications and in addition will be capable of filtering and summarising information and presenting this to the clinical team along with triggers that support work processes and care plans.”

122 Paul K. Courtney, “Data Liquidity in Health Information Systems,” *Cancer Journal*, Jul-Aug; 17(4) (2012): 219–221.

123 Bill Schrier, “Chapter 28: Toads on the Road to Open Government Data,” *Open Government (Mountainview: Creative Commons) 2010*, http://chimera.labs.oreilly.com/books/123400000774/ch28.html#ancient_media.

124 Sally Weston, Martin Kretschmer and Jenifer Piesse, “Open Standards in Government IT: A Review of the Evidence,” *Centre for Intellectual Property Policy & Management (CIPPM)*, October 25, 2012, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78891/Review-of-the-Evidence_CIPPM.pdf.

125 HSCIC, “Information Standards,” *Systems.HSCIC.Gov*, accessed March 26, 2014, <http://systems.hscic.gov.uk/data>.

II.2.4. GOOD ANALYSIS

It may be true, as so much of the technical and business literature argues, that a decision is only as good as the underlying data. However, good data does not guarantee a good decision. Another way of saying this is that data does not exist or “act” in a vacuum: it is mediated through human judgments and subjectivity. Those judgments may be flawed; data can be misinterpreted and misunderstood, often due to personal biases or organizational or cultural pressures. Arguably, such mistakes are more likely to occur in situations where large amounts of data are suddenly made available: a data glut, as much as data paucity, can pose severe analytical challenges.¹²⁶

Open data can also be deliberately misinterpreted or misused. A recent report cites the (hypothetical) example of MRI providers who use open data to “target consumers for treatment based on their characteristics, rather than actual need.”¹²⁷ Similarly, open data can be used to target or market to patients from disadvantaged and less-informed demographics.¹²⁸

A comprehensive data transparency policy will therefore embed its approach to open data within a human and social context. It will take account of ways data can be abused or misused. It will focus on developing not only technical capacity within an organization, but also human analytical capacity. As a recent *Harvard Business Review* article entitled “Good Data Won’t Guarantee Good Decisions” puts it: “IT needs to spend more time on the ‘I’ and less on the ‘T.’”¹²⁹ Training, education, mentoring and general human capacity building are all part of this process, and essential components of any effort to increase data transparency.

126 According to a recent study by the IDC, only 3% of data in the world is currently tagged and ready for analysis. See “Study: Less than 1% of the world’s data is analysed, over 80% is unprotected,” *The Guardian*, December 19, 2012, accessed February 23, 2014, <http://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume>.

127 Manyika et al., “Open Data: Unlocking Innovation,” 86.

128 Jeni Tennison, “Four Challenges for Open Data,” *Open Data Institute*, September 18, 2012, accessed February 23, 2014, <http://theodi.org/blog/four-challenges-open-data>.

129 Shvetank Shah, Andrew Horne, and Jamie Capella, “Good Data Won’t Guarantee Good Decisions,” April 2012, accessed February 23, 2014, <http://hbr.org/2012/04/good-data-wont-guarantee-good-decisions/ar/1>.



Part III
Towards
a Conceptual
Framework
for Open Data
in Healthcare

Part I and **Part II** of this paper have highlighted some of the potential and challenges of open data in healthcare. In this section, we seek to develop a conceptual framework that will help to achieve that potential and limit those challenges. A conceptual framework can help to measure the impact of open data. It offers a model that will allow the NHS to become more *goal-oriented*, permitting it to set clear priorities for the use of open data and targets for change. In addition, a conceptual framework can help the NHS become more *agile*, allowing it to focus on continuous improvement by periodically measuring results, thereby providing the basis for adjustments (tactical and strategic) to keep open data projects on track. Our overall goal in this section is to develop a roadmap and research agenda for the implementation of open data policies and applications in a healthcare setting.

III.1. What is a conceptual framework and why do we need one?

Although there is now widespread recognition of the possibilities offered by open data, we are still at the early stages of understanding how best to use it. This is true not just of open data in a healthcare setting, but also a wide range of other settings and governance contexts. Policymaking is often easier than policy implementation: it is one thing to say what *should* be done, and quite another to know *how* to do it. In addition, even when we know how to implement a policy, it is often challenging to measure the impact of that policy.¹³⁰

A large part of the problem—a problem we seek to address in this section—stems from the fact that we lack a conceptual vocabulary to describe open data interventions.¹³¹ As Francis Fukuyama, writing on the poor state of assessing governance worldwide, put it in a recent paper: “[W]e cannot measure what we cannot adequately conceptualize, we have to start with the concept first.”¹³²

A conceptual framework can be understood as a theory of change, or a logic model, that helps us understand what specific interventions or actions are needed to achieve a desired outcome. It establishes metrics to measure, and a methodology to achieve, success. It does so by establishing the parameters and components of any intervention, allowing us to test assumptions, identify the relationships among resources available for an intervention, and establish the activities necessary to achieve a particular goal.

The logic model we have developed for the NHS is illustrated in Figure 2, and described more fully below. Its key components and the logical chain it establishes can be explained as follows:

The use of certain kinds of inputs and data (INPUTS: OPEN DATA), by certain kinds of users (USERS), for certain kinds of activities (ACTIVITIES), will achieve certain outputs (OUTPUTS) and outcomes (INDICATORS) that indicate impact (IMPACT). Specific methodologies (METHODOLOGIES) will be used to collect and measure indicators, helping to assess impact.

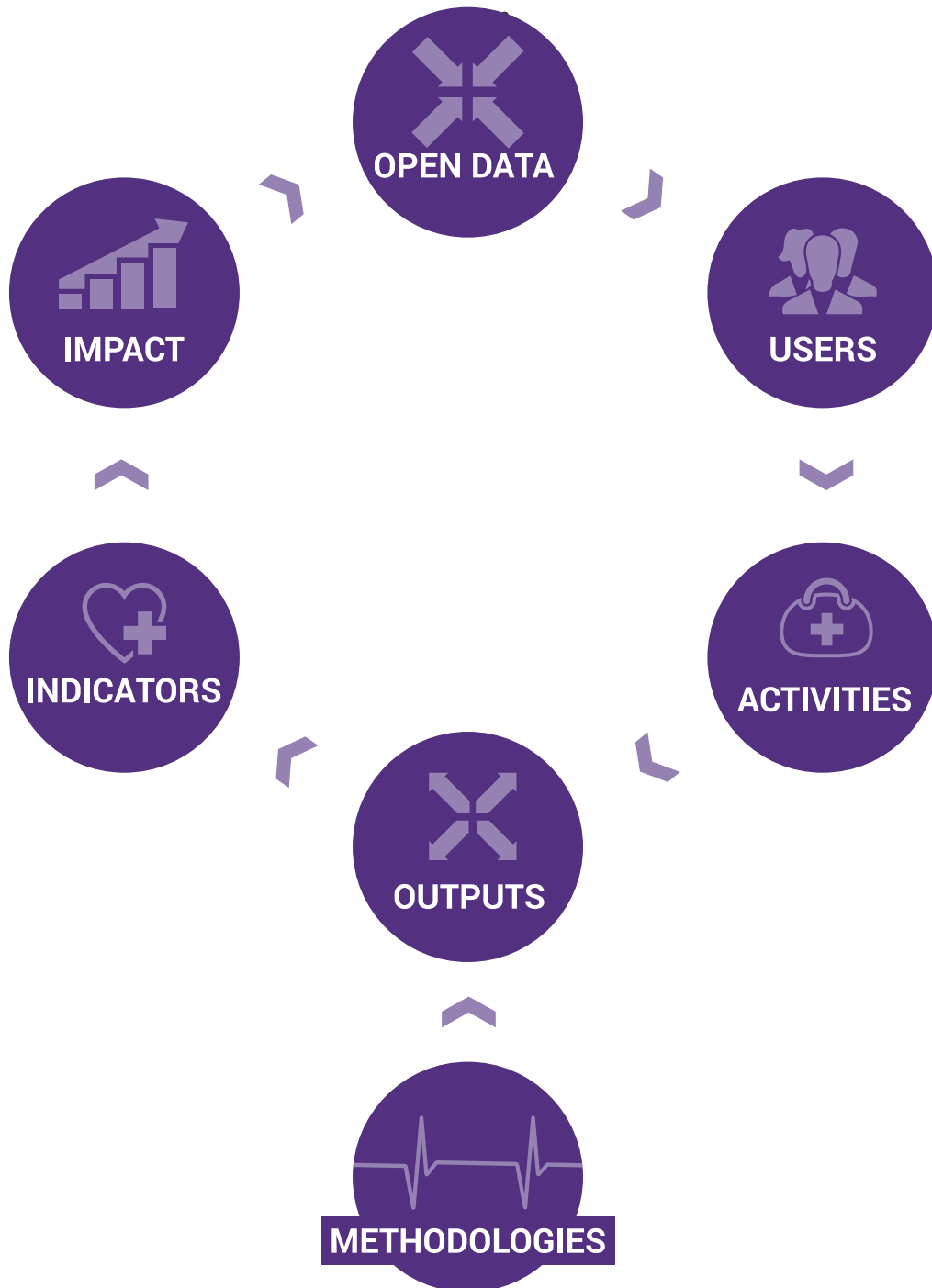
130 See also Karolis Granickas, “Understanding the impact of releasing and re-using open government data,” EPSI Platform, Aug 29, 2013, accessed March 23, 2014, <http://www.epsiplatform.eu/content/understanding-impact-releasing-and-re-using-open-government-data-0>.

131 See also Muriel Foulonneau, Sébastien Martin, and Slim Turki, “How Open Data Are Turned into Services?” *Exploring Services Science: Lecture Notes in Business Information Processing* 169 (2014):31-39.

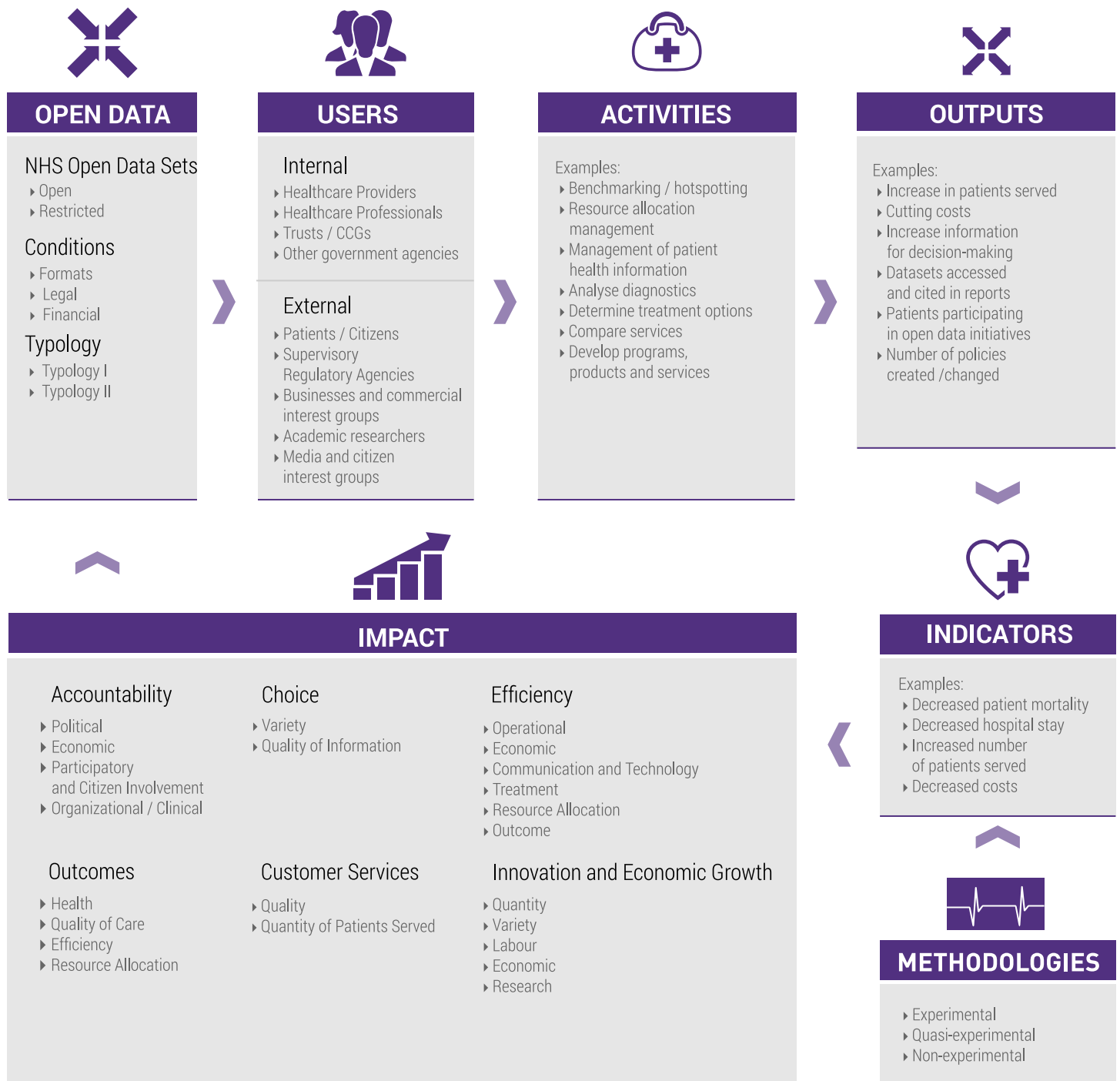
132 Francis Fukuyama, “What Is Governance,” Working Paper 314, Center for Global Development, January 2013, accessed March 23, 2013, http://www.cgdev.org/files/1426906_file_Fukuyama_What_Is_Governance.pdf.

In developing this conceptual framework, we were guided by two overriding principles, each of which is central to a successful implementation of open data by the NHS:

- ▶ Focus on results, determining impact and effectiveness. We want to allow for clear attribution of outcomes resulting from the use of open health data.
- ▶ Architect metrics into design, embedding them from the outset for systematic use at each stage of the open data management cycle: planning, implementation, monitoring, reporting, and evaluation.



CONCEPTUAL FRAMEWORK



OPEN DATA



III.2. Inputs: What open data set was used?

At the core of our conceptual framework lies the data that is accessed and used. This data (and more generally the set to which it belongs) will determine the extent and quality of openness, and by extension will set the boundaries of what can be achieved through greater transparency.

As a reminder, we have suggested that there are several categories of data within the NHS, and that data should be seen as existing on a spectrum of “accessibility.” At one end of the spectrum, there is **open health data**, which can be accessed directly by anyone through data portals such as *Data.gov.uk* and the Health and Social Care Information Centre (HSCIC) found at *Hscic.gov.uk*. Such data is focused on financial and statistical data for individual NHS departments and hospitals, trusts, and CCGs, and is available in .CSV format. Open data provided through the HSCIC includes information such as prescription data, the number of patients registered at GP practices, and audits of chronic disease prevalence such as cancer and diabetes. Open data is all data published under the transparency agenda, and the Open Government License.¹³³

We also described a category of data called “**restricted**” data. For instance, certain forms of patient-centred data are offered through the NHS Choices website and other NHS sites, but is not provided in fully downloadable format. Restricted health data is also frequently available to researchers and other organizations upon request, and upon receipt of a fee. The key difference between restricted and fully open data is that the latter is available in fully-downloadable format, whereas the former must be requested through *HSCIC.gov.uk*.¹³⁴

For the purposes of the conceptual framework, we include only open data. However, when examining the impact of open data, there exist several other conditions and standards that determine what kinds of data can be included within our framework:

1) Formats: In order to be considered open, data should be published in a standard format, without proprietary conditions, be machine readable, fully downloadable and available online.

2) Legal: Open health data should conform to the European Union Data Protection Directive 1998 (DPA), the Public Records Act 1958, the Access to Health Records Act 1990, and the Access to Medical Reports Act 1988.¹³⁵

¹³⁶ Since not all open health data concerns medical records, such as financial records of NHS bodies, the Freedom of Information Act 2000 that created a public “right of access” to information held by public authorities is also relevant. These acts require the anonymization of most data, and, failing that, the setting of privacy safeguards for individuals, and the protection of identifiable data from disclosure to third parties. Records collected by the NHS will be anonymized by removing information such as names, and through a process of de-identification.¹³⁷

133 HSCIC, “Supporting Transparency and Open Data,” Health & Social Care Information Centre, accessed February 23, 2014, <http://www.hscic.gov.uk/transparency>.

134 HSCIC, “Data Linkage and Extract Service Applications, Approvals, and Charges,” accessed February 23, 2014, HSCIC.Gov.UK, Health & Social Care Information Centre, <http://www.hscic.gov.uk/dlesaac>.

135 Information Commissioner’s Office “Draft Anonymisation code of practice,” May 2012, http://ico.org.uk/news/latest_news/2012/~/_/media/documents/library/Corporate/Research_and_reports/anonymisation_cop_draft_consultation.ashx.

136 HSCIC, “NHS Records Management: Guidance for Access to Health Records Requests,” HSCIC.Gov.UK, accessed February 23, 2014, <http://systems.hscic.gov.uk/infogov/records>.

137 NHS Choices, “Your Records: Better Information Means Better Care,” accessed February 23, 2014, <http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/care-data.aspx>.

Though patients must provide consent to the collection of their data under the acts, they can also choose to “opt-out” of data collection at a later point, further increasing their opportunities to protect their privacy.¹³⁸

TYOLOGY

A number of different types of open and restricted data are held and made available by the NHS. Different types of data will determine what kinds of activities and outcomes are possible. Therefore, when assessing the impact of open data, researchers should pay close attention to the type of data under consideration. Broadly, two different typologies can be identified:

TYOLOGY 1: Internal Data

Following a survey of the open data available at the time of our analysis (on Data.gov.uk, HSCIC, NHS Choices, and the regional and local sites), we developed the following typology of “internal” data. Such data is less likely to be accessed by patients than by researchers, agencies, watchdog organizations or groups seeking to assess or audit the performance of the NHS. In this typology, we can include the following forms of data:

- 1) **Financial Data:** Financial data on NHS departments, regional trusts and CCGs, much of which can be found on *Data.gov.uk*. This data normally includes information on costs and spending, and is held in .CSV format.
- 2) **Administrative Data:** Administrative data is available on NHS departments through *Data.gov.uk*, and is available in .CSV format.
- 3) **Statistical/Diagnostic Data:** The NHS frequently publishes statistics on chronic disease rates, preventable infections, and a wide range of other topics. Open health data, for instance, held in HSCIC.gov.uk’s archives (NHS Information Centre) includes statistics on a variety of issues including data quality, hospital care, illnesses and conditions, mental health, patient experience, prescribing, primary care services, public health, social care, workforce data.
- 4) **Audits:** The majority of data published under the Open Government License takes the form of audits on NHS departments and special divisions. In a transparency and open data letter to the Cabinet Ministers in July of 2011, Prime Minister David Cameron committed to making audit data available through the National Clinical Audit and Patient Outcomes Programme.¹³⁹ These audits engage healthcare professionals across England and Wales in the systematic evaluation of their clinical practice standards, to support improvement in the quality of treatment and care.¹⁴⁰

138 “How to opt-out of care.data,” *Optout.care.data*, accessed March 8, 2014, <http://optout.care-data.info/>.

139 HSCIC, “National Bowel Cancer Audit, Open Data – 2011-2012,” accessed February 21, 2014, <http://www.hscic.gov.uk/searchcatalogue?productid=12273&infotype=0%2fOpen+data&sort=Relevance&size=10&page=2#top>.

140 Healthcare Quality Improvement Partnership, “National Clinical Audits (Including NCAPOP),” accessed February 21, 2014, <http://www.hqip.org.uk/national-clinical-audits-including-ncapop/>.

TYOLOGY 2: STREAMS OF HEALTH DATA

Open health data includes information about a wide-range of players within the NHS system. Therefore, certain types of open health data are more focused on some groups than others. For example, audits place an emphasis on the quality of treatment and care provided by healthcare practitioners to their patients in cases of chronic diseases such as cancer and diabetes. To understand the impact of open health data, it is important to know the target of the data set that is being opened up by the NHS.

- ▶ **Open Health Data on Practitioners:** Examples include doctors, administrators, midwives, nurses, etc.
- ▶ **Open Health Data on the System/Infrastructure:** Examples include hospitals, clinics, equipment, and waiting times, etc.
- ▶ **Open Health Data on Patients:** Examples include diagnoses, prescribing rates, mortality rates, treatment options, etc.

USERS



III.3. Users: Who is using the data?

What is their relationship to the NHS?

When assessing the impact of open data initiatives, researchers should identify both the user (or category of user, such as an organization), and its relationship to the NHS. Different users of data may have different criteria for determining the success of an activity or programme. In addition, different users – whether they are internal or external - may value some outputs over others, or may be solely interested in the impact or outcome of an open data initiative on their particular interest or sector. For this reason, it is also important that users themselves identify their particular areas of interest, and help determine their own set of indicators upon which to judge success.

This report separates users into two broad categories, depending on their relationship to the NHS system:

(1) Internal Users

Internal users are those who work within the health system, including healthcare providers, such as physicians, nurses, administrators, and other healthcare professionals, health funds (Trusts/CCGs), and other government agencies, such as the UK Department of Health and Ministry of Finance.

(2) External Users

External users are those who interact with open data from a position that is partly or completely external to the NHS. This includes patients and members of the public, external supervisory and regulatory agencies (such as the Royal Society for Public Health), watchdog organizations such as media and citizen interest groups, businesses and commercial interest groups, academic researchers, and policy scholars.¹⁴¹

141 Royal Society for Public Health, "NGO Forum Launches New Planning Guide," November 5, 2009, accessed February 23, 2014, <http://www.rsph.org.uk/en/about-us/policy-and-projects/projects/ngo-forum-separate-cfm>.

ACTIVITIES



III.4. Activities: How is data used?

Ultimately, how and for what purposes data is used will determine the impact of an open data initiative. As illustrated by the logic model, users use open data and restricted data, possibly in conjunction with or independently of privately held data, and this use leads to specific outputs or results. Each unique activity (use) leads to its own measurable output indicator.

Open data will be used for different purposes, depending on the position and intention of the user accessing the information. For instance, internal and external users might use the same data set for a different set of activities. An internal user (such as a NHS administrator or executive) could use open data for the purposes of benchmarking, hotspotting, or to compare performance against a standard in order to improve the quality of care. A user external to the NHS (such as a watchdog organization) could use the same data to compare actual health outcomes against the standards proposed by the NHS or the government. The primary uses of open data will be used to improve management practices within the NHS, as well as to improve policies geared towards public services. Though the activity may appear similar, the impact of the use is likely to be different. As open data is collected on an ongoing basis, the same dataset can be used to address a variety of problems, dependent on the needs of the researcher or organization. The internal user intends to improve efficiency within the NHS; the external user seeks to increase accountability.

A number of other possible uses for data exist. A more complete list can be found in Appendix VI. For our discussion here, the important point is that researchers seeking to understand the effectiveness of an open data initiative should always consider the use (or activity) for which data is accessed. Understanding use can help determine impact, and ultimately will guide our assessment of the success or failure of an initiative.

OUTPUTS



III.5. Outputs: What was the value produced by the open data initiative (activity)?

Outputs are basic metrics of success in quantitative terms, such as the number of additional patients served, or the number of pounds saved. Outputs themselves are basic indicators of a sort, though they are tied very closely with the activity, or open data initiative. More high-level indicators are also needed to measure how well an initiative achieves the long-term goals of the NHS.

There are a number of examples of potential outputs. Referring to the example given in the **Activities** section (III.4), an output for an external user engaging in benchmarking might be the difference between a desired outcome and a measurement of the actual outcome. For instance, an administrator using open data to assess the rate of preventable in-hospital disease would consider the output to be the difference between the expected rate of preventable disease and the actual reported result. More examples of outputs are listed in the **Appendix VII**.

INDICATORS



III.6. Indicators: To what end?

The selection of indicators and metrics is a value-based assessment based on the desired outcomes of the NHS. A metric can be thought of as a numerical unit of measurement, such as number or percent of increase or decrease, while an indicator is a metric that is tied to one or more targets, such as an outcome like rates of premature mortality. Open data initiatives target both the impact of open health data on metrics and indicators, as well as the role of open data in supporting the values and guidelines outlined here for the use of open data. To this end, measures and indicators are assessed with an eye towards how open data can help the NHS achieve the values of *accountability, efficiency, choices, outcomes, customer service, and innovation and economic growth*.

Indicators build on outputs, or basic measures of success for any given activity undertaken through the use of open health data. They are generally short-term indicators that can determine whether an activity led to a desired result. More high-level indicators make use of these outputs to determine whether the activities led to any long-term changes within the NHS system.

An indicator is therefore the result of an output. For instance, an indicator can be a decrease in premature patient mortality, a decrease in the average hospital stay, or a decrease in hospital spending (the output would be the absolute number or percentage pertaining to mortality, hospital stays, and spending). Indicators are essential for researchers seeking to assess the impact of open health data—they provide information on success or failure, and help us understand the extent to which any programme has achieved its stated goals. More examples of indicators can be found in **Appendix VIII**.

METHODOLOGIES



III.7. Methodologies to Measure Impact: According to what measure?

Once indicators have been established, data can be collected using a number of different methodologies. These include, for instance, direct observation, questionnaires, document review, focus groups, interviews, and an analysis of administrative data. Ultimately the method used will depend on the specific programme, and on the desired outcome.

THREE CATEGORIES OF MEASUREMENT:

Traditionally, there are three broad categories of research design to evaluate societal intervention programmes:

- 1) *Experimental designs*: Researchers randomly assign one group of people to be participants and another group to be non-participants, and compare results between the two groups to establish the effect of the programme. Experimental designs are commonly referred to as randomized controlled trials (RCTs) and are considered the most rigorous research methodology.

There are many examples of potential experimental designs. For example, researchers might be interested in finding out whether an awareness of hospital waiting times affects whether or not patients visit a hospital in their locality. After obtaining a list of people who wished to take part in the programme, they could randomly assign the participants to two groups. The experimental group would differ from the control group *only* in that they had access to information about hospital wait times. Researchers would then observe whether the two groups differed in their visits to local hospitals, and the results would provide an indication about the impact and potential usefulness of releasing data about hospital wait times. Similar experiments can be designed for a range of other situations, all designed to measure differences in behavior, impact or outcomes as a result of access to open data.

2) *Quasi-experimental* designs may use groups of “participants” who have not officially signed up for an experiment or programme, or may consider a single group, measuring the effects before and after that group has access to data. Quasi-experimental designs may also differ from experimental designs in that participants in the programme and comparison group are *not* randomly assigned; rather, a comparison group may be selected after a programme group has been established. This technique is less rigorous because there is no certainty that the two groups are comparable across all characteristics. For example, those who choose to sign up for a waiting times application may be more technologically savvy, and may have other obligations, such as inflexible work schedules or childcare, that would at least partially determine their actions within the programme.

3) *Non-experimental* techniques, considered the least rigorous method for measuring programme success, may involve a single survey delivered after a programme, in order to gain descriptive information.¹⁴² Non-experimental techniques are less effective in establishing causality. For example, researchers using a design survey evaluating the effectiveness of open data internally within the NHS will not be able to conclusively determine whether an increased efficiency resulted from the open data, or from other organizational practices.

MIXED METHODS APPROACH

To study the impact of open data, it may be best to employ a combination of the three approaches described above. Frequently used in the social sciences, this combination is referred to as a *mixed methods* approach, a mix of qualitative and quantitative research design.¹⁴³ This approach can be taken when there are a number of data streams available, each with their own limitations. In the case of open data, a quantitative variable could be data download rates. When combined with qualitative data, such as a user survey held on the site with the relevant dataset, researchers seeking to evaluate the effectiveness of a particular dataset can learn more from users about the purpose of use for the data, the effectiveness of the presentation of data, and other information dependent upon the research questions being posed. Given that most research studying the impact of open data will be conducted in a complex social environment, rather than an enclosed controlled setting, a mixed method approach can be used to increase both the validity and reliability of the data being collected.

There are four major types of mixed methods design:

1) **Triangulation:** Merging qualitative and quantitative methods, such as in the example provided above.

2) **Embedded:** Using either quantitative or qualitative methods to answer a research question within a largely quantitative or qualitative study. For instance, this can include the use of interview techniques to make sense of a portion of quantitative data, to gain a deeper understanding of quantitative indicator that was being measured.

142 William Trochim, “Research Methods Knowledge Base,” *Social Research Methods*, accessed March 8, 2014, <http://www.socialresearchmethods.net/kb/>.

143 Viswanath Venkatesh, Susan Brown, and Hillool Bala. “Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems.” *MIS Quarterly* 37 (2013), accessed April 4, 2014.

- 3) **Explanatory:** Using qualitative data to help explain or elaborate upon quantitative results.
- 4) **Exploratory:** Using quantitative data to test or to explain a relationship that has been found through qualitative methods.

Possible Tools: A variety of tools exist to conduct all three forms of research-- experimental, quasi-experimental, and non-experimental that can be used in a mixed methods approach. A non-exhaustive list includes the following:

A) Data Collection Methodologies:

1) **Surveys/Questionnaires:** Both a quantitative and a qualitative measure that can be used to determine outcomes such as patient satisfaction, use of data to determine treatment options, and choice in services. They can include both close-ended questions or open-ended questions that can be transformed into quantitative data. Surveys can also be used to assess the effectiveness of an open data programme at the level of healthcare administrators and professionals.

2) **Site Analytics and Download Rates:** Quantitative measures that can be used to determine the effectiveness of open data. Part of the impact of open data can be determined solely through the number and types of individuals who access it.

3) **Other Numerical Data:** Insight can be gained from budgetary information from before or after the introduction of an open data programme, or other performance measures such as items produced, patients seen, wait times, prescribing rates, diagnoses, and other relevant measures.

4) **Content Analysis:** Researchers examining the impact of open health data through the frequency of citations should score the reports, applications, and entrepreneurial activities in relation to the six Open Data value propositions outlined above (Part II). The more mentions of accountability, efficiency, choices, outcomes, customer service/patient satisfaction, and innovation/economic growth, the higher the score. Special attention should be paid to evidence suggesting that open data had real-world impact and was used to achieve one of these value propositions.

a) **Number of media and academic mentions of open data:** The impact of open data can be partially determined by the extent to which it is used or referred to in academic and media reports.¹⁴⁴

b) **Number of Open Data Sets Mentioned in Grant Applications and Policy Proposals:** The impact of open health data can also be seen in the number of citations in grant applications and policy proposals, especially those that could lead to change in government transparency and accountability, social equality, healthcare reform, and other similar initiatives.

c) **Number of Open Data Sets Used in Entrepreneurial and Business Activities:** In addition, the impact of open data can be assessed through its usage in the development of new products, services, and applications.

B) Qualitative Methodologies:

1) **Interviews:** Can be either structured or semi-structured (or a combination of the two), and can be used to gain a deeper understanding of the view of the participants of an open data programme, or users of open data.

2) **Observation/Focus Groups:** To evaluate the effectiveness of open data programmes, it may be useful for researchers to enter into departments or organizations using open data to observe participants in their

144 Tim Davies, "Open Data Barometer: 2013 Global Report," Open Data Institute and World Wide Web Foundation, 2013, <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>.

use of open data, to evaluate its use and effectiveness in context.

3) **Interpretive Methods:** There are a variety of interpretive methods that can be used by researchers to gain a deeper understanding of the impact of open data. Included within this (but not limited to) are critical studies, legal interpretive analysis, historical analysis, and policy analysis.

C) Other Methodologies:

1) **Design Science:** A method normally employed in engineering and computer science, design science is increasingly being integrated into information systems research. Researchers use the act of design (such as a web application) as an opportunity for theory-testing and building.¹⁴⁵ Researchers identify a problem, define their objectives for a solution, design a product or program to address the problem and lead to the solution, and then evaluate the program in use, noting the observed results from those that were intended. This method should be used in conjunction with many of the data collection or qualitative methods described above.

2) **Computational Social Science:** A burgeoning interdisciplinary field that studies complex social systems through computational modeling and related techniques.¹⁴⁶ Computational approaches can include (but are not limited to) social network analysis, social geospatial modelling (GIS) methods, and automated content analysis.

It should be stated that open data is itself changing the way research is conducted, leading to new methodologies and new forms and sources of data. Traditionally, social sciences research has relied on data that is collected one time, which is often self-reported by research participants through the use of surveys and interviews.¹⁴⁷ Data has expanded opportunities for both quantitative and qualitative data in the social sciences, offering access to content-related information such as anonymized emails, public forums, and search records, as well as quantitative information that can enable social scientists to analyze macro trends. In addition, greater use of data and new technologies increases the potential for data collected over a period of time, possibly as the result of a series of interactions with the researcher, thus making it easier to make comparisons across weeks and months and even years. In these ways and many others, open data is pushing at the boundaries of research methodologies; this process has just begun, and is likely to lead to many new and innovative methodologies.

Recommendations for Building Metrics-into-Design

To increase the ease and effectiveness of research programmes studying the impact of open data, it may be useful to consider what metrics or incentives can be built into the design of sites housing open data. In light of the tools listed above, the following recommendations can be made:

- ▶ **Standardize the citation of your open dataset.** Ask that users of open data cite the dataset in a particular way that will enable researchers to find examples of its use. Though normally employed only in academic research, encourage users of open data, such as regulators, or entrepreneurs to be transparent about their use of open data on their website, including a link back to the open dataset whenever possible.
- ▶ **Build surveys into the download design, and incentivize users to fill these out before downloading.** Though it is not feasible to require the completion of a survey prior to the downloading of data, government agencies can incentivize survey completion. One way this can be done is through offering users of open data an email alert to notify them when new data becomes available.

145 Ken Peffers, Tuure Tuunanen, Marchus Rothenberger, and Samir Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* 24(3) (2008): 45-77.

146 Krasnow Institute for Advanced Study, "About: Computational Social Science," George Mason University, accessed April 4, 2014, <http://www.css.gmu.edu/node/2>.

147 David Lazer, et al. "Computational Social Science." *Science* 326 (2009):721-723, <http://gking.harvard.edu/files/gking/files/LazPenAda09.pdf>.

- ▶ **Encourage open data users to be transparent about their research method to facilitate reproducibility.** Ask that users of open data clearly report their research method when using open datasets. Include on your site a set of guidelines for the reporting of research design when publishing their results, and emphasize that open data users should be open about the dataset they have used, the analytic techniques they have employed, and the other methodologies they have used to triangulate their research.

Can randomized controlled trials (RCTs) be used to evaluate policy?

A recent report published by the UK Cabinet Office's Behavioural Insights Team advocates for using an old methodology to reinvigorate the testing of public policy interventions.¹⁴⁸ Long established as a tool in science and medicine, the authors of the report argue that randomized control trials (RCTs) should also be used to test the effectiveness of policies implemented in the UK.

According to the report, new policy interventions should be evaluated using evidence from RCTs to determine whether or how a policy or intervention is working within a community. The RCTs would compare the intervention with control groups that were not exposed to the same condition, using this comparison to eliminate potential bias or confounding variables that could influence policy effectiveness (such as general economic conditions at the time of implementation).

Using RCTs in policy programmes enables researchers to develop counterfactual conditions, comparing old and new policies, and gaining insight into what might occur if policy interventions are implemented. These conditions will help to determine the cost effectiveness of intervention programmes, and to assure citizens that public money is being spent on policies that have been proven to deliver value for the cost.

IMPACT



III.8. Impact: To add what value to the NHS and the UK as a whole?

The six value propositions outlined in II.1 should be used to guide discussion and evaluation regarding the impact of any open data programme. Occasionally, an open data programme can achieve multiple impacts. For instance, a programme using comparative data across hospitals to increase accountability may also increase efficiency. Researchers examining the impact of any open data initiative should consider whether the initiative achieved one or more of the following six aims:

- 1) **Accountability:** Opening comparative data on health outcomes will ensure that physicians and hospitals are held accountable for unwarranted variation (for example, in the use of certain medical procedures) shown in the data.
 - **Political:** Improved transparency and accountability in terms of reaching desired outcomes for public health.
 - **Economic:** Improved accountability and transparency in spending, cost, waste, and fraud.
 - **Participatory/Citizen Involvement:** Increasing public participation in healthcare services.
 - **Organizational:** Internal NHS objectives are met with regard to patient outcomes.

148 Laura Haynes, Owain Service, Ben Goldacre and David Torgerson. "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials," Cabinet Office: Behavioural Insights Team, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf

2) **Efficiency:** Connected information for integrated care as well as electronic records will accelerate data-sharing practices and reduce unnecessary duplications in care.

- **Operational:** Improving the speed at which patients are processed, treated, and discharged.
- **Economic:** Decreasing the costs associated with each hospital or service by comparing open data across regional and local contexts.
- **Communication and Technology:** Identifying spots in the healthcare system where communication needs to be improved both internal to the system, and external in terms of reaching the broader public.
- **Treatment:** Determining which treatments work effectively, and for which populations.
- **Resource allocation:** Determining where and when services are needed most.
- **Outcomes:** Determining whether an open data initiative has had an impact on outcomes, and particularly whether this impact is in line with expected or desired outcome goals set within the NHS.

3) **Choice:** Open data on patient experience and health outcomes will allow patients to seek out the services that best meet their individual healthcare needs. Furthermore, increased access to health data will allow patients to choose between providers and treatments, both of which will drive patient satisfaction and quality improvements.

- **Variety:** Increasing the amount of information available to patients for their decision-making.
- **Quality:** Increasing the quality, completeness, and timeliness of information provided to patients and the broader public to determine whether treatment is needed and/or available.

4) **Outcomes:** Publishing “outcomes data” drives competition between health-care professionals, which in turn can lead to quality improvements and innovation. Opening health data to analysts and researchers improves both the quality and consistency of evaluation.

- **Health outcomes:** Assessing whether the entire NHS, or its individual hospitals and regions, have achieved NHS health outcome measures.
- **Quality of treatments:** Increasing the quality of treatment while achieving health outcomes.
- **Efficiency:** Enabling hospitals and regions to compare their services and outcomes, to identify weak spots within their individual systems.
- **Resource allocation:** Identifying high-risk patients, directing resources to high-need communities, in order to improve overall outcomes.

5) **Customer service:** Improving patient access to data and providing convenient customer-oriented services such as online appointment booking and prescription refills will enable the NHS to achieve world-class customer service and improve efficiency.

- **Quality:** Ensuring positive experiences of care between patient and physician, and between the patient and the general healthcare system.
- **Quantity of patients served:** Decreasing wait times, and increasing the number of patients served through emergency services, general practitioners and specialists. Shorter wait-times for GPs and specialists

6) Innovation and Economic Growth

- **Quantity:** Increasing number of services provided, and increasing competition between different businesses and innovators.
- **Variety:** Increasing the variability of new services and applications.
- **Labour:** Increasing the number of individuals employed through the development of new services and applications.
- **Economic:** Increasing the amount of financing given to private sector initiatives that can benefit the healthcare system.
- **Research:** Ensuring that researchers are made aware of available data, and that it can contribute to their research programmes effectively.



Part IV

Building an Open Data Learning Environment (ODLE)

The preceding discussion has outlined the key value propositions for the implementation of open data as well as a conceptual framework that can help achieve (and measure) those value propositions. For the moment, the conceptual framework remains just that—a concept or theory. Future research and variations on this paper can help bridge the gap between theory and practice, and build on the conceptual framework to develop a series of specific recommendations and real-world activities that can help the NHS achieve its goals.

As a starting point towards this effort, we offer one specific recommendation in this section. We suggest that the NHS create an Open Data Learning Environment (ODLE), both to explore and develop potential applications of open health data, and to assess how effectively it is currently being used. As much as the NHS might intend or desire to open all its data, there do of course exist practical constraints. An ODLE can help prioritize which data to make open, and in what way. It serves as a foundation for future open data initiatives, guiding the NHS in its ongoing efforts to achieve greater transparency and the six value propositions outlined in this paper.

In what follows we propose ten steps that can help in building such a learning environment. A variety of ways exist for each of these steps to be implemented; here, we outline a few pathways towards implementation.

1) BUILD AN OPEN DATA LEARNING CAPACITY AND CULTURE WITHIN THE NHS.

To create a deeper understanding of the current and potential uses of open data, the NHS will need a commitment from leadership to institutionalize action and learning across its various open data projects and initiatives. Such a broad, comparative view will make it easier to determine what works and what doesn't, and to use cross-programme insights to build a more effective institutional strategy. Importantly, each individual programme should be designed with this broader view in mind. Mechanisms to learn from individual programmes should be architected into the design of those programmes. Measuring impact should not be an afterthought; it must be considered up front at the design stage of every project.

Possible Pathways:

- ▶ Develop an NHS Open Data “Do and Learn Tank.” Unlike a traditional think tank, this would be structured to help launch new projects using lessons from existing programmes and initiatives and evaluate their results.
- ▶ Set up an NHS Data Geek Squad to create a corps of volunteer data geeks and researchers from Britain's best universities to work with open health data. This could be modeled on Datakind in the U.S., or on Code for America (although Code for America focuses its work city by city rather than by datasets).
- ▶ Connect research organizations already existing within the NHS, such as the NIHR.ac.uk, that are already examining the impact of open data.¹⁴⁹
- ▶ Use prizes to stimulate new solutions to public problems using open health data, as the UK and U.S. governments have already begun doing.
- ▶ Develop campaigns designed to promote data sharing, and raise awareness about the release of open datasets, their locations online, and their potential uses by the public.
- ▶ Set up an open health data academy that trains people to use open health data and measure its impact, using online learning, project-focused instruction, and mentoring.

149 NIHR, “Research,” NHS: National Institute for Health Research, accessed March 18, 2014, <http://www.nihr.ac.uk/research/Pages/default.aspx>.

- ▶ Fund the creation of NHS open health data fellowships for students and graduates with compelling ideas and practical ways to implement them.
- ▶ Develop an open health data mentor network to encourage and train new recruits and younger members of the NHS in the use of open data.

2) ENGAGE THE PUBLIC IN DEFINING METRICS

Engage the public in defining metrics of success with regard to open data. The public also has valuable insights and lessons to contribute to this proposed ODLE. In addition, since all indicators and uses of data are inextricably linked with priorities and values—i.e., we measure what we care about--open data policies should reflect a larger social and political perspective. The NHS should therefore ensure that direct stakeholders (e.g., patients and providers) are included in decisions about open data and how it is used.

Possible Pathways:

- ▶ Develop a citizens' open health data panel (similar to Hackney's online citizen panel) to review metrics on a regular basis.
- ▶ Facilitate user-led design exercises to better understand how open data can support stakeholders' work and ultimately improve people's lives.
- ▶ Design and implement online mechanisms such as ratings and feedback tools to gauge public opinion and solicit insights from citizens.
- ▶ Create a "Metrics Bank" of important indicators, with input from stakeholders, researchers, and experts in the field who have studied the relevant literature.
- ▶ Prioritize the creation of an inventory of datasets by departments within the NHS, and invite experts to curate the datasets released to the public.

3) STAY FOCUSED ON WHAT REALLY MATTERS.

Too often, indicators used to measure open data only quantify the level of use (e.g., the number of datasets used or money saved). The metrics that really matter, however, are those that are focused on improving people's health and lives. The NHS must not lose sight of these truly important metrics.

Possible Pathways:

- ▶ Develop NHS Open Data Stories that will allow stakeholders to share how open data improved people's lives in real time.
- ▶ Use surveys, social media, and sentiment analysis to learn what dimensions of healthcare improvement are most important to the public, and ensure that success metrics and indicators capture those priorities.

4) DEVELOP A COMMON ASSESSMENT FRAMEWORK

Develop a common assessment framework with partners in other sectors and other countries. As the NHS is developing a research agenda to measure the impact of open data, it is important to coordinate with other organizations, sectors and countries that are also designing analytical frameworks for the use of open data across sectors. This coordination is an essential part of developing an ODLE. Among other things, it will require participants to agree on definitions of key terms in the sector; make assessment methods comparable and context sensitive; assess and compare outputs, outcomes and impacts; and above all, limit duplication and pool efforts to reduce costs and generally enhance efficiency.

Possible Pathways:

- ▶ Set up an annual meeting/listserv/monthly hangouts on open health data research to trade best practices and ideas.
- ▶ Create a directory (perhaps in wiki format) of other assessment frameworks across countries and sectors. Such a directory would also include a list of key contacts and organizations.
- ▶ Use online and offline meet-ups and other approaches to create a culture that encourages knowledge sharing and collaboration with other organizations.

5) STAY FLEXIBLE AND ADAPTIVE IN MEASURING IMPACT.

As the use of open data evolves and becomes more dispersed, the NHS will need to find a balance between a centralized measurement function (which sets measurement standards, aggregates data across programmes, and monitors data quality) and a more diffused structure that empowers users at different levels of the organization. Recent studies have shown that a mix of methodologies—e.g., quantitative, qualitative, randomized control trials, and other methods--is needed to evaluate new programmes that provide public goods. Ultimately, we must adopt an approach that does not rely on a single method but that selects the best of various methods.

Possible Pathways:

- ▶ Hold regular “What Works Camps” that connect various users and researchers.
- ▶ Ensure that all metrics (especially those to measure use and impact) are categorized by different levels and can be analyzed separately.
- ▶ Continue ongoing research into new and existing approaches to impact measurement. This research effort should be considered a core part of the NHS open data programme.

6) SHARE WHAT IS LEARNED (INCLUDING FAILURES) WITH EVERYONE.

The best way to learn and improve is from the NHS' own efforts to work with open data. Failures and mistakes are in this regard as important as successes. By sharing what is learned on a regular basis, the NHS can ultimately enhance the way open data is used.

Possible Pathways:

- ▶ Develop visualizations of how open data has made an impact--for example, through maps that show changes in healthcare quality, efficiency, or cost in different geographical areas.
- ▶ Develop an Open Health Data 10 – a listing of the ten most impactful uses of open health data.
- ▶ Consider wikis, seminars and other means for stakeholders within the NHS to share stories and experiences with open data.

7) BUILD A RESEARCH NETWORK AND EXPERT NETWORK

Build a research network and expert network to build capacity quickly and broker debate around methodology and findings. Understanding the best knowledge and thinking in the field will be key in all stages of the research cycle to assess the impact of NHS open data.

Possible Pathways:

- ▶ Hold an annual summit to exchange what has been learned, along with regular meet-ups.
- ▶ Set up an NHS Open Data listserv.
- ▶ Consider establishing an external advisory or consulting board of experts to whom the NHS can turn for advice and guidance.

8) DEVELOP OPEN HEALTH DATA ECOLOGY MAP

Develop open health data ecology map, including a dictionary of all open health data sets in the NHS along with the variety of uses and users. This map should include visualizations and other graphical representations to fully represent how open data is being accessed and used within the NHS.

Possible Pathways:

- ▶ Develop a conceptual map of users, uses and impact.
- ▶ Develop a site inventorying all available datasets within the NHS, making it easy for users to access datasets from a variety of online spaces and departments.
- ▶ Develop a crowdsourced resource of innovative start-ups and new applications that use open health data.

9) PUBLISH, INTEGRATE AND FINE-TUNE THE OPEN DATA CONCEPTUAL FRAMEWORK

Publish, integrate and fine-tune the open data conceptual framework for feedback and improvement. The conceptual framework presented in this report is an attempt to guide the systematic review and study of the open data value chain. A framework, as we have noted, represents just a concept or theory. In order to bridge the gap between theory and practice, it is essential that various stakeholders participate in refining the framework. The framework must itself be responsive and flexible; its design must constantly be refined and implemented using the same level of transparency as embodied by the concept of open data.

Possible Pathways:

- ▶ Develop an interactive version of the conceptual framework that can be annotated.
- ▶ Create an expert, online advisory network to vet and review the conceptual framework.
- ▶ Create channels for feedback and review by various stakeholders.
- ▶ Research and evaluate similar frameworks used in other sectors or countries and build on insights or lessons learned.
- ▶ Build on the Memorandum of Understanding between the United States' HHS, and the NHS, and include impact assessment as a joint activity undertaken by these two countries.¹⁵⁰

10) ENGAGE STAKEHOLDERS IN SHAPING THE OPEN HEALTH DATA PROGRAMME

While the NHS' release of open health data will benefit both public and private-sector entities, those benefits can be enhanced if the beneficiary entities can help participate in shaping the government's open data programme. The NHS should engage key stakeholders on a regular basis to determine which datasets have the highest priority for them; what new datasets should be released as open data; and which open data collections are particularly easy or difficult to use.

Possible Pathways:

- ▶ Set up a wiki, forum, or combination of online tools for stakeholders to provide this feedback.
- ▶ Develop a subcommittee of the Open Data User Group to focus on health data specifically.
- ▶ Hold roundtables with different groups of stakeholders--health-related businesses, advocacy groups, and patient groups--to help shape government policy on the release of open health data.

¹⁵⁰ "A Memorandum of Understanding Between The US Department of Health and Human Services, The Secretary of State for Health, NHS England, and the Health and Social Care Information Centre," January 1, 2014, http://www.healthit.gov/sites/default/files/hhsnhs_mou_final_jan_21.pdf.



Part V

Conclusion

The Open Data Era in Health and Social Care

As the NHS continues its ambitious plans to release unprecedented amounts of data, it is clear that open data can have a powerful, transformative effect on how healthcare is delivered and consumed. This paper has outlined six key Value Propositions for data transparency and openness in a healthcare setting:

- ▶ **Accountability:** Using data to hold healthcare organizations and providers accountable for treatment outcomes.
- ▶ **Choice:** Providing open data to help patients make informed choices from among the healthcare options available to them.
- ▶ **Efficiency:** Improving the efficiency and cost-effectiveness of healthcare delivery.
- ▶ **Outcomes:** Improving treatment outcomes by using open data to make the results of different treatments, healthcare organizations, and providers' work more transparent.
- ▶ **Customer service:** Using open data to educate patients and their families and make healthcare institutions more responsive.
- ▶ **Economic growth and innovation:** Using open data to fuel new healthcare companies and initiatives and to spur innovation.

As leaders in the open data movement, the NHS and more generally the United Kingdom are well-positioned to reap the benefits of these Value Propositions. These benefits can have powerful ripple effects on the broader economy. Yet as we have also argued, there are important challenges to be overcome, and choices to be made. The benefits and costs of open data must be weighed against those of other programs. Priorities must be established, especially at a time of constrained budgets. To borrow a phrase from the open source movement, open data is free not in the sense of free beer, but of air. Huge sums of money are involved in transitioning from a paper-based system to an open data universe with real time feeds. Cost-effective and efficient means of implementation need to be developed.

Implementation is critical, one of the most important challenges confronting the NHS. It is not enough just to know that data transparency has tremendous positive potential. We must also know how to maximize that potential, and how to overcome (or at least minimize) the risks that accompany it. Without a roadmap for implementation, the NHS—and the wider public—risks drowning in the “tsunami of data” that the government has promised.

Therefore, one of the central goals of this paper has been, precisely, to develop a roadmap for implementation of open data policies. We have outlined a conceptual framework, or logic model, that can help in this process. It breaks down the open data chain into seven component pieces: inputs, users, activities, outputs, outcomes, impacts, and methodologies. These seven components allow us to better analyze risks and opportunities, and to understand where (and how) an open data initiative might succeed or fail.

Perhaps the most important contribution of a conceptual model is that it allows us to measure the impact of open data. For all the hype surrounding the field, it remains in fact nascent, dominated by unproven hypotheses and assumptions. The model presented here allows us to move beyond guesswork and into the realm of evidence-based policymaking. Importantly, the model allows us to continue gathering, analyzing and disseminating evidence about the *positive* impact of open data. The risks (e.g. to privacy) are real, and cannot be overlooked or minimized. But it is equally important, even as we acknowledge and seek to address these risks, that we continue to evaluate and better understand the tremendous positive potential of data transparency. Only such a positive agenda can help us design more effective open data policies.

Ultimately, a conceptual model remains just that—a concept, or theory, that will require further analysis and data in order to become fleshed out and more concrete. Towards that end, we have suggested that the NHS should foster and create what we call an Open Data Learning Environment (ODLE). An ODLE represents the first step towards operationalizing the conceptual model. If implemented, it can lead to a culture of learning and improvement across the NHS. It would encourage responsiveness, agility, and flexibility in the way the NHS designs and implements its open data initiatives. In a

word, an ODLE would help the NHS *learn* from its own failures and successes.

We submit this paper in the same spirit of learning, responsiveness and flexibility. What we have included here is just an initial proposal—the outlines of a model, the beginnings of a system for evaluation and measurement. Much work (much research, much data collection, much analysis) remains to be done, and this work can and should be used to refine the conceptual framework we have presented.

Part of this process of refinement also includes collecting and incorporating feedback from readers of this document. We very much look forward to your comments and suggestions. They will help improve our recommendations, and ultimately improve the way open data is collected, disseminated and used within the NHS.

APPENDICES

Appendix I: The GovLab Open Data Index

VALUE AND IMPACT

- ▶ Potential global value of open data estimated by McKinsey: **\$3 trillion annually.**¹⁵¹
- ▶ Potential yearly value for the United States: **\$1.1 trillion.**¹⁵²
- ▶ Europe: **\$900 billion.**¹⁵³
- ▶ Rest of the world: **\$1.7 trillion.**¹⁵⁴
- ▶ How much the value of open data is estimated to grow per year in the European Union: **7% annually.**¹⁵⁵
- ▶ Value of releasing UK's geospatial data as open data: **13 million pounds per year by 2016.**¹⁵⁶
- ▶ Estimated increase in GDP in England and Wales in 2008-2009 due to the adoption of geospatial information by local public services providers: **+£320m.**¹⁵⁷
- ▶ Estimated worth of business reuse of public sector data in Denmark in 2010: **more than €80 million a year.**¹⁵⁸
- ▶ Estimated worth of business reuse of public sector data across the European Union in 2010: **€27 billion a year.**¹⁵⁹
- ▶ Total direct and indirect economic gains from easier public sector information re-use across the whole European Union economy, as of May 2013: **€140 billion annually.**¹⁶⁰
- ▶ Economic value of publishing data on adult cardiac surgery in the U.K., as of May 2013: **£400 million.**¹⁶¹

151 Manyika et al, "Open Data: Unlocking Innovation."

152 Ibid.

153 Manyika et al, "Open Data: Unlocking Innovation."

154 Ibid.

155 Capgemini Consulting, "The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data," February 7, 2013, http://www.capgemini.com/resource-file-access/resource/pdf/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf.

156 "The Value of Geospatial Information to Local Public Service Delivery in England and Wales," Local Government Association Group, ConsultingWhere and ACIL Talisman, 2010, http://www.local.gov.uk/c/document_library/get_file?uuid=b6875678-4150-4d74-8b16-bdd9653f774d&groupId=10180.

157 Ibid.

158 Catherine Lippert. "Topic Report No. 20 Public Sector Information Reuse in Denmark 2010," European Public Sector Information Platform, 2010.

159 Deloitte, "Market Assessment of Public Sector Information," Deloitte, U.K. Department for Business innovation & Skills, May 2013.

160 Ibid.

161 Deloitte, "Market Assessment."

The Open Data Era in Health and Social Care

- ▶ Economic value of time saved for users of live data from the Transport for London apps, as of May 2013: **between £15 million and £58 million.**¹⁶²
- ▶ Average decrease in borrowing costs in sovereign bond markets for emerging market economies when implementing transparent practices (measured by accuracy and frequency according to IMF policies, across 23 countries from 1999-2002): **11%.**¹⁶³
- ▶ Open weather data supports an estimated **\$1.5 billion** in applications in the secondary insurance market – but much greater value comes from accurate weather predictions, which save the U.S. annually more than **\$30 billion.**¹⁶⁴
- ▶ Estimated value of GPS data: **\$90 billion.**¹⁶⁵

EFFORTS AND INVOLVEMENT

- ▶ Number of **U.S. based companies identified by the GovLab** that use government data in innovative ways: **500**
- ▶ Number of open data initiatives worldwide in 2009: **2.**¹⁶⁶
- ▶ Number of open data initiatives worldwide in 2013: **over 300.**¹⁶⁷
- ▶ Number of countries with open data portals: **more than 40.**¹⁶⁸
- ▶ Number of U.S. cities with Open Data Sites in 2013: **44.**¹⁶⁹
- ▶ U.S. states with open data initiatives: **40.**¹⁷⁰
- ▶ Number of cities globally that participated in 2013 International Open Data Hackathon Day: **102.**¹⁷¹
- ▶ Membership growth in the Open Government Partnership in two years: **from 8 to 59 countries,**¹⁷²
- ▶ Number of time series indicators (GDP, foreign direct investment, life expectancy, internet users, etc.) in the World Bank Open Data Catalog: **over 8,000.**¹⁷³
- ▶ How many of 77 countries surveyed by the Open Data Barometer have some form of Open Government Data Initiative: **over 55%.**¹⁷⁴

162 *Ibid.*

163 Rachel Glennerster and Yongseok Shin. "Does Transparency Pay?" *International Monetary Fund* 55(1) (2008).

164 "Value of a Weather Ready Nation." *National Weather Service and the National Oceanic and Atmospheric Administration*, October 13, 2011, <http://www.ppi.noaa.gov/wp-content/uploads/PPI-Weather-Econ-Stats-10-13-11.pdf>.

165 Philip Yam, "How to Kick-Start Innovation with Free Data," *Scientific American*, March 23, 2013, accessed August 29, 2013, <http://www.scientificamerican.com/article.cfm?id=how-to-kick-start-innovaton>.

166 Jose Alonso. "Announcing the Global Open Data Initiative," *World Wide Web Foundation*, June 11, 2013, accessed August, 29, 2013, <http://webfoundation.org/2013/06/announcing-the-global-open-data-initiative-godi/>.

167 *Ibid.*

168 "Open Data in the United States," *Data.Gov*, accessed December 12, 2013, <https://www.data.gov/open-gov/>.

169 *Ibid.*

170 "Open Data in the United States."

171 *Ibid.*

172 Alonso. "Announcing the Global Open Data Initiative."

173 "Data," *The World Bank*, accessed August 29, 2013, <http://data.worldbank.org/>.

174 "Open Data Barometer: 2013 Global Report," *ODI, OpenData Barometer, World Wide Web Foundation*, 2013.

The Open Data Era in Health and Social Care

- ▶ How many OGD initiatives have dedicated resources with senior level political backing: **over 25%**.¹⁷⁵
- ▶ How many countries are in the **Open Data Index**: **70**.¹⁷⁶
- ▶ How many of the 700 key datasets in the Index are open: **84**.
- ▶ Number of countries in the **Open Data Census**: 77.¹⁷⁷
 - ▶ How many of the 727 key datasets in the Census are open: **95**.
- ▶ Those who have machine-readable data available: **25%**.¹⁷⁸
- ▶ Top 5 countries in Open Data rankings: **United Kingdom, United States, Sweden, New Zealand, Norway**.¹⁷⁹
- ▶ The different levels of **Open Data Certificates** a data user or publisher can achieve “along the way to world-class open data”: **4 levels, Raw, Pilot, Standard and Expert**.¹⁸⁰
- ▶ The number of data ecosystems categories identified by the OECD: **3, data producers, infomediaries, and users**.¹⁸¹

EXAMINING DATASETS

- ▶ How many datasets have been made open by governments worldwide: **more than 1 million**.¹⁸²
- ▶ Number of datasets on the U.S. site **data.gov**: **more than 90,000**.¹⁸³
- ▶ How many released key datasets are truly open for re-use and can be used to hold government accountable, stimulate enterprise, and promote better social policy: **fewer than 1 in 10**.¹⁸⁴
- ▶ Percentage of datasets published in both machine-readable forms and under open licenses: **less than 7%**.¹⁸⁵
- ▶ How many countries surveyed have formal data policies in 2013: **55%**.¹⁸⁶
- ▶ Number of datasets on the Australian government’s open data website that were found to be unusable: **one-third**.¹⁸⁷
- ▶ Number of financial datasets in the World Bank Open Data Catalog: **over 850**.¹⁸⁸
- ▶ Out of 23 countries surveyed by Capgemini, those who share comprehensive data that includes both breadth and granularity: **22%**.¹⁸⁹

175 *Ibid.*

176 “**Open Data Index**,” *Open Knowledge Foundation*, Accessed December 2013.

177 “**Open Data Census**,” *Open Knowledge Foundation*, accessed December 2013, <https://index.okfn.org/>.

178 “**Open Data Census**.”

179 “**Open Data Census**.”

180 “**Open Data Certificate**,” *Open Data Institute*, accessed December 2013.

181 Barbara Ubaldi. “**Open Government: Data Towards Empirical Analysis of Open Government Data Initiatives**,” *OECD Working Papers on Public Governance No. 22*, OECD Publishing, 2013.

182 Jose Alonso. “**Announcing the Global Open Data Initiative**,” *World Wide Web Foundation*, June 11, 2013.

183 “**Open Government**,” *Data.gov*, accessed December 12, 2013.

184 “**Open Data Barometer: 2013 Global Report**,” ODI, *OpenData Barometer*, World Wide Web Foundation, 2013.

185 “**Open Data Barometer**.”

186 “**Open Data Barometer**.”

187 Allie Coyne. “**Govt finds one third of open data was ‘junk’**,” *ITnews*, Nov 12, 2013 accessed February 23, 2014, <http://www.itnews.com.au/News/363834,finance-finds-one-third-of-open-data-was-junk.aspx>.

188 “**Data**,” *The World Bank*, accessed August 29, 2013., <http://data.worldbank.org/>.

189 *Capgemini Consulting*, “**The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data**.”

- ▶ Those who lacked enhanced search capabilities: **over 60%**.¹⁹⁰
- ▶ Countries who share data that is not regularly updated: **96%**.¹⁹¹
- ▶ Those who are not utilizing user participation capabilities: **87%**.¹⁹²
- ▶ Average score of evidence of impact in 43 countries with some form of open data policy: **1.7 out of 10**.¹⁹³
- ▶ Percentage of impact questions for which no evidence could be found: **45%**.¹⁹⁴

Appendix II: Open Data Definitions

1. Open Definition

(referenced by *Open Data Handbook*, *ODI*, *Open Data Census*, and *OECD Open Data Analytical Framework*)¹⁹⁵

“Open data is data that can be freely used, reused and redistributed by anyone--subject only, at most, to the requirement to attribute and share-alike.”

2. The White House, 2013 OMB Memorandum¹⁹⁶

“Open data refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users.”

3. Data.Gov.UK¹⁹⁷

“Open data is data that is published in an open format, is machine readable and is published under a license that allows for free reuse.”

4. Dbpedia: A nucleus for a web of open data¹⁹⁸

“Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.”

5. Open Data Institute¹⁹⁹

“Open data is information that is available for anyone to use, for any purpose, at no cost. Open data has to have a license that says it is open data. Without a license, the data can't be reused. These principles for open data are described in detail in the Open Definition.”

190 Capgemini Consulting, “The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data.”

191 Capgemini Consulting, “The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data.”

192 Capgemini Consulting, “The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data.”

193 Capgemini Consulting, “The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data.”

194 Capgemini Consulting, “The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data.”

195 Open Knowledge Foundation, “Open Definition.” accessed February 23, 2014, <http://opendefinition.org/od>.

196 Sylvia Burwell et al. “Memorandum for the Heads of Executive Departments and Agencies,” *Open Data Policy-Managing Information as an Asset*, March 9, 2009, <http://www.whitehouse.gov/the-press-office/memorandum-heads-executive-departments-and-agencies-3-9-09>.

197 Data.Gov.UK, “A Simple Intro to Open Data,” April 11, 2013, accessed February 23, 2014, <http://data.gov.uk/blog/simple-intro-open-data>.

198 Auer Sören, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. “Dbpedia: A nucleus for a web of open data.” *Lecture Notes in Computer Science* 4825 (2007): 722-735.

199 Open Data Institute, “Guides: What is Open Data?,” accessed February 23, 2014, <http://theodi.org/guides/what-open-data>.

6. LinkedGov²⁰⁰

“Open data is non-personally identifiable data produced in the course of an organization’s ordinary business, which has been released under an unrestricted license. Open public data is underpinned by the philosophy that data generated or collected by organizations in the public sector should belong to the taxpayers, wherever financially feasible and where releasing it won’t violate any laws or rights to privacy (either for citizens or government staff).

7. McKinsey Global Institute²⁰¹

“Machine-readable information, particularly government data, that’s made available to others. These open datasets share the following 4 characteristics:

Accessibility: A wide range of users is permitted to access the data.

Machine readability: The data can be processed automatically.

Cost: Data can be accessed free or at negligible cost.

Rights: Limitations on the use, transformation, and distribution of data are minimal.”

8. Open Data Now²⁰²

“Open Data is accessible public data that people, companies, and organizations can use to launch new ventures, analyze patterns and trends, make data-driven decisions, and solve complex problems.”

9. Open Data Barometer²⁰³

Excerpt from report indicates that researchers assessed datasets based on the “full Open Definition requirements of being machine readable, accessible in bulk, and openly licensed.”

200 LinkedGov, “What is open data?” accessed February 23, 2014., <http://linkedgov.org/what-is-open-data>.

201 Manyika et al. “Open data: Unlocking innovation and performance with liquid information.”

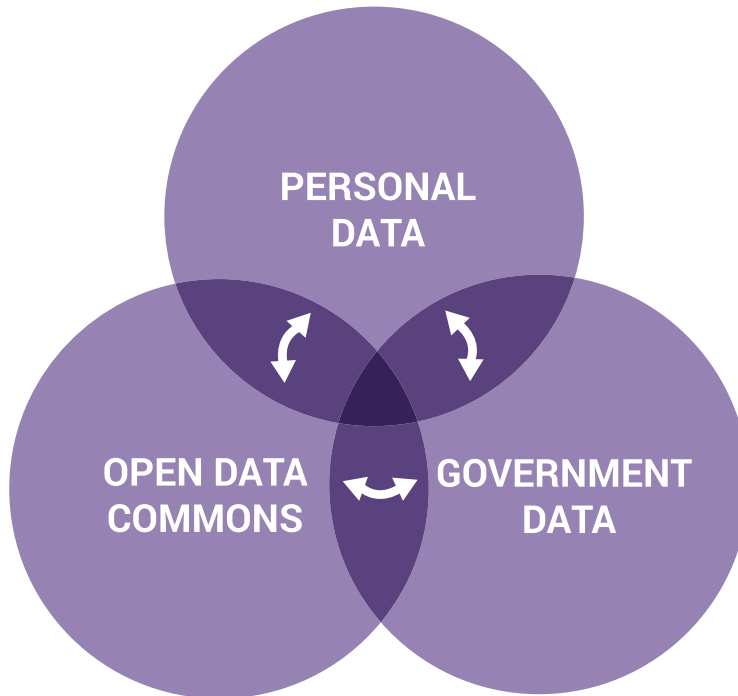
202 Gurin, *Open Data Now*.

203 Davies, “Open Data Barometer.”

Appendix III: Open Data and Big Data

Open health data exists within a broader context. Big data taxonomies explain the relationship between open data and other types of data such as personal data. One example distills data into categories based on control: Open data commons, personal and proprietary data, and government secret data.²⁰⁴ Categories within big data tend to overlap, signifying that in practice insights can be found by using each type in combination with one another.

Figure 3:

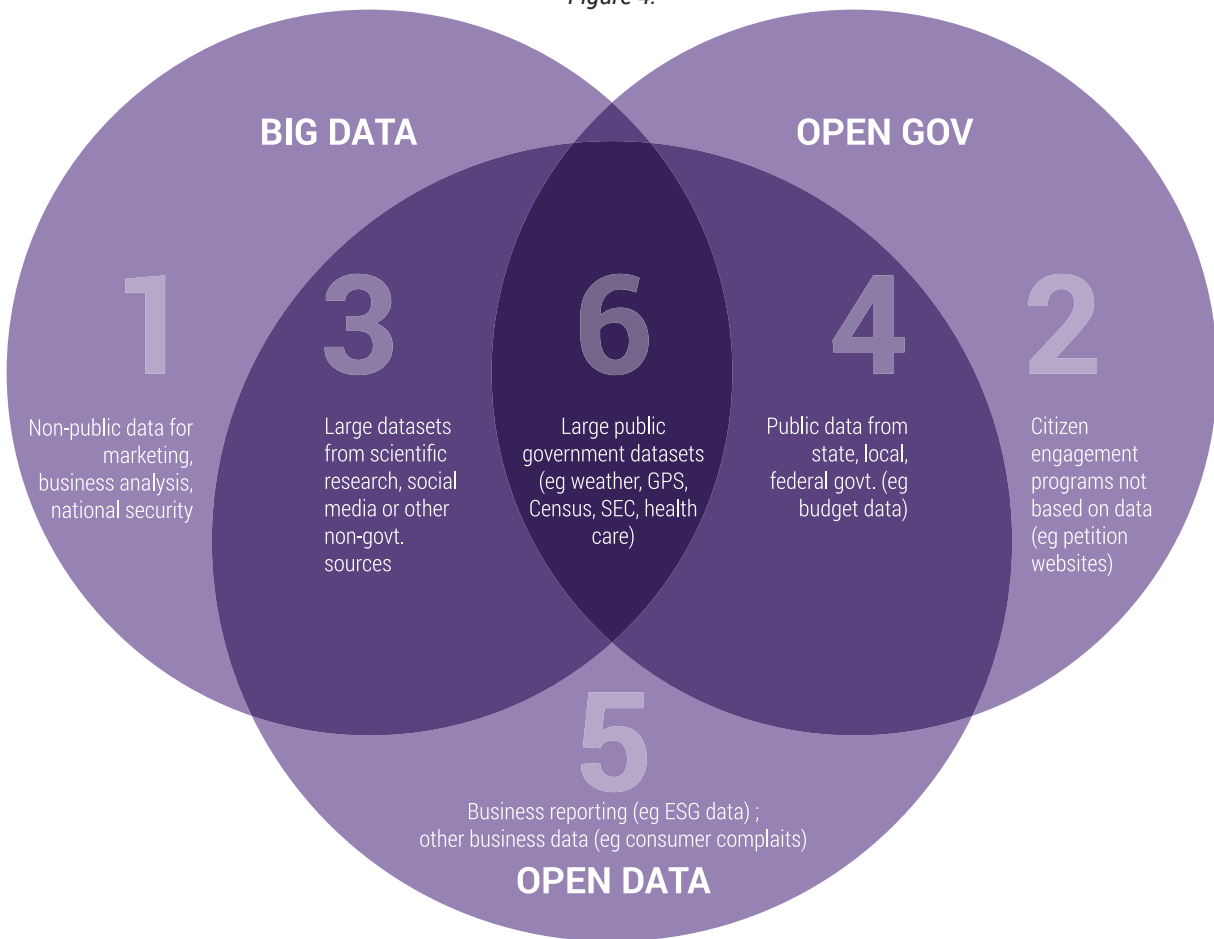


Source: Pentland, A., et al. *Big Data and Health: Revolutionizing Medicine and Public Health*. December 2013. <http://www.wish-qatar.org/app/media/382>

In his book, Joel Gurin fills in the overlapping combinations among big, government, and open data, all of which could fall in one of the three major categories below. For example, big data may not be open data and vice versa, which would then fall into personal or proprietary data or government secret data, depending on the source of control.

204 Alex Pentland, Todd Reid and Tracy Heibeck, "Big Data and Health: Revolutionizing Medicine and Public Health," WISH Big Data and Health Report 2013, <http://www.wish-qatar.org/app/media/382>.

Figure 4:



Source: Gurin, J. <http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/#.Uv5w7newLP1>

Further, big, open, non-governmental data and big, open, governmental data both fall within the Open Data Commons, as openness is their defining factor. For the purposes of this paper, big data will be compiled into the three control-based categories here.

Open Data Commons are made available to all with very little or no limitations on use. Most commons have been developed to help monitor the performance of specific policies. Pentland, et al explain that big data holds significant opportunity for the development of many more commons, which can be captured without human intervention and as a byproduct of other activities. Existing commons include geo-located and time-stamped statistics about aggregate costs, health outcomes and behaviors (e.g., mobility, smoking, drinking, crime, accidents).

Open Government Data is the most important and promising subcategory of open data, and is released by the government for anyone's use. Whether classified under the umbrella of big data or not, it can have a major impact when released. Examples include weather, census and GPS data, as well as agency or departmental data that helps people use city services like public buses or health clinics.²⁰⁵

Personal and Proprietary Data are intrinsically more restrictive, as they are controlled by individuals or companies and require more stringent infrastructure, use, and auditing controls. They include digital breadcrumbs (mobility, call, and purchasing patterns, etc.), personal annotations about eating, subjective variables, and standard health data (temperature, glucose, genomic, etc.).

205 Pentland, Reid, and Heibeck, "Big Data and Health: Revolutionizing Medicine and Public Health."

The Open Data Era in Health and Social Care

Government Secret Data typically includes private data such as tax data, detailed census and expenditure data, detailed healthcare system costs, and individual performance ratings. It is kept secret under the assumptions that the government is protecting the public and these types of information would be damaging if they were made public. Challenges arise with the collision of big data and health systems, as governments increase the amount of individual behavior data and other personal information.²⁰⁶

206 Open Data Now, "Big Data vs. Open Data – Mapping it Out," November 8, 2013, accessed February 23, 2014, <http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/#.Uv5w7newLP1>.

Appendix IV: Interviews

John Billings, Director, NYU Wagner Graduate School of Public Service - Health Policy and Management Programme

Beverly Bryant, Director of Strategic Systems and Technology, NHS England

John Coulthard, Director Customer Relations, NHS England

Emma Doyle, Head of Open Data and Transparency, NHS England

Francois Grey, Head of Citizen Science, NYU Center for Urban Science and Progress

Xanthe Hannah, Strategic Intelligence Projects Lead, NHS England

Mark Headd, Chief Data Officer, City of Philadelphia

James Hendler, Director, Rensselaer Institute for Data Exploration and Applications

Cesar Hidalgo, Assistant Professor, MIT Media Lab

Tim Kelsey, National Director for Patients and Information, NHS England

Heather Kirkham, Senior Analyst, Walgreens

Geraint Lewis, Chief Data Officer, NHS England

Samantha Riley, Director of Insight, NHS Commissioning Board

Bryan Sivak, Chief Technology Officer, United States Department of Health and Human Services

Roger Taylor, Director of Research and Public Affairs, Dr Foster Intelligence and National Professional Advisor, Care Quality Commission

John Tolva, President, PositivEnergy Practice

Ian Townsend, Strategic Intelligence Projects Lead, NHS England

Appendix V: Open Data in the NHS

TYPE	EXAMPLE	DESCRIPTION	PUBLISHER
Government			
Organograms	Organogram and staff pay data for Department of Health	A list of most Senior Civil Service posts in the Department of Health including title, contact details, their line manager, and where disclosed, the name of the officer.	Department of Health
Population Health Estimates	Healthy Life Expectancy at birth for Upper Tier Local Authorities: England	Estimates of healthy life expectancy by upper tier local authority, using a number of data sources including national surveys, population estimates and death registrations over a range of time periods.	Office for National Statistics
Workforce Data	Sickness Absence Rates in NHS	Statistical bulletin that relates to sickness absence rates for staff at NHS organisations.	Health and Social Care Information Centre
Government Spending			
Expenditure Data	Expenditure on Healthcare in the UK	Expenditure on Healthcare in the UK	Office for National Statistics
Financial Transactions	Spend over £25,000 in the Department of Health	A monthly-updated list of all financial transactions spending over £25,000 made by the Department of Health, as part of the Government's commitment to transparency in expenditure.	Department of Health
Health			
Official Statistics	Primary Care Trust Prescribing Data	Presents information on prescribing practices by primary care trust.	Health and Social Care Information Centre
Surveys	Friends and Family Test	The responses of maternity patients, and patients who have been treated in an acute setting as inpatients or in A&E, to the question "how likely are you to recommend our ward/A&E department to friends and family if they needed similar care or treatment?"	NHS England
Central Returns	Diagnostic Waiting Times and Activities	Central Returns Data Sets support: Information requirements for national and local performance management, planning and clinical governance Assurance of the quality of health and social care services.	Health and Social Care Information Centre

TYPE	EXAMPLE	DESCRIPTION	PUBLISHER
Practice Extracts Clinical and Administrative	Hospital Episode Statistics (HES)	HES data is collected during a patient's time at hospital and is submitted to allow hospitals to be paid for the care they deliver. The HES data that is made available as Open Data is aggregated and anonymized and is formatted to enable uses for non-clinical purposes.	Health and Social Care Information Centre
Indicators	Public Health Outcomes Framework	Compendium of public health outcomes indicators presented at England and upper tier LA level.	Public Health England
Audits	National Diabetes Audit	Audit participation by Primary Care Trusts (PCTs). Aggregate measures about the process of care given to patients. Information about care outcomes and treatment. Information about complications and mortality	Health and Social Care Information Centre
National Tariff Information	NHS Payment by Results 2010-11 National Tariff Information	This dataset contains the mandatory admitted patient, outpatient attendance, outpatient procedure, accident and emergency (A&E) and Best Practice tariffs applicable in 2010-11.	Department of Health
Society			
Demographics	Healthy lifestyle behaviours	Provides 'synthetic estimates' and confidence intervals for a range of healthy lifestyle variables. The variables are: 1) Current smoking among adults (aged 16 or over). 2) Binge drinking for adults (aged 16 or over). 3) Obesity among adults (aged 16 or over). 4) Consumption of 5 or more portions of fruit and vegetables a day among adults (aged 16 or over)	Office for National Statistics

Appendix VI: Activities

Activities exist at the level of the open data initiative. As illustrated by the conceptual framework, users use open data and restricted data, possibly in conjunction with or independently of privately held data, which leads to specific outputs or results. Each unique activity leads to its own measurable output indicator.

Open data will be used for different purposes, depending on the position and intention of the user accessing the information, and whether the activity is done internally (within the NHS and related departments), or externally (by media organizations, citizens, policymakers, or other researchers).

Internal Activities and Outputs

Internal activities can include, but are not limited to:

- ▶ *Benchmarking*: Comparing measures against a standard for the purposes of improving quality or level of care within the NHS system.
- ▶ *Resource allocation and management*: Identifying where resources, such as emergency services, will be most useful at different times in varying locales, and decreasing wait times in various stages of the healthcare process.
- ▶ *Managing health information at the population level*: Tracking illnesses and disease to identify risk factors for members of the general population.

Analyzing diagnostics: Assessing treatment methods and risk factors for various diseases, both treatable and chronic.

External Activities and Outputs

Activities: Examples

- ▶ *Patient managing personal health information*: By collecting health data, patients can be given access to their own information, and compare it against desired healthcare outcomes.
- ▶ *Determining treatment options*: Given access to more open data, patients can determine which treatment options are available, and which have higher success or failure rates.
- ▶ *Comparing and choosing services*: As part of the NHS Choices initiative, patients have the right to choose their doctor, choose their healthcare service providers, and make other healthcare decisions. Open health data can allow patients more access to information they need to make these choices.
- ▶ *Researching disease and public health*: Open health data will make tracking diseases easier for the general population, and will enable them greater control over preventable illnesses that could affect them or their families.
- ▶ *Developing programmes, products, and services (apps and online tools)*: Businesses and outside commercial interests can use open data to develop applications that can be used by the general public or the healthcare industry.
- ▶ *Benchmarking*: Citizens can use open data to compare actual healthcare outcomes against the standards proposed by the NHS in their policy frameworks.
- ▶ *Advocating for better services*: Open health data can be used by watchdog agencies, external regulators, and citizen interest groups, to ensure that healthcare is being distributed with an eye towards guaranteeing equality across a variety of needs and interests in the general population.

Appendix VII: Outputs

Outputs are the result of the open data initiative. They will vary depending on the purpose and type of activity, and are basic measures, such as number or percent of increase or decrease, that are produced out of open data activities.

Outputs: Examples

- ▶ *Number of additional patients served (post-open data use/application):* By gathering data on resource allocation, the NHS can direct resources more effectively so as to serve more patients.
- ▶ *Number of pounds saved:* Using open data, resources can be allocated more effectively, waste can be tracked, and dollars can be saved or invested into new programmes.
- ▶ *Number of patients using data tools for decision making (i.e. NHS Choices):* Open data can increase the amount of patient and public participation in decision-making for both individual health concerns, and to determine the direction of services.
- ▶ *Percent decrease contraction of preventable in-hospital disease (C-Diff):* Open data can be used to track the change in number of preventable infections in hospitals, enabling a comparison between rates between hospitals.
- ▶ *Number of datasets accessed by NHS staff and other internal users:* The effectiveness of open data is in part determined by whether it is of use to both internal and external users.
- ▶ *Number of policies created/changed using insights gained from open data:* The real-world impact of open data can be measured in part based on its policy impact.

Outputs/Measures: Examples

- ▶ *Patient Satisfaction:* Patients should be given surveys to determine the impact of open data in determining their treatment, and choice of services.
- ▶ *Number of patients using data to determine treatment options/choice in services and providers:* Release of data should increase the amount of patient participation in their own treatment process.
- ▶ *Number of datasets accessed by external users:* Release of data should increase the amount of public participation in healthcare services by citizens, regulators, researchers, and other stakeholders and interest groups.
- ▶ *Number of products created by private sector entities:* Release of data will increase the number of products, services and applications developed by businesses and other commercial interest groups.
- ▶ *Number of datasets cited in academic articles or media reports:* Release of data will increase research in healthcare and other disciplines in the social sciences.
- ▶ *Number of bills/legislation introduced or passed informed by open data insights:* Release of open data will increase the number of policy initiatives undertaken by policymakers and other scholars.

Appendix VIII: Indicators

Indicators build on outputs, or basic measures of success for any given activity undertaken through the use of open health data. They are generally short-term indicators that can determine whether an activity led to a desired result. More high-level indicators make use of these outputs to determine whether the activities led to any long-term changes within the NHS system.

Collective Indicators: Examples

- ▶ *Decreasing patient mortality*: Release of data should be done with an eye towards improving healthcare outcomes, for example reduced premature mortality rates.
- ▶ *Decrease average hospital stay*: Release of data should decrease average hospital stays by enhancing resource allocation and increasing treatment options.
- ▶ *Decreasing spending in hospital*: Open data can be used to lower spending by ensuring that healthcare services are being used as intended, and that unnecessary treatments are performed less frequently.
- ▶ *Increase/decrease in patients seen within a given time period (week, month, quarter/term) by specialists*: Release of data should be used to determine the optimal number of patients that can be seen by specialists in the NHS system.
- ▶ *Staff productivity*: Open health data can be used to ensure that healthcare professionals are meeting goals and desired outcomes. Additionally, open data can be used to identify when and where extra staff is needed.
- ▶ *Decrease in waste, fraud (number of dollars saved)*: The impact of open data will be seen by measuring cost and spending.
- ▶ *Decrease in preventable disease and infections*: The impact of open data will be seen in a decrease in preventable diseases, if the data is used to communicate dangers and risk factors effectively, and if it is used to determine better treatment options.
- ▶ *Increase in patients using data to determine treatment and choice*: Surveys and site analytics should demonstrate that patients are using data to determine their treatment options and choice in services.



GOVLAB

The GovLab is a global action-research center seeking new ways to solve public problems using advances in technology and science. Our goal is to advance our understanding of how 21st century citizen engagement can make governance more effective and legitimate.