SMARTER HEALTH

BOOSTING ANALYTICAL CAPACITY AT NHS

Beth Simone Noveck
Stefaan Verhulst
Andrew Young
Maria Hermosilla
Anirudh Dinesh
Juliet McMurren

GOVLAB

thegovlab.org

> *It would be impractical and prohibitively expensive to fulfill all of NHS's analytical needs through more hiring*

EXECUTIVE SUMMARY

Public institutions such as the National Health Service in England increasingly want—and are expected—to base their actions on nationally agreed standards rather than anecdote. The collection and analysis of data, when done responsibly and in a trusted manner, has the potential to improve treatment and drive towards value, both social and economic, in healthcare. However, the goal of using data to improve the NHS and social care is hampered by a "talent gap" – a lack of personnel with data analytical skills – that stands in the way of uncovering the rich insights expected to reside in the NHS' own data. The NHS is not unique among public and even private sector institutions who are struggling to identify, hire and retain people with data science skills, and, above all, with the ability to apply new technological tactics to advancing the institution's priorities.

Take two examples. An independent organisation called the *Healthcare Quality Improvement Partnership (HQIP)* conducts forty annual audits comparing hospital and physician outcomes, and the implementation of National Institute of Clinical Excellence standards across England and Wales. But, says HQIP Director Dr. Danny Keenan, although we have the expertise to do the analysis, "we are woefully inadequate at *translating such analysis into improvements*. What's the takeaway for the hospital or community provider Board or the medical director? They cannot understand what they have to do."

Dr. Geraldine Strathdee chairs the National Mental Health Intelligence Network, based at Public Health England, and which, together with stakeholders and partners, launched the Fingertips Mental Health data *dashboard* of common mental health conditions in every locality. Strathdee points out there is a tremendous need for such benchmarking data: to plan local services based on local need, build societal community assets, design and improve NHS services and value. Without it, NHS resourcing is just based on historical allocations, guesswork or the 'loudest voice'. An *example* is psychosis: "you can spend sixty percent of your budget on poorly treated psychosis with people ending up in hospital beds, homeless or in prison and less than 10% ever get employment," says Strathdee. The data dictates investment in early intervention psychosis teams, which dramatically improve outcomes and value (50% of patients get back to education, training, employment), however, there is a shortage of people, she explains, to draw these insights. "We have a major capability level problem everywhere," says Strathdee.

Over the past twenty years, the NHS has amassed increasing amounts of data. There is almost universal *consensus* that, when used ethically and responsibly respecting and protecting confidentiality of all patients (especially, I might add, that of vulnerable populations such as immigrants), better use of data could bring about more high quality, accessible and effective services and better outcomes. But this requires more than just technology. The NHS needs data analytical talent, which comes in different forms and comes from a variety of disciplines.

Such know how includes the ability to ask the questions and wield the tools to identify patterns in both structured and unstructured records. For example, Leeds Teaching Hospitals *analyses* approximately one million unstructured case files per month. They have identified 30 distinct areas for improvement in either costs or operating procedures by using natural language processing to identify wasteful procedures such as unnecessary diagnostic tests and treatments.

Data analysis also requires people who know how to apply diverse statistical methods to determine future outcomes. EPSRC has funded *five new research centres* around the UK that will apply mathematics and statistics to prediction models in order to help clinicians tackle serious health challenges such as cancer, heart disease and antibiotic resistant bacteria.

Improving public institutions with data also requires strong communications, design and visualisation skills. Talented digital designers are needed who know how to turn raw data into *dashboards* and other feedback mechanisms to aid managers' decisions and actions.

And none of this is possible without the legal and ethical training that allows for the development of policies, platforms and procedures that enable data to be shared and algorithms to be used responsibly and effectively – toward the goal of improving people's lives and protecting confidentiality.

GOVLAB

Hence the NHS needs to be able to tap into those with a wide range of data analytic know-how. These include credentialed computer scientists, statisticians, economists, ethicists and social scientists. But credentials alone may not reveal the many people in the organisation who have practical skills in using data analysis and visualization platforms and, perhaps more important, experience navigating the NHS and its institutions. Both the technical skills and the organisational ones are needed to apply data toward transformation in the context of a bureaucracy.

It is *impractical and prohibitively expensive* to fulfill all of NHS's analytical needs through more hiring. But there are promising and inexpensive ways that the NHS can match its demand for data science expertise to the supply of knowledgeable talent both within and outside the organization, especially using new technology.

For example, to make better use of sensitive administrative data to measure what works, the NHS should expand upon and invest in its efforts already underway to construct an NHS Data Lab modelled on the *Ministry of Justice Data Lab*. The Justice Data Lab is a secure facility that accepts requests from anyone wishing to test the effectiveness of a program by using the government's administrative data to assess outcomes. Although it employs only three people, the Lab has already conducted over 173 analyses over the past three years. The Justice Data Lab enables what New Philanthropy Capital's David Pritchard describes as "mass produced, low quality, quasi-experimental studies." NHS England and the Health Foundation have started planning *such a service* and more needs to be done.

The NHS should also create a variety of online knowledge networks of those inside and outside of NHS, especially in UK universities, who possess the skills and willingness to help the NHS with their data analytical questions. For example, last week the Rockefeller Foundation launched the *Zilient platform* to connect resilience practitioners, and the GovLab and Justice Management Institute launched DataJustice. Both are designed to connect networks of professionals for mutual learning. NHS Improvement and the Health Foundation created an online network for improvement practitioners called the *Q Network*. Now NHS England is taking steps to create a *Population Health Analytics Network*, bringing academics, commercial and charitable organizations together with NHS to help accelerate learning. But it can do more and expand the agility and scale of such groups by using new technology.

Such networking platforms help to catalogue relevant skills and make them more searchable. For example, the *World Bank's SkillFinder* tracks expertise and experience across three dimensions: technical expertise, geography, and business processes. In New York City, the *Mayor's Office Volunteer Language Bank* tracks only a single skill—translation—to match the supply of multilingual civil servants to the demand for translation services.

Whether the NHS wants to know which techniques to use to spot the most high-risk patients or how to allocate beds during a particularly cold winter, or explain why a given community has a disproportionate incidence of a certain condition, it can build upon these models of online knowledge networks to find the talent hiding in plain sight

inside and outside the health and social care system and query those willing to lend their expertise. The difference between active and successful online networks and those that start with a bang and end with a whimper of inactivity, however, is how relevant they are to institutional decisionmaking practices. The process of querying distributed expertise needs to become part of how decisions are taken on a systematic and regular basis with feedback provided to the community as an incentive to engage.

Regardless of the platform the NHS adopts, it has an urgent need to convene its different bodies who are already undertaking data analytical work to catalog and compare research agendas, share learning, consider trust building provisions and methods, and identify gaps in order that the NHS can effectively and responsibly understand its need for data science skills. The NHS can then take advantage of new models like the Data Lab and the knowledge network to match the supply of talent to the demand.

Informed by both a literature review and analysis as well as over fifty interviews with NHS and other experts, this paper offers a multiplicity of proposed recommendations for meeting the data analytic talent gap and achieving greater institutional readiness without full-time hiring. These recommendations complement those made by the Health Foundation in its December 2016 focused on improved training and cross-organisational teams to enhance data analytical insight. No one approach alone will be enough.

These recommendations fall into four categories.

▸ Using new technology to coordinate *distributed talent* already present in the NHS, including project marketplaces.

▸ Using new technology such as talent banks and skill finders, to find *talent hiding in plain sight*—namely those with the relevant skills but who are not classed as analysts and match them to projects.

▸ Using expert networks to connect with empirical *expertise outside the NHS.*

▸ Creating *cost-effective incentives* to bring talent in from outside, including prize-backed challenges and foundation-funded fellowships.

RECOMMENDATION 1

The NHS should build a Project Marketplace like the environmental protection agency's one EPA Skills Marketplace and help supply find the demand

By enabling the posting of project descriptions together with requests for help with specific assignments, the skills marketplace helps to match talent to those opportunities to use it. The marketplace encourages teams focused on outcomes rather than affiliation.

RECOMMENDATION 2

The NHS should construct an NHS Data Lab modeled on the Ministry of Justice Data Lab to make better use of sensitive administrative data

The Justice Lab accepts requests for analysis from anyone wishing to test the effectiveness of their interventions against a specific data set. The Lab is a secure facility that provides reports on recidivism to help measure the success of interventions.

RECOMMENDATION 3

The NHS should build an employee expert network like health and human services' HHS Profiles and help the demand find the supply across the NHS

The American equivalent of the Department of Health – Health and Human Services – is building an expert network to match the supply of certain kinds of biomedical expertise to demand. HHS Profiles imports and analyses "white pages" (detailed contact) information, publications, and other data sources to create and maintain a complete, searchable library of web-based electronic CVs within HHS. Such an expert network could try to be comprehensive regarding analytical skills, but focus on a small part of the NHS or, as discussed in Recommendation 5, focus on a single skill across the broader enterprise

RECOMMENDATION 4

The NHS should set up a web-based help desk like the World Bank's Skill Finder to accelerate the process of matching demand to supply

The World Bank has catalogued and made searchable the know-how of its 27,000 employees, consultants and alumni. To accelerate adoption, the Bank is testing its ability to target and pinpoint expertise by launching a web-based help desk. The combination of help desk with expert network helps to match demand to supply.

GOVLAB

RECOMMENDATION 5

The NHS should build a single skill expert network modeled on the New York City's Mayor's Office volunteer language bank or the data justice network focused on data science

The Volunteer Language Bank matches the supply of multilingual civil servants to the demand for translation services. Volunteers are self-selected. Currently, more than 1200 volunteers, who speak approximately 70 languages, have joined. Data Justice connects state and local criminal justice practitioners to develop peer to peer learning about data science.

RECOMMENDATION 6

The NHS should use two-stage prize-backed challenges like the Harvard Catalyst Diabetes Challenge to augment the supply of internal data science capacity with external know-how

Harvard Catalyst opens each aspect of the research process to more input from a wider audience. In the Catalyst challenge, contributors supplied the questions rather answering questions from the organizers. Many had expertise that diverged markedly from diabetes investigation.

RECOMMENDATION 7

The NHS should use more foundation funded fellowships modeled on the Presidential Innovation Fellowship to augment the supply of internal data science capacity with external know-how

Fellowships complement traditional hiring to bring in more talent from outside, including from the private sector. In this case, the time to hire is unusually quick and the fellowships are focused on solving a specific problem articulated by a public agency.

RECOMMENDATION 8

The NHS should create an Open Data Learning Hub like the Commerce Data Usability Project for data scientists to grow its data science community

The Usability Project offers guided video tours concerning how to turn the Department's raw data into valuable knowledge. Each toolkit includes the original data and step-by-step instructions for using the data to conduct sample analyses, create visualisations from the data and connect interested data scientists.

RECOMMENDATION 9

The NHS should establish a process to explore the creation of Data Collaboratives

This is a new form of collaboration in which participants from different sectors—including private companies, research institutions, and government agencies—exchange data to help solve public problems. These collaborations focus both on sharing data but also on sharing talent.

RECOMMENDATION 10

The NHS should connect UK university know how systematically to the NHS using a public-facing expert network

A "public data science brain trust" combines a directory of people and what they know with communications tools that facilitate questions and discussion. This adds a "conversational" infrastructure to a secure physical infrastructure for managing data to tap the best know-how outside NHS on an ongoing basis.

Regardless, of the institutional arrangement that NHS adopts to accelerate its ability to obtain more expertise, there is an urgent need to convene the different bodies of the NHS currently undertaking data analytical work to catalog and compare research agendas, avoid duplication and share learning.

As NHS Chief Clinical Information Officer Professor Keith McNeil has urgently *articulated*: "It is evident the benefits of investing in and optimising use of digital technology to improve efficiency and enhance care is more widely understood but we are not yet realising these benefits at scale or sufficiently quickly." In order to realise this vision for data-driven clinical excellence and organisational efficiency, the most pressing and important challenge of all, however, is not creating a supply of analytical capability, but articulating and communicating why data analytics are essential to improved operation, policy and change.

There is now every reason to articulate a leadership vision and implement organisational practice that weaves together the NHS's strategic goals for patient outcomes with the tactics and tools of data science. Undertaking initiatives in some or all of these areas will enable the NHS to identify and exploit the deep reservoirs of expertise within and outside the

organisation. This will enable the NHS to invest more in predictive analysis that maximizes wellness and health outcomes rather than simply be limited to measuring compliance and past performance. To create a data-driven culture, however, it is not enough to have analysts buried three layers deep in bureaucracy. Rather, data scientists and technologists need to be 'at the table' for all major policy decisions; enabling new policies and services to be informed by evidence of past practice and modeling of future scenarios.

The paper contains:

▸ An overview of the benefits provided by data collection and analysis

▸ A taxonomy of skills required for data analysis, developed through consultations with the NHS and a survey of selected staff

▸ A description of data analysis capabilities currently available within the NHS, the Department of Health, and other health organisations in the UK

▸ Examples of data use and the barriers to more effective exploitation of its benefits in NHS

▸ Recommended projects that utilise resources already within the NHS, and in the community, to improve data use without straining budgets to the executive summary so people know what's in the report if you paste into the end of the exec summary, I'll review that to see if it looks right

ABOUT THE GOVLAB

The GovLab's mission is to improve people's lives by changing the way we govern.

Our goal is to strengthen the ability of institutions – including but not limited to governments – and people to work more openly, collaboratively, effectively and legitimately to make better decisions and solve public problems.

We believe that increased availability and use of **data**, new ways to leverage the capacity, intelligence, and expertise of **people** in the problem-solving process, combined with new advances in **technology and science** can transform governance.

We approach each challenge and opportunity in an **interdisciplinary, collaborative** way, irrespective of the problem, sector, geography and level of government.

GOVLAB

SMARTER STATE

SMARTER HEALTH

BOOSTING ANALYTICAL CAPACITY AT NHS

Beth Simone Noveck
Stefaan Verhulst
Andrew Young
Maria Hermosilla
Anirudh Dinesh
Juliet McMurren

thegovlab.org

CONTENTS

Public institutions such as the National Health Service in England increasingly want—and are expected—to base their actions on nationally agreed standards rather than anecdote. The collection and analysis of data, when done responsibly and in a trusted manner, has the potential to improve treatment and drive towards value, both social and economic, in healthcare. However, the goal of using data to improve the NHS and social care is hampered by a "talent gap" – a lack of personnel with data analytical skills – that stands in the way of uncovering the rich insights expected to reside in the NHS' own data. The NHS is not unique among public and even private sector institutions who are struggling to identify, hire and retain people with data science skills, and, above all, with the ability to apply new technological tactics to advancing the institution's priorities.

Take two examples. An independent organisation called the *Healthcare Quality Improvement Partnership (HQIP)* conducts forty annual audits comparing hospital and physician outcomes, and the implementation of National Institute of Clinical Excellence standards across England and Wales. But, says HQIP Director Dr. Danny Keenan, although we have the expertise to do the analysis, "we are woefully inadequate at *translating such analysis into improvements*. What's the takeaway for the hospital or community provider Board or the medical director? They cannot understand what they have to do."

Dr. Geraldine Strathdee chairs the National Mental Health Intelligence Network, based at Public Health England, and which, together with stakeholders and partners, launched the Fingertips Mental Health data *dashboard* of common mental health conditions in every locality. Strathdee points out there is a tremendous need for such benchmarking data: to plan local services based on local need, build societal community assets, design and improve NHS services and value. Without it, NHS resourcing is just based on historical allocations, guesswork or the 'loudest voice'. An *example* is psychosis: "you can spend sixty percent of your budget on poorly treated psychosis with people ending up in hospital beds, homeless or in prison and less than 10% ever get employment," says Strathdee. The data dictates investment in early intervention psychosis teams, which dramatically improve outcomes and value (50% of patients get back to education, training, employment), however, there is a shortage of people, she explains, to draw these insights. "We have a major capability level problem everywhere," says Strathdee.

Over the past twenty years, the NHS has amassed increasing amounts of data. There is almost universal *consensus* that, when used ethically and responsibly respecting and protecting confidentiality of all patients (especially, I might add, that of vulnerable populations such as immigrants), better use of data could bring about more high quality, accessible and effective services and better outcomes. But this requires more than just technology. The NHS needs data analytical talent, which comes in different forms and comes from a variety of disciplines.

Such know how includes the ability to ask the questions and wield the tools to identify patterns in both structured and unstructured records. For example, Leeds Teaching

Hospitals *analyses* approximately one million unstructured case files per month. They have identified 30 distinct areas for improvement in either costs or operating procedures by using natural language processing to identify wasteful procedures such as unnecessary diagnostic tests and treatments.

Data analysis also requires people who know how to apply diverse statistical methods to determine future outcomes. EPSRC has funded *five new research centres* around the UK that will apply mathematics and statistics to prediction models in order to help clinicians tackle serious health challenges such as cancer, heart disease and antibiotic resistant bacteria.

Improving public institutions with data also requires strong communications, design and visualisation skills. Talented digital designers are needed who know how to turn raw data into *dashboards* and other feedback mechanisms to aid managers' decisions and actions.

And none of this is possible without the legal and ethical training that allows for the development of policies, platforms and procedures that enable data to be shared and algorithms to be used responsibly and effectively – toward the goal of improving people's lives and protecting confidentiality.

Hence the NHS needs to be able to tap into those with a wide range of data analytic know-how. These include credentialed computer scientists, statisticians, economists, ethicists and social scientists. But credentials alone may not reveal the many people in the organisation who have practical skills in using data analysis and visualization platforms and, perhaps more important, experience navigating the NHS and its institutions. Both the technical skills and the organisational ones are needed to apply data toward transformation in the context of a bureaucracy.

It is *impractical and prohibitively expensive* to fulfill all of NHS's analytical needs through more hiring. But there are promising and inexpensive ways that the NHS can match its demand for data science expertise to the supply of knowledgeable talent both within and outside the organization, especially using new technology.

The paper contains:

▸ An overview of the benefits provided by data collection and analysis

▸ A taxonomy of skills required for data analysis, developed through consultations with the NHS and a survey of selected staff

▸ A description of data analysis capabilities currently available within the NHS, the Department of Health, and other health organisations in the UK

▸ Examples of data use and the barriers to more effective exploitation of its benefits in NHS

▸ Recommended projects that utilise resources already within the NHS, and in the community, to improve data use without straining budgets

GOVLAB

INTRODUCTION
DATA AND ITS HUMAN COMPONENT

The health care field, like other areas of business, government, and research, is being thoroughly transformed by data collection and analysis. A few examples of the trend:

▸ The Cameron Administration launched a high-profile big data genomics initiative to map one hundred thousand human genomes by 2017. Similarly, the Obama Administration launched a high-profile *Precision Medicine Initiative* to cull behavioral, environmental, and genomic data on large numbers of people in a secure manner for health research.

▸ Leeds Teaching Hospitals analyse approximately one million unstructured case files per month, and have identified 30 distinct areas for improvement in either costs or operating procedures by using natural language processing (NLP). This unstructured data analytical capability enables Leeds to improve efficiency and control costs through identifying wasteful procedures such as unnecessary diagnostic tests and treatments.

▸ Kaiser Permanente, the largest private health provider network in the United States, maintains a *Division of Research* using numerous data sets for health care improvement.

▸ The EPSRC Centre for Mathematics of Precision Healthcare at Imperial was launched to link up mathematical, computational and medical capabilities for data-rich precision healthcare. It is one of *five new research centres* around the UK that will explore how mathematics and statistics can help clinicians to tackle serious health challenges such as cancer, heart disease and antibiotic resistant bacteria.

In recent years, NHS England has invested substantial financial and intellectual resources in its own path toward data-driven transformation. An incipient ecology of innovation, analysis and research has begun to emerge from these investments. There is clear evidence that better information—used appropriately and responsibly, with respect for personal privacy and security—can help the NHS commission, manage and govern health care. Data can also help make patients and citizens in Great Britain healthier, and ensure that they live longer lives.

Data collection and analysis can identify fraud, waste and abuse in spending. Better algorithms can spot the most high-risk patients, allocate beds during a particularly cold winter, or explain why a given community has a disproportionate incidence of a certain condition. Overall, NHS data-driven efforts have the potential to transform the delivery and experience of healthcare, spurring greater efficiencies, more innovation, and—perhaps most importantly—better health outcomes.

However, as the NHS and local Clinical Commissioning Groups (CCG) scale up their use of data—a trend documented later in the section—they are hitting a ceiling caused by a lack of access to experts in analysing and interpreting data. As Professor John Newton Chief Knowledge Officer Public Health England and Deputy Chair of the National Information Board comments, the NHS has a great history of collecting data but it is not so good at "getting it out." It is essential to have the resources in place to ensure that the NHS is using data to its maximum potential in primary care, hospital, mental health and community care settings alike and that social care is part of the information landscape. This report cites cost-effective, innovative measures that can be taken to uncover resources within NHS and in the larger community in which it resides.

New technologies like hardware, platforms, and analytical tools will be important in this process. But the central premise of this report is that human capacity and skills will play the most important role going forward. Consequently, there is a vital need to build human analytical and data capacity within the NHS in the coming years.

The draft NHS Data Strategy outlines a vision of increased data availability, better data quality, and more effective information governance.

The Strategy also describes in detail what it will take to realise this vision, and the many shortcomings that need to be overcome. In many cases, as we discuss below, necessary data is simply not collected; e.g., the system routinely discards appointment data that might yield insights into poor utilisation of existing resources.

Often the data collected is of low quality and significant gaps and delays frustrate the collection of patient-level mental health data. In addition, unclear governance structures and policies slow the process of collecting and reporting patient-level data, impeding the access of those who could make most use of it.

In short, much must be done if the NHS is to achieve its commitment to give all citizens electronic access to their own records by 2020 and, at the same time, become a successful consumer of its own data, able to realise operational improvements as a result of analytical insights.

Technology alone will not be sufficient to extract value and realise improvements from the treasure trove of information upon which NHS sits. Technical infrastructure–more data availability, better data quality, and data governance–provide only part of the response. Even with perfect and comprehensive data flows, more attention must be paid to questions of institutional readiness, and specifically to human capital and capacity, to translate raw data into actionable insights. Even as organisations collect more and more data at a granular level, and even as technologies improve for collecting, manipulating, storing and visualising such data, the real key to achieving a data-driven healthcare system is trained personnel who know how to work with data.

HARNESSING INTERNAL AND EXTERNAL HUMAN RESOURCES FOR ANALYSIS

Although the Cabinet Office's 2015 Talent Action Plan calls for building a world-class, 21st-century civil service and work has been done to build a data science community in the public sector, the government has a long way to go to improve numeracy and skills in evidence-based policymaking within the public sector.[1] The Department of Health has announced plans for a data science centre of excellence to be based at NHS Digital (formerly the Health and Social Care Information Centre (HSCIC)). But hiring and, in particular retaining, analytical talent with knowledge of computer science, statistics, and data management is challenging, especially given competition from the private sector and budget cuts in the public sector. Data analytical capacity, too, has to be understood not simply as a headcount of those with 'data' or 'analyst' in their job title but through a more refined understanding of the demand for different kinds of data science expertise and the supply of such know how independent of title and organisation.

---

1 Health Foundation, *Improvement Science and Our Plans for a New Institute, accessed 30 March 2016,* *https://goo.gl/M1R7at*. Paul Maltby and Sue Bateman, *Growing a Government Data Science Community, Civil Service Quarterly (Jan 2017), available at* *https://goo.gl/RNhyOh*.

Problem with the lack of capacity and the mismatch of demand to supply are not unique either to the NHS or the United Kingdom. In its analysis of nineteen open government data projects around the world, the GovLab repeatedly encountered a lack of readiness and capacity among government officials that hampered the ability of public institutions to make data available and gain insights from it.[2] Although government institutions are very experienced at collecting data, they are less accustomed to using it to change and innovate upon how they work today.

The talent shortage goes beyond the public sector. In its 2011 study on big data, McKinsey reported that by 2018, "the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions."[3]

It would thus be impractical and prohibitively expensive to fulfill the NHS's analytical needs exclusively through more hiring. Hiring must be accompanied by aggressive training and recruitment of existing talent within NHS and among those in the wider community who are dedicated to its goals.

Just as individuals use only part of their brain to solve most problems, most institutions make far too little use of the skills and experience of those inside and outside their walls—those with scientific credentials, practical skills, and ground-level street smarts in the emerging field of data science. Given the need for NHS England to increase its operational efficiency while reducing its budget, it is even more imperative to identify and make use of the skills and expertise already present within its organisation, and create networks across specialisations between healthcare workers, administrators and other healthcare professionals to address the data skills gap.

---

2   *Open Data's Impact, The Governance Lab, accessed 30 March 2016,* http://odimpact.org.

3   *James Manyika et al, Big data: The next frontier for innovation, competition, and productivity, May 2011, accessed 30 March 2016,* https://goo.gl/Wd68Y7

DEFINING DATA ANALYSIS AND ITS VALUE

This section takes the themes in the overview to greater depth. It covers key benefits provided by the use of data in health care, and then offers a taxonomy of skills in data analysis for health care. These skills draw on many academic disciplines and include more loosely defined forms of experience and personal ability.

THE BENEFITS OF DATA IN HEALTH AND SOCIAL CARE

Research conducted and curated by the GovLab in collaboration with the NHS has emphasised numerous reasons to use data in health and social care. Briefly, access to more and better data leads to the following potential benefits:

▸ **Improved outcomes and equities:** Comparisons of performance, outcome and spending data can identify what's working and what's not—and can identify populations to see for whom an intervention works.

▸ **From problems to operational solutions:** Data can explain the source of shortcomings in a healthcare setting and translate problems into solutions. For example, when data reveals unusually high rates of a condition in a specific place among a given population, policy-makers and clinicians can target their attention on investigating the anomaly and potentially developing more effective solutions. In addition, with better and more real-time data, the NHS can pilot, evaluate, and implement new strategies quickly based on evidence, and thus innovate more rapidly.

▸ **Lower costs and reduced inefficiencies:** Calculations by Monitor, NHS England, and independent analysts all predict a budgetary shortfall between current allocations and patient needs of nearly £30 billion a year by 2020/21. The sources of this mismatch—and the best ways to address it—may be identified through data analysis. In addition, benchmarking and estimating costs of care, especially for increasingly prevalent conditions such as autism, cancer or Alzheimer's disease, can enable more effective planning.

▸ **Increased choice:** An increase in the quality, completeness, and timeliness of information provided to patients can enable them to make better decisions in treatment plans and to choose appropriate and cost-effective services. More comprehensive data about individual conditions and care can lead to increasingly personalized outcomes information, which in turn, can lead to the development of more targeted clinical interventions as well as facilitating the distinction between harms caused by treatment versus illness.

▸ **Greater accountability:** Better data and data analytics can ensure that health regulators are meeting standards, and that physicians and hospitals are held accountable for unwarranted variation and poor practices, For example, in the number of individuals dying from preventable diseases or in rates of hospital-acquired infections. Active use of data is what helps to lay bare problems with data quality and support continuous improvement in data collection and use practices.

▸ **Stimulate innovation and economic growth:** More data and data analytics will allow more analyses at the population level. Increased data collection and distribution will enhance the UK's standing as a leader in research and innovation.

▸ **Increased patient satisfaction and customer service:** Using data to provide 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.

SKILLS IN THE DATA REALM

As the NHS assesses its capacity and needs for data analysis, it must define data analytical expertise to enable more comprehensive mapping. Big data is a popular phrase with many meanings, while "data analysis" is a catch-all term for a host of data assessment approaches and tools. Furthermore, these terms all touch on skills that are scattered among many academic disciplines.

When GovLab began its research with the NHS, we conducted a small *survey* among Expo 2015 attendees about NHS's data needs and capabilities, described in detail in Appendix 2. The "data analytical capacity" gap covers a broad range of activities and abilities:

▸ Which questions can be asked given the available data, such as how to frame a question about cost savings in light of gaps in information?

▸ How to elicit insights from data even without knowing what the questions are, such as how to identify disproportionate or unusual patterns of use of healthcare services that might be targeted for special interventions ("hot spots"), a research project where the variables being identified are unknown in advance?

▸ How to translate insights into evidence-based operational improvements, such as how to use data-driven learning about hospital infection in devising strategies to combat it?

▸ How to collaborate with clinicians and researchers to devise more individualised and effective healthcare?

▸ How to publish data responsibly for use by researchers and app makers while safeguarding personal privacy?

- How to combine healthcare with demographic and environmental data about educational attainment or employment, for example, to assess impact of innovative services on specific populations?

- How to make data visual and useful to citizens to inform their own decision-making and make patients more effective partners in care?

- How to collect and make use of data contributed by patients and citizens scientists, who can contribute highly granular sources of new data?

- What the urgency and frequency are for analysing different kinds of data so that the NHS can plan for and respond to emergencies, such as extreme weather conditions, major accidents or unforeseen events or infectious disease outbreaks?

- How to make the decisions about data governance, such as who should have access to which data and for what purposes?

Martin Bardsley in his report on analytical capacity for the Health Foundation summarizes the need as follows:[4]

- Understand and structure the problems faced by managers/clinicians

- Access and understand the evidence and information that can be brought to bear on a problem

- Apply appropriate and robust methods to manipulate information and data

- Communicate the findings accurately and clearly.

Informed by our interviews and bench research, in particular work done by scholars at Auburn University, we have summarized an initial taxonomy of data analytical expertise for healthcare along two axes:

- **Types of data analytical skills:** structured analysis, unstructured analysis, predictive, decision support, traceability, governance, implementation/communication/visualisation. and

- **Types of expertise:** credentials, skills, experience using data.

This taxonomy follows.

---

4   Martin Bardsley, *Understanding analytical capability in health care Do we have more data than insight?* *Health Foundation, December 2016, 4. See https://goo.gl/Nv455F.*

TYPES OF DATA ANALYTICAL SKILLS

Data analytical skills fall into the following primary categories:[5]

▸ **Structured analysis of patterns of care:** Analytical capabilities in healthcare can identify patterns of care and discover associations culled from massive healthcare records, thus providing a broader view for evidence-based clinical practice.

▸ **Unstructured data analytical capability:** Unstructured and semi-structured data in healthcare refer to information that can neither be stored in a traditional relational database nor fit into predefined data models. The ability to analyse unstructured data plays a pivotal role in the success of big data in healthcare settings because 80 percent of health data is unstructured.

▸ **Predictive capability:** Predictive capability is the ability to apply diverse methods from statistical analysis, modeling, machine learning, and data mining to both structured and unstructured data to determine future outcomes.[6]

▸ **Decision support capability:** Decision support capability aims to produce reports about daily healthcare services to aid managers' decisions and actions.

▸ **Traceability:** Traceability is the ability to track output data from all the system's IT components throughout the organisation's service units. Healthcare-related data such as activity and cost data, clinical data, pharmaceutical R&D data, patient behavior and sentiment data are commonly collected in real time or near real time from payers, healthcare services, pharmaceutical companies, consumers and stakeholders outside healthcare.

▸ **Data governance:** Data governance is the ability to manage overall availability, usability, integrity, and security of the data employed in an enterprise and to develop policies, platforms and procedures to enable data to be shared and algorithms to be used responsibly and effectively.

▸ **Implementation, Communication and Visualisation:** To be of value to the enterprise, data analysis must be translated into insights for managers. This requires a degree of clinical and institutional literacy as well as strong communications, design and visualisation skills.

---

5 *See Yi Chuan Wang, et al., Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care, 2015 48th Hawaii International Conference on System Sciences, accessed 30 March 2016,* *https://goo.gl/bIXPwl*

6 *Data Analytic Capabilities for Medicare Program Assessment, Center for Medicare and Medicaid Services, September 2014, accessed 30 March 2016,* *https://goo.gl/uUl59n*

TYPES OF EXPERTISE

As with other types of expertise, data analytic know-how in any of the previous topics comes in three different forms, which imply different strategies for searching for and uncovering this know-how: credentials, skills and experience.

▸ **Credentials:** Credentials signal technical knowledge gained through education and publication. Indicia are credentials issued by third party institutions such as universities or publishers. Data Science credentials include diplomas in computer science, statistics, economics, computation, social science, and informatics. These degrees and certificates can be useful, albeit incomplete, proxies of knowledge. Credentials can often be misleading, however, as degrees alone say little about actual skill and ability, or may indicate the possession of theoretical skills that fit poorly with current practice. Furthermore, many social science programs that were once qualitative have become heavily quantitative and might provide as much training in data analytics as a computer science degree. Many experts have published articles in peer-reviewed journals; others publish in magazines, but some of the best analysts do not publish at all.

▸ **Skills:** Skills are methods and tools developed and deepened through a combination of training and practice. Data science skills include the ability to use tools such as R, D3, Hadoop, Postgres, MapReduce, iPython for visualisation, natural language processing, graph analytics for analysis, or parallel databases for data management. In addition to data management, analysis, storage and visualisation, skills in data governance, data ethics, and domain expertise are equally important and relevant. The list of tools within each domain is growing and changing.

▸ **Experience using Data:** This includes NHS-specific experience, heuristics or routines/methods. Experience using data can be of greater relevance to a manager than any degree or tool. The know-how developed by the team in Leeds Teaching Hospitals, which analysed unstructured case files to identify unnecessary procedures, is as important as the specific database or visualisation they used.

Any effort to map analytical skills needs to consider both the data analytic know-how and the different categories of expertise. For example, limiting a search to LinkedIn will yield primarily credentials, without as much insight into skills and experience. An effort to track those who can do predictive modeling could miss those with the translational capacity to derive practical recommendations from raw data.

Healthcare data in England is collected, published and used by a variety of institutions, each of which has its own cadre of statisticians, analysts, and economists.

DATA IN THE NHS: THE CONTEXT

This section lays out the state of data analysis in the NHS today. We cover:

▸ Key organisations in the UK health space

▸ A brief chronology of milestones in the NHS's collection, use, and sharing of data

▸ Shortcomings in the use of current data analytics within the NHS

NAVIGATING THE HEALTH AND SOCIAL CARE DATA LAKE

In order to better understand, and ultimately meet, the analytical requirements of the NHS, we need to understand how and by whom its data is currently collected, stored, shared and used, and by whom as well as some key shortcomings in this process.

Healthcare data in England is collected, published and used by a variety of institutions, each of which has its own cadre of statisticians, analysts, and economists. This section offers an overview of where capacity currently resides.

NHS

Health and care data in England is collected, published and used by a variety of institutions, each of which has its own cadre of statisticians, analysts, and economists. As set out in the Health and Social Care Act 2012, the NHS holds the contracts for GPs and NHS dentists and manages (together with the CCGs, see below) primary care commissioning. As a result, it oversees the process though it does not collect or directly use the data generated by GPs. NHS England's 6500 mostly non-clinical staff members across fifty sites oversee the budget, planning, delivery and day-to-day operation of the commissioning side of the NHS in

England. Among its staff are approximately 150 people who hold the title of analyst for NHS England, including those in an operational research and evaluation unit created in 2014. NHS England publishes indicators on performance and satisfaction data about patient experience, bed availability, and wait times, and administers friends and family satisfaction testing.

Clinical Commissioning Groups

Under the Health and Social Care Act, devolved management of most hospital services, including emergency and urgent care, has been devolved to regional Clinical Commissioning Groups (CCGs). These clinically-led bodies include the GP practices in their geographical area. They operate by buying healthcare services, including elective hospital care, rehabilitation care, urgent and emergency care, most community health services, mental health services and learning disability services. The aim of the CCG system is to give GPs and other clinicians the power and autonomy to shape commissioning decisions for their patients, consistent with local priorities and values and informed by their clinical expertise.

As originally established, CCGs did not have any responsibility for managing the budgets and performance of primary care practices, which are managed by NHS England. In November 2014, however, CCGs were invited to become co-commissioners of primary care in their areas, and some CCGs have now either taken over complete control of primary care commissioning, or jointly manage it with NHS England.

Although many clinicians have analytical skills and conduct empirical research of their own, anecdotally, it appears that most CCGs likely do not have dedicated teams of analysts or formal data analytical functions (there is no existing inventory of the analytical capacity in the CCGs to enable verification of this statement). We know there is lots of information on quality out there but don't know if all NHS staff, such as clinicians, commissioners and service managers, are equipped to access and analyse it for both operational and improvement purposes, nor do we know if it meets people's requirements, says one of NHS England's clinical leads.

NHS Digital

NHS Digital is the statistical and informatics arm of the health system and is responsible for the secure transmission and storage of patient-level data. NHS Digital publishes over a thousand indicators, ranging from quality to population health and treatment outcomes, and including examples like prescribing data by GP practice, hospital mortality rates, and incidence of female genital mutilation. In addition, it produces over 250 annual official publications on topics from GP earnings to drug use among school children. These reports are accompanied by data in formatted Excel spreadsheets and machine-readable comma separated variable (csv) text files. Anonymised population-level data is available both on the NHS Digital website and on data.gov.uk.

NHS Digital has a statutory monopoly over the collection of certain kinds of data, and over 300 professionals work on the collection and analysis of primary and secondary, mental health, community, population health, prescribing services, clinical indicators, indicator and methodology assurance, clinical audit, and social care data. Of this group, approximately 250 are classified as analysts, whose work is focused on this routine and statutorily-mandated data collection and the publication of statistical reports. To date, as a result of time and work pressures to create these statistical reports on past performance, NHS Digital has not made much use of predictive modeling to evaluate innovations, conduct experiments or design new models of clinical care. Nor have there been active attempts to share or and mash up data with that from other departments such as education, labour, and justice, or to assess health outcomes by other social variables. There is no time—and no mandate—to translate that into policy recommendations nor to do research, according to Dr. John Varlow, Director of Information Analysis at NHS Digital. In total, the organisation has 2000 professional staff, most of whom focus on information technology and the maintenance of NHS websites.

Health and Social Care Quality Improvement Partnership (HQIP)

The Health and Social Care Quality Improvement Partnership (HQIP) conducts data-driven audits of the quality of care provided by hospitals, GPs, and specialists, publishing the results as comparative league tables. It also publishes the data underlying these national audits by subject, such as the national hip fracture, bowel cancer, and diabetes databases. HQIP contracts with third parties with subject matter expertise to conduct the audits: the Royal College of Psychiatrists' Centre for Quality Improvement, for example, collected data about the quality of care received by people with dementia in general hospitals. HQIP's 30 staff members include highly numerate individuals, responsible for writing and evaluating the tenders for the audits and communicating their results.

Care Quality Commission

The Care Quality Commission is the independent group monitoring the safety, effectiveness, responsiveness and concern of hospitals, care homes, GPs, dentists, ambulances, prison care services, hospices, and home care agencies in England. Teams of inspectors, clinicians and patients (experts by experience) conduct inspections with input from the public and staff. The Commission publishes reports and ratings designed to increase accountability and help people choose between services. It publishes a complete directory of the places in England where care is regulated by the CQC, with the latest ratings for each service, and encourages third parties to use this open data to create new tools for patients and their families.

Department of Health

The Department of Health is responsible for government policy on health and adult so-
cial care matters in England, including funding for the healthcare system, and oversees
NHS England and HQIP at arm's length. Although the Secretary of State for Health must
answer to Parliament for NHS performance, NHS England is an autonomous public body
not subject to ministerial or political control and with its own governance bodies and
structures. The Department of Health employs approximately 150-180 statisticians, or-
ganisational researchers and economists across all policy units under the direction of
Department of Health Chief Analyst Tony O'Connor. It publishes abortion statistics,
mental health and local health profiles, and patient experience scores.

In collaboration with NHS Digital and the Cabinet Office, the Department of Health is
setting up a Centre of Excellence for big data and data science. The idea was initiated
under the government's Data Science Programme, set up in 2014 under the leadership
of Paul Maltby, to help the government make better use of data science in policy-mak-
ing and service delivery and address some of the ethical challenges surrounding the use
of big data in governing.

An additional 600 analysts sit across the Cabinet Office, representatives of whom have
recently begun to start meeting on a regular basis and to start the conversation around
developing programs and initiatives to foster data science training within the broader
public policy profession.

Public Health England

Public Health England (PHE) is an executive agency within the Department of Health.
It focuses on public health priorities, on protecting and improving the nation's health
and wellbeing, and reducing health inequalities. PHE has eight directorates, includ-
ing the Health Protection Directorate, Health Improvement and Population Health,
Knowledge and Intelligence, which includes PHE's statistical operations. This knowl-
edge unit focuses on "delivery of a new national evidence and intelligence service that
supports transparent assessment of need; tracks performance and progress against
key outcomes; in partnership with the National Institute of Health and Clinical Excel-
lence makes available authoritative assessments of interventions to improve and pro-
tect health and supports the development and evaluation of innovative approaches and
techniques; enables the agency to develop profound knowledge in relation to the pub-
lic's health and the possibilities for improvement; and secures Public Health England as
an evidence and intelligence led organisation, ensuring a consistent high-performing
approach to the collection, collation and quality assurance of data, including in relation
to disease registration and the efficacy of drug and alcohol treatment."[7]

---

7 *Department of Health, Structure of Public Health England: Factsheet (26 July 2012), available at* *https://*
goo.gl/DXceup

Public Health England is responsible for researching, collecting and analysing data to improve our understanding of health and come up with answers to public health problems. Although NHS Digital has primary responsibility for data collection, Public Health England has statutory authority to use that data for scientific purposes. Public Health employs 5,000 full-time equivalent staff, comprising mostly scientists, researchers and public health professionals and, because it does a great deal of genomics and cancer research, Public Health England also has significant computational infrastructure. Hence, for example, Public Health England runs the impressive UK Cancer Data Dashboard (*https://www.cancerdata.nhs.uk/dashboard)*, which is a partnership between NHS England, Public Health England, and the non-profit Cancer Research UK. However, for the most part, the Public Health infrastructure is not connected to operational functions of the NHS. Its infrastructure is used to further population-level epidemiology and basic science, which has an indirect and more attenuated impact on improvements in patient care.

Other actors in the UK health and social care space

There are also a wide variety of individuals and organisations outside the NHS that engage with health data. This includes patients and members of the public; external supervisory and regulatory agencies such as the Royal Society for Public Health; watchdog organisations such as media and citizen interest groups; businesses and commercial interest groups and vendors; university and NGO-based academic researchers; entrepreneurs and app makers; foundations and philanthropists.[8]

8   *Royal Society for Public Health, accessed 30 March 2016,* https://www.rsph.org.uk/

Milestones in Data Use at the NHS

This section provides some background on the growth of data gathering practices within the NHS in order to illustrate the increasing demand for data analytical talent.

The 1980s-1990s
FOUNDATIONS FOR DATA COLLECTION

- Systematic collection of data in the NHS can be traced to the 1980s, but it wasn't until the 1990s that a robust data collection framework was established. The decision to upgrade and refine the way data was collected in the nineties was driven not so much by a desire for greater transparency, however, 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 policies, and were designed to foster competition inside the NHS, separating out 'providers' from 'purchasers'.

- It was during the 1980s, too, that transparency began to be understood globally as a tool for better governance. However, despite growing conceptual recognition of the value of transparency, it was really around the turn of the millennium, and particularly with the advent of new technologies and an Internet-led data boom, that the true potential of data began to be recognised within the NHS.

2000-2009
DATA COLLECTION AND STANDARDISATION

- The early 2000s saw a substantial increase in the number and types of datasets that NHS organisations were required to collect. These datasets included, for example, information on waiting times, performance evaluation of general practitioners, registers for diabetes and strokes, and a system for collecting and reporting adverse events to the then National Patient Safety Agency.

- In 2005, the NHS Information Centre was established as the authoritative repository of NHS health information, making a wide range of data available to providers, researchers and patients. It took over from the NHS Information Authority and was later replaced in 2013 by the HSCIC.[9]

- 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 45 million monthly visits and studies have indicated its important (and largely positive) role in changing patient behaviour.[10]

---

9   *Recently reborn as the Health and Social Care Information Centre (HSCIC) and subsequently renamed NHS Digital in July 2016.*

10   *Nicolaus Henke, Tim Kelsey, and Helen Whately, Transparency the most powerful driver of health*

2010-2013
OPEN DATA

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[11].

In 2011, 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."[12]

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, the Act required the Health and Social Care Information Centre, now known as NHS Digital, to publish all except for specific categories of the data it possessed (in anonymised 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."[13]

The NHS appointed Tim Kelsey as its first National Director for Patients and Information in July 2012. The position combined the responsibilities of a chief technology and information officer with oversight of patient and public participation and communications. Kelsey held the position until December 2015.

On 5 July 2013, the Secretary of State for Health announced the creation of Genomics England, a wholly owned company of the Department of Health. The company was established to deliver the 100,000 Genomes project, which will sequence 100,000 whole genomes from NHS patients with rare diseases, and their families. Patients with

---

care improvement? 2011, accessed February 23, 2014. https://goo.gl/LpbE3k

11   Open Data White Paper: Unleashing the Potential, 2012, accessed 28 March 2016, https://goo.gl/u1JsXs

12   Prime Minister David Cameron, speech on life sciences and opening up the NHS, 2011, accessed 23 February 2014. https://goo.gl/a8vjeo

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

common cancers can also take part on a voluntary basis. The genomic sequences will be combined with NHS medical records to help researchers deliver improved diagnosis and treatment.[14]

2013-present
PARTNERSHIPS AND PILOTS

In September 2013, the NHS launched the CareConnect Pilot, a phone and web service that allowed citizens to interact with the NHS, get information, and provide real-time data through feedback and other mechanisms. It has since been replaced.

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, recognising the importance of establishing a collaboration to allow both countries to learn from the experiences of the other, and to align their approaches.

Also in January 2014, the NHS announced the launch of _care.data_, a programme that aimed, in its first phase, to securely connect information together from a variety of healthcare settings such as general practices, hospitals and care homes to be used for purposes beyond direct care.

The _NHS Five Year Forward View_ was published on 23 October 2014 and sets out a new shared vision for the future of the NHS based around the new models of care, including innovation through technology and the use of data. As it relates to the use of data analytics the roadmap proposes: "Working with NIHR and the Department of Health we will expand NHS operational research, RCT capability and other methods to promote more rigorous ways of answering high impact questions in health services redesign. An example of the sort of question that might be tested: how best to evolve GP out of hours and NHS 111 services so as to improve patient understanding of where and when to seek care, while improving clinical outcomes and ensuring the most appropriate use of ambulance and A&E services."[15]

In November 2014, the National Information Board released _Personalised Health and Care_, a plan to improve patient outcomes through data and technology by building on the _NHS Five Year Forward View_'s vision. The plan laid out a proposal to "give care professionals and carers access to all the data, information and knowledge they need", including real-time digital information on patient health and care by 2020 for all NHS-funded services, and comprehensive data on service outcomes.[16] " The NIB brings together national health and care organisations from the NHS, public health, clinical science, social care and local government, along with appointed independent representatives to develop the strategic priorities for data and technology.

14   _Genomics England is delivering the 100,000 Genomes Project, accessed 30 March 2016, https://goo.gl/ysxWqd_

15   _NHS Five year forward view, 2014, accessed 28 March 2016, https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf, pp 34–35._

16   _Personalised health and care, 2014, accessed 29 March 2016, https://goo.gl/xWZUI1_

On 13 November 2014, the Secretary of State for Health appointed Dame Fiona Caldicott to the new position of National Data Guardian for health and care, with responsibility for overseeing the security of personal medical information.[17]

In September 2014 the Secretary of State for Health launched the MyNHS platform designed to bring key information about health, social care and public health together to further support the ability of transparency to bring about meaningful improvements

In March 2015, the HSCIC, who are now known as NHS Digital, released Information and technology for better care, its data strategy for 2015-2020. This committed the organisation to five objectives: ensuring the protection of individual data; establishing shared architecture and standards for the healthcare sector; implementing technology and data services to meet local and national needs; helping healthcare organisations plan wisely and maximise the value of their information technology investments; and opening, using, and analysing more data about the sector.[18]

In September 2015, the Care.Data programme was paused further to the Secretary of State for Health's announcement that the National Data Guardian for health and care would carry out a review of data security, consents and opt-outs, including proposing what choices should be given to people about how their information is used for purposes beyond direct care Further development of Care.Data paused in September 2015 over confidentiality concerns.

In December 2015 Matthew Swindells was announced as the new Director for Commissioning, Operations and Information for NHS England

In March 2016, HSCIC appointed NHS informatics specialist Daniel Ray as its first Director of Data Science, to head its newly-created centre of excellence for big data and data science. The Centre was established jointly by HSCIC, the Department of Health, and the Cabinet Office as an initiative of the government's Data Science programme, which was intended to facilitate data-driven policymaking.[19]

Published in July 2016, National Data Guardian's review of Care.Data recommended a different consent and opt-out model to that which was planned and encouraged the Government to consider the future of programme. In light of these recommendations, NHS England closed the programme and its ambitions were realigned to the National Information Board, a collection of national health and care organisations and independent representatives working together to put data and technology safely to work.

In July 2016, NHS England announced the appointment of Professor Keith McNeill as NHS Chief Clinical Information Officer and Will Smart as NHS Chief Information Officer. Keith McNeill has taken over Tim Kelsey's role as the chair of the National Information Board.

17   Department of Health, National Data Guardian appointed to safeguard patients' healthcare information, 13 November 2014, accessed 29 March 2016, https://goo.gl/HpfnTP

18   Information and technology for better care, 2015, accessed 27 March 2016, https://goo.gl/BKy827

19   Sooraj Shah, HSCIC appoints director of data science, Computing, 8 March 2016, accessed 30 March 2016, https://goo.gl/fAon8N

Current Shortcomings in Data Capabilities within the NHS

The NHS must employ people with data and analytic capacity and marshal them effectively to realise the benefits that data can bring. But does the NHS have the data and analytic capacity that it needs to make use of its own data? Some of the questions we have asked during our work with the NHS are:

‣ What is the unmet demand for data skills in the NHS?

‣ Are missing skills identified, and is there a follow-up to hire the necessary people?

‣ Are people with the necessary skills being motivated to remain in the organization, and are their skills tapped when they would be of value?

‣ What opportunities are there for engaging the public outside the NHS to fill in the gaps in skills?

The *NHS Five Year Forward View* and the *National Information Board Framework for Action 2020* describe how the NHS plans to transform health and care services to ensure high quality care for patients and good value for money for the taxpayer. Both strategies acknowledge the fundamental role of data in achieving these aims.

In 2014, NHS England and The Governance Lab at New York University created a blueprint—called *The Open Data Era in Health and Social Care*—to assess and accelerate the use of open data in healthcare settings.[20] The blueprint identified four key challenges toward more data use within the NHS:

‣ Cultural and institutional barriers

‣ Privacy

‣ Standards and interoperability

‣ Good analysis

The Blueprint stated that a meaningful data strategy should "focus on developing not only technical capacity within an organisation, 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.' Training, education, mentoring and general human capacity building are all part of this process, and essential components of any effort to increase data transparency." [21]

Complementing the lessons learned from our previous work, we conducted numerous interviews and additional research to identify where human data capacity currently resides and how it is matched to need, including how data science and domain experts interact. Our analysis has resulted in the following key observations about the way the NHS uses information.

---

20   *The Open Data Era in Health and Social Care*, accessed 30 March 2016, *http://thegovlab.org/nhs/*

21   Shah, Shvetank, Andrew Horne, and Jamie Capella, *Good Data Won't Guarantee Good Decisions*, *Harvard Business Review*, 2012, accessed 30 March 2016, *https://goo.gl/s7T9Rt*

There is a great demand for data and data analytical capacity

In our initial *survey* (see Appendix 2) conducted among a small sample of Expo 2015 attendees, NHS England asked respondents if they needed to use data in their work, and if they felt they lacked the training and skills to do so, or lacked access to those who did. The responses were resounding:

▸ Over 81 percent of those responding to the survey, and the majority of those we interviewed, said data would improve the quality of care. A full two-thirds of survey respondents said that data would drive improvements in population health.

▸ The survey group, comprised primarily of policymakers from NHS England, said they wanted access to more statisticians and analysts, and to form more teams combining clinicians with statisticians and health informatics experts.

▸ Without exception, those the GovLab spoke with said more data science and analytical capability were essential. Interviews reveal a near-universal perception of a lack of capacity, or failure to organise analytic capacity, to draw insights from data for improved outputs and outcomes. Data science expertise or experience is perceived to be in short supply or difficult to acquire, both internally and externally.

Resources in data analytical capacity are unrecognised

Knowledge about the workforce and how it is organised to meet the demand for analytical skills is a vital missing ingredient to realising performance improvements, better quality clinical experiences, more equitable treatment across the system, and efficiency gains. The NHS does not keep human resource records about the knowledge of data science among its staff, such as the categories described by our taxonomy. NHS Digital's own copious statistics don't include information about who works at HSCIC and what they know. There are no analytics about data science skills at NHS Digital or across the wider NHS.

But identifying who knows what even inside bureaucracies is hard. Within most government institutions, titles such as "director" and "manager" disclose little about lived experience. Projects are branded as institutional achievements, making it difficult to identify individual contributions and who knows what. People with statistical and computational training might very well be deployed doing jobs that do not utilise these skills. Except for the clinicians, who are regulated by a different body, there is very little systematic knowledge of the talents, skills and abilities of those within the NHS.

"Data analytical capacity" is in short supply

There are approximately 150-180 statisticians, organisational researchers and economists distributed across all policy units from obesity to cancer to hospital efficiency within the Department of Health. Across the Cabinet Office, there are a total 600-800 people with numerate backgrounds. NHS Digital in Leeds has 300 analysts

working on patient level data focused entirely on the collection, transporting, storing, analysing and disseminating the nation's health and social care data to comply with statutory obligations.

Even as NHS England has developed a much clearer understanding of the data it collects, uses and publishes, there is no systemic picture of where the whole NHS stands today in terms of its analytic capabilities, available skills and roles.

Although NHS England has a corps of 150+ talented analysts and tens of thousands of statisticians, social research experts, economists, and data savvy people across the health and care system, demand is outweighing supply of expertise, especially in the latest big data tools and techniques that turn data into improved health outcomes.

Data scientists with this knowledge are in low supply and high demand everywhere in the public sector, which competes with industry to attract the best and brightest. NHS is beefing up its capacity in data science, but must develop this skillset faster and on a larger scale if it is to realise the benefits of using data to improve people's lives. Although there has been a great deal of attention to the availability of data, opening of population-level data, and giving citizens back their own health care data, short shrift has been given to strategies for using that data, whether by the NHS to innovative or by patients to achieve better wellness.

All of the investment in creating the infrastructure for data access needs to be complemented by investing in the know-how and talent needed to draw insights from all this data and to manage its responsible use.

Previously, the public sector could compete with the private sector by offering job security in lieu of higher salaries but, under current circumstances, that is no longer the case.

An anticipated budget shortfall of thirty percent over the next three years, mean the NHS is no longer the secure place to work that it once was and analysts are subject to the same restructuring and cuts as the rest of the NHS. The health and care system is at risk of losing the data and analytical capacity it has, report many of the analysts interviewed, as a result of low morale and the perceived absence of adequate recognition. In contrast, growing numbers of private firms understand well how analysts add value to an organisation.

The desire to retain people against these pressures has given rise to the practice of moving people between teams. Rotation is common practice at the Department of Health, NHS Digital and NHS England, where human resources is moving toward the creation of professional 'pools' to ensure more fluidity in assignments. "They will be in survey design one week and in prescribing another week," says Dr. John Varlow about the analysts in his team. New resource managers in NHS Digital, for example, will be responsible for career development for the personnel in her pool. So in addition to an assignment manager, who is triaging who does what work based on expertise to create work teams, there will be a constant pressure to put professional development first when designing assignments.

The current emphasis on performance analytics needs to expand to predictive analytics, simulation and modeling

Most analytical capacity is currently dedicated to performance measurement and compliance activities—pursuing mundane data manipulation or what the Health Foundation refers to as 'lifting and shifting'—rather than to the more advanced analytical and inter-disciplinary modelling work needed to innovate and improve how things work and that is becoming more common in the defence and other sectors.[22] For example, at the University of Chicago Data Science for Social Good, the university is partnering with local agencies to develop predictive models that can flag individuals who are frequent utilisers of multiple public systems such as mental health and criminal justice. By using advanced statistical and computational methods, University of Chicago Data Science for Social Good will analyse behaviour patterns to understand what factors best predict interactions with these systems and develop models so employees of these systems may reduce future interactions while providing quality services.

"Analytical modelling isn't the norm," says Martin Caunt, Senior Analytical Lead at NHS England and Head of National Operational Research in the Operational Research and Evaluation Unit. Or, as Adam Steventon, Chief Data Scientist of the Health Foundation recently wrote: "Our best guess is that around 20,000 people use statistics as part of their day-to-day job in the NHS in England, most of whom are focused on financial management rather than improving quality more generally."[23] There is a real gap in predictive modelling capacity and the skills needed to know how to do things differently. The greatest demand is for the models that make it possible to test whether making a change leads to different outcomes. This is relatively simple to do and adds massive value, assert the analysts to whom we spoke. "A GP on the ground isn't interested in what has happened but what will happen," says John Varlow. Furthermore, the job of routinely collecting data for statistical purposes, as distinct from collecting data for a specific research project, creates challenges when it comes to using this data to drive learning and improvements.[24]

Especially in the face of budget constraints and the need to do more with less, these data analytical skills, which can point the way to innovations and improvements, are an especially valuable investment toward becoming more cost effective. While backward facing analysis of unnecessary variation and wasteful spending can help to cut costs and improve efficiency, far bigger savings will be realised from predictive modelling of innovative models of care and new ways of working. These can be accomplished only

---

22   Martin Bardsley, *Understanding analytical capability in health care Do we have more data than insight? Health Foundation, December 2016, 5.*

23   See Adam Steventon, *The Era of Big Data?, The Health Foundation blog, 18 Jan 2016, accessed 30 March 2016,* *https://goo.gl/pqXuJO*

24   Sarah Deeny and Adam Steventon, *Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data, BMJ Quality and Safety, doi:10.1136/bmjqs-2015-004278, accessed 30 March 2016,* *https://goo.gl/fvSPOa*

through collaboration between clinical and subject matter experts, on the one hand, and those with the skills to do predictive and prescriptive analytical work on the other.

Predictive analysis, in turn, emphasises maximising wellness and health outcomes.[25] "How professionals put data to best advantage for their patients at the moment seems to me something that is undervalued by health systems," says Tim Kelsey, former NHS National Director for Patients and Information. "There has been an historic failure to invest in those capabilities, which are at the root of personalised and precision medicine. The ability to act effectively and creatively in the clinical setting will depend upon a capacity for data analysis that is not necessarily clear to people yet."

That capacity is in short supply.

25   Eric Topol, The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Healthcare, Basic Books, 2013.

USING DATA STORIES TO UNDERSTAND THE NEED FOR DATA ANALYTICAL CAPACITY

The need for greater analytical capacity within the NHS is evident. This need is pervasive, general, and spread across every corner of the NHS. Understanding and ultimately resolving such a general need can be a challenge. In this section, we use stories to help break down the NHS's requirements into discrete analytical categories. These stories, based on real-life data use and decisions within the NHS, can help us better understand who is accessing data, how they are using it, and what changes could lead to greater innovation and efficiencies across the NHS.

We are woefully inadequate at translating such analysis into improvements.

Dr. Danny Keenan

STORY 1

Supporting front line decision making by hospital managers

The Healthcare Quality Improvement Partnership needs more analytical capacity to translate audit insights into recommendations for hospital directors.

Britain's National Health Service set up the *Healthcare Quality Improvement Partnership (HQIP)* as a central location to coordinate improvements in the delivery of health care in England and Wales in 2008. An independent organisation led by the Academy of Medical Royal Colleges, The Royal College of Nursing and National Voices, HQIP has conducted forty nationally funded audits and acts as an advisor on another twenty-five designed to compare and contrast hospital and physician outcomes across England and Wales.

For example, there is an HQIP audit on stroke, hip fracture and five different cardiovascular audits as well as assessments of cancer and ophthalmology. Every year, HQIP adds new studies, particularly in areas identified as having problems, such as mental health and maternal and foetal health. Whereas previously the results of these audits were confidential (league tables simply showed Hospital A, B, C), now the public can compare doctors and hospitals by name to make more informed choices about their own care. Institutions can also spot outliers requiring attention.

Each audit is designed in consultation with patient groups and public representatives as well as with expert clinical input. In many cases, the membership association of the relevant profession conducts the audit because of its familiarity with the clinical area of care.

The risk algorithms used to evaluate performance are revised every year, taking into account new indicators published by third party research organisations, such as the National Institute for Outcomes Research (NIOR). The data isn't perfect, says HQIP Director Dr. Danny Keenan. No algorithm is going to take account of the challenges of

treating a patient who is a Jehovah's Witness who will not accept blood. The algorithms do not distinguish between a neurologist, for example, who chooses to treat only accident victims and trauma patients, most of whom are likely to die before he lays a finger on them, and one who treats only elective cases under ideal conditions in a hospital setting. Despite their imperfections, these studies help to give some picture of what's happening in healthcare across the nation.

Managing the HQIP process demands analytical capacity both inside and outside of the NHS. NHS England will propose an idea for an audit. Then there's a formal meeting of all stakeholders in that field. In this preliminary meeting they determine whether it is an area amenable to doing an audit, and if so, the HQIP drafts and evaluates a tender for the evaluation. Because the first year's audit is generally fairly crude and leads to more learning, the audit needs to be improved before the following year.

Results are ultimately provided in several formats. A number of reports can be downloaded from the web site. However, a hundred-page report is hard for a clinician or a patient to turn into productive action. "What's needed above all," says Keenan, "are the two pages of recommendations for what a hospital manager and a patient should do differently as a result." National clinical audit data is collected and published regularly but is not always optimised by local clinical teams to improve care for a variety of reasons.

To provide more concise and actionable information, HQIP is moving toward the creation of dashboards with 3-5 recommendations against which hospitals can grade themselves. A number of activities are proposed in their *quality guide*, some familiar (such as root cause analysis) and some more tailored to their audience, such as "Plan, do, study, act" (PDSA) cycles or "situation, background, assessment, recommendation" (SBAR).

"We have the expertise to do the analysis," Keenan explains, "but we are woefully inadequate at translating such analysis into improvements. What's the take away for the hospital or the medical director? They cannot understand what they have to do."

> *There are many people who crunch data, but few who know what it means...We are achieving 1% of the potential to improve people's lives.*
>
> **Dr. Geraldine Strathdee**

STORY 2

Translating fingertips mental health data into local strategies

Turning data into interventions into crisis scenarios requires more analytics.

Dr. Geraldine Strathdee is NHS England's former National Clinical Director for Mental Health. She continues to chair the Public Health England and NHS England Mental Health Intelligence Network, which was responsible for launching the Fingertips Mental Health *dashboard*. Fingertips uses anonymised, population-based data to provide a profile of social determinants of mental health, predicted incidents and prevalence rates, numbers and rates of identification, quality of care given, access and wait times for psychosis and other common mental health conditions in every locality. Fingertips brings together forty-two national open datasets, including census, equality, employment, educational achievement, crime rates, psychiatric morbidity survey, suicide rates and more to create a detailed picture of a place and a community when it comes to mental health.

Especially in inner cities, Strathdee points out, there is a tremendous need for these kinds of early warning systems, if local councils and healthcare providers are to put preventive measures in place. Without them, "you spend sixty percent of your budget on psychosis and people end up in hospital beds, homeless or in prison," when use of this data to invest in prevention and early intervention would dramatically improve outcomes and value.

With the availability of such place-based population data, a community health care strategy can recognise the needs and the resources in a given place. For example, when a manager can spot that a majority of emergency room cases in some regions are due to alcohol-induced conditions in specific communities and that those are

triggering domestic violence, self-harm and suicide and other morbidities, the local hospital can work more effectively with the community, street triage, and police department to develop an early intervention approach that works.

But despite the availability of data, there is a shortage of people who know how to use it. "There are people crunching data in the NHS," she explains, "but they can't tell you what it means." There is a dearth of translators who can interpret the data, especially at the local level. Many government agencies want and need to use mental health data but lack the capacity and, above all capability, to do so. Thus, local plans are made without data or evidence, says Strathdee, wasting money and opportunity. Furthermore, a general practitioner, not a specialist, sees eighty percent of the people with mental health conditions. Yet primary care physicians and specialist mental health clinicians are not using this data even though they create it, each one spending an average of 52% of her time entering data. The data that is available does not yet translate into improvements on the ground.

"It's not enough to just give people the numbers," she says, "they also need to be told what the numbers mean." That means more people in more positions need to learn how to interpret this valuable data, especially at the local level. "We have a major capability level problem everywhere," says Strathdee.[26]

STORY 3

26   Shah Shventank, Andrew Horne and Jamie Capella, *Good Data Won't Gaurantee Good Decisions, Harvard Business Review 2012, accessed 30 March 2016,* https://goo.gl/glnX3o *(April 2012).*

New models of care

Culture and capacity for analytics hamper the launch of a key quality tracking program.

The *NHS Five Year Forward View* published in October 2014 sets out a new shared vision for the future of the NHS based around new models of care. To date, 50 vanguard sites across England have been announced that seek to experiment and implement this vision. In reference to data analytics, the *New Models of Care roadmap* outlined in the *Five Year Forward View* proposes: "Working with [National Institute for Health Research] NIHR and the Department of Health we will expand NHS operational research, randomised controlled trials (RCT) capability and other methods to promote more rigorous ways of answering high impact questions in health services redesign." An example of the sort of question that might be tested: "how best to evolve out-of-hours general practice and NHS 111 services so as to improve patient understanding of where and when to seek care, while improving clinical outcomes and ensuring the most appropriate use of ambulance and A&E services."[27]

The Operational Research and Evaluation (OR&E) Unit within NHS England is leading the evaluation of how those new vanguard sites are improving health and health care as well as supporting several transformation programmes and the wider work of NHS England.

The OR&E approach is based upon rapid cycle evaluation, built into the implementation

---

27 *NHS Five Year Forward View, "How Will We Get There?" Chapter 4, accessed 30 March 2016,* https://www.england.nhs.uk/ourwork/futurenhs/nhs-five-year-forward-view-web-version/5yfv-ch4/.

of the programs. Sharing the lessons learned from these evaluations and experiments regarding the enabling conditions can help improve the likelihood of rapid transformation across NHS England. Particularly promising is the creation of the Health Data Lab (see Recommendation 2 below), a collaboration between NHS England and the Health Foundation to use nationally available data to rapidly apply advanced statistical techniques to construct counterfactuals to measure the impacts of specific inteventions within NHS England's transformation programmes, in particular.

At the time of writing, the Lab was under construction. The metrics used for the Health Data Lab will be dependent upon the specific interventions examined and the expected outcomes they will achieve. Martin Caunt, Senior Analytical Lead, NHS England and Head of National Operational Research in the Operational Research and Evaluation unit, shared three important perceived challenges:

There is limited access to data that is current (as opposed to annual reports) and that would make it possible to follow the patient across settings (by linking data sets).

Human analytical capacity. Most data analysts are evaluating performance, hardly anyone does predictive analytics, yet this has the biggest opportunity for impact.

There is a lack of a demand or culture where modeling is the norm, as it is in the defense sector, where simulation and predictive modeling have become the default.

STORY 4

Supporting policy making at the national level

Current demands on analysts' time leave little scope for the rapid ability to mash up and draw insight from data across multiple domains.

Imagine a policy director in the Department of Health wanting to model the likely impact of a change in policy, such as raising the alcohol tax in Britain. Imagine that the Prime Minister asks the Department of Health to develop workable strategies quickly for how to improve poor rates of dementia diagnosis in England that lag behind those of Ireland and Scotland, not to mention Europe. Or imagine that the Secretary of State wants to know how to improve the way key citizen services are delivered nationwide, which calls for combining education, employment and health data.

To create its November 2013 *Dementia: a state of the nation report* and its online tools for checking the availability and performance of local dementia social care services, the Department of Health had to pull together local and nationally available data on dementia services from across the country. But its capacity to do so is limited by the shortage of analysts throughout the health system.

For the most part, says Tony O'Connor, the Chief Analyst in the Department of Health, that expertise is focusing on compliance rather than operational analysis. Consistent and reliable access to this expertise is also lacking because, in order to retain staff, there is an emphasis on rotating people across units to aid their professional development.

The members of his immediate team, who are focused on scanning the horizon for new issues, such as microbial resistance or the impact of the dark web on health or doing more near term policy analysis and design in response to a government priority, number only six to eight. Their ranks are augmented by rapid research contracts of one to four months, but the budget for those contracts is relatively small. Compared with the larger pot of money spent on clinical research, research on institutional performance— what some call improvement science—is negligible.

TEN RECOMMENDATIONS FOR BUILDING INSTITUTIONAL READINESS AND TAPPING DATA ANALYTIC CAPACITY INSIDE AND OUTSIDE THE NHS

Given the potential for cost savings and better health outcomes down the line, a strong case can be made for meeting the demand for data analytics by hiring more staff. But given the underdeveloped data culture within the NHS, the lack of understanding of how data can help to improve health outcomes and, above all, the constrained budgetary climate with across the board cutbacks, our proposed recommendations for meeting the data analytic talent gap focus on alternatives to full-time hiring that we describe in detail in this section.

These recommendations fall into four categories.

▸ Using new technology to coordinate *distributed talent* already present in the NHS, including project marketplaces.

▸ Using new technology such as talent banks and skill finders, to find *talent hiding in plain sight*—namely those with the relevant skills but who are not classed as analysts and match them to projects.

▸ Using expert networks to connect with empirical *expertise outside the NHS*.

▸ Creating *cost-effective incentives* to bring talent in from outside, including prize-backed challenges and foundation-funded fellowships.

Figure – The Skills Marketplace Homepage

RECOMMENDATION 1

The NHS should build a project marketplace like the environmental protection agency's one EPA Skills Marketplace and help supply find the demand [28]

Model: One EPA Skills Marketplace

The United States Environmental Protection Agency, a Cabinet-level Department, is facing the same challenge as so many public institutions, namely how to discover and make better use of the skills employees have to advance the agency's priorities. Previously people have worked on projects in their own divisions. The Skills Marketplace encourages cross-agency interaction and the creation of teams focused on outcomes rather than affiliation.

Why is this interesting for the NHS?

Skills Marketplace relies on an open call to employees to participate. It may be an interesting model for finding and better deploying the analytical capacity across NHS by helping those with analytical (and other) know how find projects to which they could contribute. It could respond to the need, described by Dr. John Varlow, Director of Information Analysis at NHS Digital, of making sure the right people are in the right place at the right while ensuring that they get professional development.

One EPA Skills Marketplace

It was accepted wisdom that organisational "silos" mitigate risk. But a recent Institute for Government study confirms what most have come to believe: that in the Internet era silos hamper the effective delivery of policies and services and lower employee morale.

To remedy the problem of silos, the Environmental Protection Agency (EPA) Skills Marketplace platform encourages managers to post projects and recruit volunteer participants to the end of fostering collaboration across the 13 headquarters offices and 10 regional offices of the agency. Employees have an incentive to sign up for projects, to learn about other parts of the organisation; to find work assignments that match with their skills and professional development goals; and to increase the organisation's productivity.

Begun in 2011, two years before President Obama published his second-term management agenda that called for "unlocking the full potential of today's Federal workforce and building the workforce we need for tomorrow," this "classified adverts" site for employee projects launched in 2014 in an effort to boost employee engagement and productivity.

Figure — List of One EPA Skills Marketplace Projects

How It Works

Supervisors post project descriptions together with requests for help with specific assignments. These projects do not exceed one year in duration and can be completed without the need to travel or relocate from the employee's home office. All employees can see all listings and all those who are interested, regardless of agency or division, can apply to any of these projects with prior permission from their home office supervisor. Borrowing a page from Google's 20% time initiative, designed to encourage employees to spend time on innovative projects other than their home office assignments, the platform enables employees to develop ideas with diverse colleagues without the cumbersome transaction costs of a formal re-assignment. Although work is performed for managers other than one's usual supervisor, projects count toward one's annual performance evaluation.

Designed through interviews with agency leaders, executives, supervisors, and its several employee unions between 2011 and 2014, the program was piloted across seven Environmental Protection Agency program and regional offices. In an evaluation, 95 per cent of the participants in the pilot said they would recommend it to others and 100 per cent of home office supervisors said they would support their staff in participating again. During just over a year of agency-wide operation, managers have posted close to 400 projects, with more than 70 per cent of projects receiving applications from 627 employees. Of those, 388 were selected to work on the projects.

Outcomes and Impact

The greatest evidence of success of the Skills Marketplace is that other federal agencies have now followed suit under the auspices of the Office of Personnel Management. The broader umbrella initiative called GovConnect is a suite of projects designed to break down silos, including EPA's Skills Marketplace and State Department's FLEX Connect. The aspiration is eventually to go beyond pilots to foster government wide collaboration.

Wing Yeung, a program analyst with the Office of Solid Waste and Emergency Response, said she was looking for a challenging project different from her daily routine. She signed up to use her statistical and analytical skills on a social media project. The experience she gained from the project led to an offer to work on a new project. As a result, her resume now showcases diverse applications of a core skill set. **Source**: FEDmanager.com, _Skills Marketplace: Building One EPA Through Professional Development_, 17 February 2015.

The Environmental Protection Agency is assessing the program's outcomes by conducting an internal survey among participating employees and supervisors for a quali-

tative analysis to understand what worked and what didn't, and a quantitative analysis of the number of projects posted and employee participation statistics. They also surveyed the employees who did not participate to find the reasons for non-participation. The majority cited a paucity of time after their regular projects as the main reason.

But these assessments provide little insight into the downstream impacts of the tool on project outcomes. Tracking the progress and outcomes of projects staffed using the Skills Marketplace would provide the agency and independent researchers more information about the effectiveness of the tool in actually solving the problem of breaking down silos in the organisation. Neither are they conducting any randomised controlled trials, staffing some projects using the tool and others without it, nor are they allowing managers to search for and reach out to employees to staff projects. The project leadership believes it is better to invest in incentives for employees to self-select.

Figure: Old Bailey, London

The NHS should build an NHS data lab modeled on the Ministry of Justice Data Lab to make better use of sensitive administrative data

Model: Ministry of Justice Data Lab[29]

The UK Ministry of Justice set up its Data Lab in 2013 to enable those in the public, voluntary and community sector who provide services to criminal offenders to obtain a robust evaluation of the effectiveness of their interventions.

Why is this interesting for the NHS? Like NHS Digital data storehouse, the Ministry of Justice's police administrative datasets are large, mostly clean and comprehensive data that must be held securely in a central location to protect privacy and confidentiality of personal data. However, by offering a "data analytical service," namely the option to *third parties* to submit requests for analysis (inside or outside) wishing to test the effectiveness of their interventions against this data, the Justice Data Lab enables what New Philanthropy Capital's (NPC) David Pritchard describes as "mass produced, low quality, quasi-experimental studies."[30] By developing a single standard method for accepting outside requests, it has been able to address a high volume of queries and provide the evidence needed to help other groups evaluate and evolve their offerings to better service the public. Although NHS Digital does have outside fellows who come in to "play with data," they must "leave their phones and notepads at the door and cannot take anything away with them," as John Varlow of NHS Digital explains.

---

29   *Thanks to Tracy Gyateng for her assistance with this case study.*

30   *Email to author from David Pritchard, April 16, 2016.*

The Data Lab model offers a useful addition to the NHS at low cost that could help make NHS more useful for testing the efficacy of interventions against patient outcomes. At the instigation of Julie Henderson, Head of Analytical Services at NHS Digital, NHS is making plans to undertake a proof of concept to test the Lab model in order to measure demand for such a service and to apply advanced statistical techniques to measure impacts against several metrics rapidly. In addition, NHS England is working with OR&E, Health Foundation and NHS England Temporary National Repository to create a health data lab to evaluate NHS Vanguard pilot sites.[31] The data lab model could be used for stakeholders outside and inside the organisation and any effort to build one or more labs should both identify the "customers" and be accelerated.

The Origins of Ministry of Justice Data Lab

In response to requests from criminal justice charities surveyed by NPC in 2012, the Ministry of Justice established the Data Lab in April 2013. The Lab is a secure facility within the Ministry staffed by three statisticians, who evaluate the recidivism rate of individuals who have received services and programs aimed at reducing their propensity to re-offend, such as interventions aimed at getting individuals into employment or prison education programs. The Lab allows those providers to obtain evidence on how the impact of their interventions differs from outcomes of a matched comparison group.

In the past, many providers of offender services have not been able to access re-offending data relevant to the offenders they work with. This has obstructed organisations from measuring the effectiveness of their rehabilitation in preventing participants from re-offending. This is similar to the situation of many voluntary sector bodies who would like access to health outcomes data but who either lack access or lack the capacity to make use of such data.

The lack of access to high-quality re-offending information has also prevented some organizations from learning from and improving the services they deliver; and has made it difficult, if not impossible, for them to demonstrate their impact to local authorities and other funders.[32]

To make use of the Justice Lab, an organization submits a minimum of sixty individual names of those who received services (smaller groups can pool their names) together with additional information about the interventions delivered. The organization must

31   Email to author from David Bull, May 9, 2016. See also, Tracey Gyateng and James Noble, Developing a Health and Care Analytical Service (HCAS): Summary (February 2016), available at

32   Ministry of Justice, Justice Data Lab: The Pilot Year, 13 March 2014, accessed 30 March 2016, https://goo.gl/01IRQF. Ministry of Justice, Justice Data Lab User Journey, April 2014, accessed 30 March 2016, https://goo.gl/xksf91. See also Fergus Lyon, Tracey Gyateng, et al., "Opening access to administrative data for evaluating public services: The case of the Justice Data Lab," Evaluation 2015, Vol. 21(2) 232–247, DOI: 10.1177/1356389015577507

have worked with the individual between 2002 and 2013 (this number is updated as new data comes in and might have been extended since the data of this writing). Full names must be given to facilitate accurate matching against the recidivism database. Therefore, to protect individual privacy, information must be emailed from a government account or using a secure Criminal Justice Secure Email system.

In return, the Justice Data Lab team matches these individuals to the re-offending datasets held within the Ministry of Justice and uses statistical modelling techniques to generate a matched control group of individuals with very similar characteristics. In the justice context, those characteristics generally include basic demographics and criminal history and may also (as of early 2016) also include education, employment, addiction, mental health, and benefit history. It is not possible, however, to match the control group on all characteristics.

The comparative analysis offered by the Justice Data Lab is based on matching two different groups that are as similar as possible, but have received different forms of intervention. Each group is similar in age, type of offense, length of prison stay, and other measures. But one group receives an intervention intended to re-integrate the offender into society and prevent re-offending. The organisation carrying out the intervention may measure direct outcomes of the intervention, while the Justice Data Lab measures the subsequent reoffending behaviour of all offenders sharing the same demographics. If these reoffending outcomes are better in the group receiving the intervention, that intervention can be flagged as a potential success and a candidate for replication in other jurisdictions.

New Philanthropy Capital, in making the case for a Health Data Lab equivalent, suggests measuring the impact of interventions against hospital admissions, readmission, length of stay, cost of care and other standard outcomes to accomplish a similarly robust and useful evaluation.

The Justice Lab sends the requestor a report and also publishes those results online. The report includes:

▸ One year proven re-offending rate and frequency of re-offending for the cohort of individuals shared with the Justice Data Lab

▸ One year proven re-offending rate and frequency of re-offending for a matched control group of offenders with similar characteristics

▸ A statistical comparison of the two re-offending rates, and frequencies of reoffending, with a conclusion indicating whether the service or intervention is associated with a change in re-offending behaviour

It is a central piece of the model that the analysis is compiled into a final report, which is posted on the Ministry of Justice Data Lab website. This public disclosure of effective interventions enhances the value-added of the service to the public and to funding organisations. At the same time, it can create a disincentive to programmes to participate unless

compelled to do so or unless everyone is participating. The more rigorous the evaluation design is, David Pritchard points out, the less effective a program will appear to be. Hence participating, absent a general mandate, might appear risky to an organization. Hence, although there are over 6,000 charities working in the health sector, NPC in its business case for a health data lab was able to pinpoint only 50 charities likely to be eligible to use such a service and 12 who affirmatively volunteered to take part in a pilot.

The Justice Data Lab is now able to provide information on the frequency of reoffending and the length of time to the first re-offence. It breaks down analyses by region and prison type. It provides additional detail on the characteristics of individuals in the analysis than it could do at the outset.[33] The analysis is compiled into a final report, which is posted on the Ministry of Justice Data Lab website.

At present, only limited indicators can be tested with a single defined methodology. It's a one size fits all service, allowing the Labs to perfect this offering with a small team and at low cost before expanding to offer additional services. So a service provider, for example, cannot ask for a statistical analysis of employability or educational attainment (yet). Nor are researchers invited to submit a request for testing a different hypothesis or applying an alternative inferential method or algorithm. But, as the Labs perfects its methods, it has been able to improve and evolve what its offerings. Frequency of re-offense, for example, was not originally measured. There are plans to bring additional data into the data sets that can be queried through the Lab.

Because of the model's simplicity, the Labs were able to output 173 analyses for social enterprises, public, and private sector organizations in England and Wales from April 2013–February 2016.[34]

By providing a report on recidivism for a cluster of specific individuals, the Ministry of Justice uses its administrative data to help others reflect on the effectiveness of their programs and understand how they can improve and develop their services. The hope is that such evidence of what works can ultimately lead to a reduction in crime in the future. Similarly, in the health context, such a service offers an easy way to make use of administrative data securely combined with the ability to learn more about the health impacts of diverse interventions.

The Justice Lab model provides an alternative to two undesirable extremes: on the one hand, openly posting personally identifiable data in violation of data privacy laws for anyone to query on the open web; and on the other hand, limiting access to such data only to MOJ internal users. Instead, the model provides a middle ground by which others can query data containing personally identifiable information about individu-

---

33   Jessica Mullen, *Five New Things the Justice Data Lab Can Tell You About Your Service*, 19 October 2015, accessed 30 March 2016, https://goo.gl/YsEjXE

34   *Ministry of Justice, Justice Data Lab Statistics March 2016*, March 2016, accessed 30 March 2016, https://goo.gl/JLhA35

als through the intermediary of the Justice Lab staff. The process enables a never-before-attempted way of using sensitive personal data for evaluation purposes while safeguarding individual privacy.

There are some similarities of this model to work done by Cancer Research England, which provides dummy data (simulacrum) to scientists to test algorithms, which the non-profit then runs itself against the real data.

The lab idea – a distinct organisational unit employing personnel with much-needed skills in short supply such as data science or design thinking or computer programming — is becoming a popular strategy to institutionalise innovation within government. Drawing on the private sector notion of the skunk works, these agile teams, which may often be short-lived, can often do what the public sector generally cannot. At the same time, leading by example, they offer a model to inspire greater innovation across the organisation.

More than a hundred labs or "i-teams" have sprung up in the public sector around the world. These innovation offices within government are designed specifically to be more open, collaborative, and experimental. Some focus on using ethnographic research methods to engage citizens in the design of services that they will subsequently use. Others focus more explicitly on the design and creation of new technologies.

The US federal government, too, has created such institutions for disseminating innovation. They function both as idea labs, to encourage employee and citizen innovation, and as new offices tasked with improving the administration's technological toolkit. The Health and Human Services Department's Idea Lab was established in 2013 to create a collaborative environment in order to encourage innovation within the department. The Lab boasts an *entrepreneurs in residence* program to learn from private sector innovation, along with programs to encourage employee innovation. The Lab hosts the Health and Human Services Ignite Accelerator, Venture Fund, and the Innovates Award to invest in and stimulate employee innovation. Although not specifically a data lab, what all these institutions have in common is an effort to institutionalize and scale more innovative and tech-driven practices.

The NHS should build an expert network like Health and Human Services' HHS Profiles and help the demand find the supply across the NHS

Model: HHS Profiles

Relinquishing the near-impossible task of hiring more data scientists, economists and organisational researchers during a time of budget cuts, health policy organisations can find the supply of internal analytical expertise and make it more searchable. At present, the NHS does not have comprehensive analytics about the workforce in its constituent bodies. With better data about who works in the NHS and what they know, including full and part-time employees, contractors, alumni and grantees, the NHS may be in a better position to mobilise existing talent more effectively. Such a platform is known as an *expert network*. The public sector has lagged behind the private sector in the use of such expert networks to understand its own talent better.

Figure – Department of Health and Human Services Logo

Developing a Model for Expert Networking Across Federal Government: The HHS Profiles Pilot

Jessica N. Hernandez and Nichole Rosamilia
Innovation and Technology Solutions Team, Food and Drug Administration
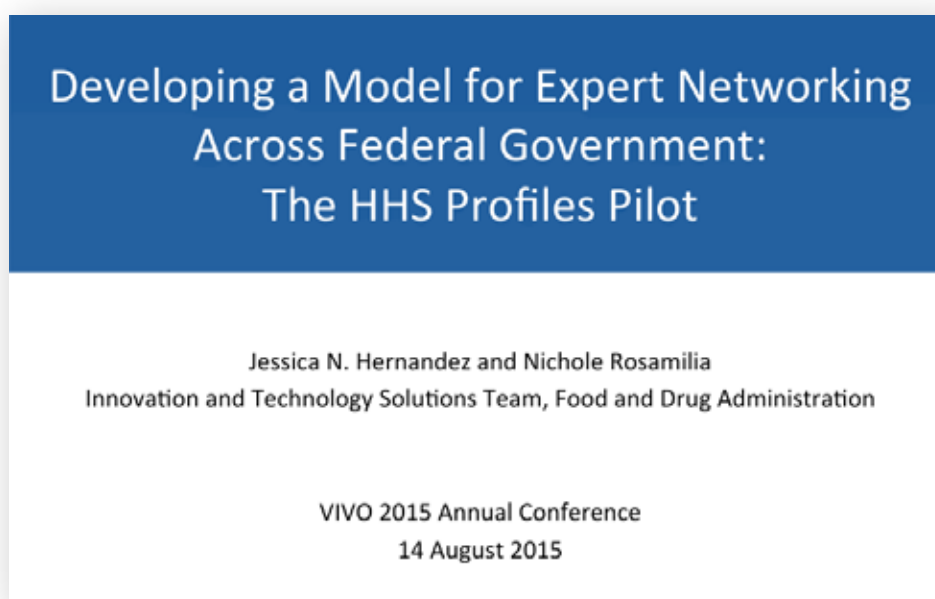
VIVO 2015 Annual Conference
14 August 2015

Figure: HHS Profiles

The American equivalent of the Department of Health – Health and Human Services (HHS) – has faced a similar challenge with its inability to quickly convene a qualified group of reviewers from among its staff, given the diversity of new medical devices coming to market. To overcome this challenge, Health and Human Services is building an expert network to match the supply of certain kinds of biomedical expertise across the whole of the Cabinet Department to the demand for it in the Food and Drug Administration, which is responsible for medical devices.[35]

Why is this interesting for the NHS?

The HHS precedent of using new technology to create a searchable database is particularly instructive because it addresses the data and information needs of the Food and Drug Administration by creating a database of the biomedical, informatics and other related areas of expertise of the much larger Cabinet Department (over 110,000 staff). In the HHS project, employee profiles are created by drawing from a number of data sources describing their skills and experiences. Managers responsible for the regulatory review of devices then use those expertise profiles to identify the most qualified individuals to participate in the regulatory review of a given medical device. Previously there was no rapid, systematic way for the ODE to identify capable reviewers within HHS, resulting on a reliance on the directory approach involving the usual suspects. Building on the HHS experience, the GovLab designed and deployed a prototype expert network as an experiment at NHS Expo 2015. (That experiment, including screenshots of the system, is described in detail in Appendix 3.)

---

35 *We would like to thank Jessica Hernandez and Nichole Rosamilia, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, for their assistance with this case study.*

The FDA's Center for Devices and Radiological Health (CDRH) manages the process of pre- market approval and post-market review of all medical devices. CDRH ensures that "patients and providers have timely and continued access to safe, effective, and high-quality medical devices and safe radiation-emitting products," and facilitates "medical device innovation by advancing regulatory science, providing industry with predictable, consistent, transparent, and efficient regulatory pathways, and ensuring consumer confidence in devices marketed in the U.S."

Every year, the Center evaluates thousands of devices of varying degrees of patient risk and complexity – many of which employ novel technology. To ensure the safety and efficacy of these products, the FDA is faced with the challenge of finding the right expertise to help it quickly and effectively assess new products. The pathway to regulatory review and compliance for low-risk items like tongue depressors is straightforward. But life-sustaining or high-risk devices such as pacemakers and breast implants require judicious and timely Premarket Approval (PMA) by those with the right know-how. If these medical devices reach patients without proper testing, there may well be very real human costs.

Unlike post-market reviews, the PMA process is undertaken exclusively by internal staff. Previously, the FDA had wrestled with the idea of allowing those outside the agency more broadly to review medical devices, but this foundered out of a fear of conflict of interest and undue influence. Hence the agency has to rely upon its own internal staff augmented by some outside experts, who are hired as temporary employees to man these review panels. As is common in many governmental practices, such as patent examination or regulatory drafting, the FDA is faced with the task of evaluating large quantities of complex information with resort only to a too-small pool of people lacking the requisite diversity and laboring under the strain of too much work.

Traditionally, reviews are undertaken by a set of "usual suspects" identified by staff from the FDA Office of Device Evaluation (ODE).

The Challenge

In its regulatory review efforts, the FDA confronts several challenges related to:

▸ **Agility:** Finding and convening a qualified regulatory review panel can take as long as nine months. The FDA has also seen a rapid increase in new medical devices being developed and submitted for review: In 2010 and 2011, an average of 49.5 high-risk medical device recalls were initiated, compared to an average of 24 over the three preceding years.

▸ **Expertise:** The current pool from which experts are identified is limited and may not include people experienced with the specific—potentially novel—technologies featured in innovative medical devices.

▸ **Diversity:** It is hard to identify people from diverse disciplines with relevant prior knowledge at each of the many stages of device review.

▸ **Complexity:** The regulatory review process lags behind the market in the ability to keep pace with advancements in technology. In addition, some devices require compliance with multiple sets of agency rules.

These challenges have real effects on the medical device industry, and potentially public health. Long review times are bad for firms due to the potential for significant financial stress after large up-front investments, as well as due to the opportunity costs sacrificed during the review timeline. Ineffective reviews are bad for the public because devices with the potential to kill or injure may be approved for widespread use. On the other hand, unnecessarily long review times for truly safe products are bad for the public, because under these circumstances life-saving products may take longer to reach patients.

As of 2012, it takes an average of 266 days for a device to pass through the Premarket Approval process. The average PMA review time for a medical device increases markedly for unique devices, which have no predicate devices. Unique, innovative new devices are reviewed over the course of, on average, 18.1 months—7.2 months longer than the approval time for the second such device. Particularly regarding new and unique devices, extended review times are somewhat expected given the danger to public health that would arise from unsafe devices reaching market. And indeed, longer approval timeframes currently correlate to fewer subsequent reports of adverse events. HHS Profiles is rooted in the belief that better targeting using the expertise within the FDA and all of HHS for reviews could lead to faster, industry-benefitting reviews that do not sacrifice thoroughness or safety.

Experimentation with HHS Profiles

As part of a broader effort across Health and Human Services , the FDA's Office of Science and Engineering Laboratories at Center for Devices and Radiological Health is launching an expert networking pilot project in response to these regulatory challenges. Health and Human Services Profiles aim to identify, quickly and intelligently, reviewers with specific areas of expertise for the pre-market review of medical devices. Under the administration of President Barack Obama, the FDA was partnering with the GovLab and its MacArthur Foundation Research Network on Opening Governance to develop and implement pilot projects to assess the effectiveness of using Profiles in this way.

Identifying and recruiting competent reviewers calls for creative and sophisticated data collection on these individuals. Thus, HHS Profiles imports and analyses "white pages" (detailed contact) information, publications, and other data sources to create and maintain a complete, searchable library of web-based electronic CVs for experts within HHS. HHS Profiles uses the Harvard Profiles Research Networking Software platform developed by Harvard Medical School with support from the National Institutes of Health. The platform's User Group includes over 300 members from institutions around the world, such as the medical and biomedical faculties at Harvard, Penn State, Boston University, and University of California at San Francisco.

The HHS Profiles pilot program—which will be launched in 2016 with funding from the General Services Administration's the Great Pitch investment contest—will evaluate several potential impacts of the introduction of this piece of expert networking software during the medical device review process. It is possible to compare the effects of using Profiles to promote participation against status quo methods to understand the impact of general matching, and of matching based on academic degrees and publications. Does targeting based on academic credentials have the downside of reaching those who are too busy to help and whose professional norms discourage such volunteer activity? Good academics may, for example, turn out to be bad participants because they are trained to contribute in ways that are unhelpful to regulators, for example, academics used to writing long articles are often not all that useful to busy government professionals.

The experimental rollout of the software will facilitate causal analysis of the software platform's impact, including its effect on review panel composition, time needed for a panel to be convened, time required for a panel to review a new medical device, and subsequent safety outcomes for all products reviewed. For example, product recalls and adverse event reports. The results of this pilot have the potential to inform and improve the process of regulatory review and will be relevant to the design of more effective regulatory review.

Moreover, adding empirical research into the rollout of HHS Profiles is a chance to run an experiment that goes beyond distinguishing between policies, to help determine how we make policy. The agile empirical manner of experimentation is meant to help develop replicable methods for studying governance innovations in the wild, and to accelerate the pace of research in government as a result.

Image by Jaakko H.

The NHS should build a web-based help desk like the World Bank's Skill Finder to accelerate the process of matching demand to supply

Model: World Bank Skill Finder

The World Bank has catalogued and made searchable the know-how of its 27,000 employees, consultants and alumni. To accelerate adoption, the Bank is testing its ability to target and pinpoint expertise by launching a help desk called Ask Water.

Why is this interesting for the NHS?

The combination of help desk with an expert network seems to be working well to match demand to supply.

World Bank Skill Finder Network and Ask Water Help Desk

In the mid-eighteenth century, Denis Diderot and Jean-Baptiste D'Alembert began publishing their twenty-eight-volume encyclopedia with contributions from enlightenment greats such as Rousseau, Montesquieu and Voltaire. The Encyclopédie endeavored to categorise and widely disseminate secular knowledge systematically in order to shift people's thinking toward a secular Enlightenment worldview. It is perhaps fitting, therefore, that Pascal Saura, a former scholar of Diderot and professor of philosophy has been among those leading the charge at the World Bank to catalog and make searchable the know-how of its 27,000 employees, consultants, and alumni.[36]

Begun in 2014, SkillFinder is an online expert and talent network. Originally intended to help employees find internal project peer reviewers across the organisation, like many tools however, SkillFinder has been "hacked." That is to say, in its first year of use, World Bankers are relying on the software more to find those they know to help with team building than to find those they don't know to conduct peer reviews. Managers and employees want to understand better the skills and skills gaps of those who sit across from them in order to inform their work plans.

What remains to be seen, however, is whether the Bank—one of the first major public institutions to deploy a company-wide expert network—can also leverage such data-driven tools to further its work of ending extreme poverty and promoting shared prosperity.

How SkillFinder Works

The World Bank's SkillFinder tracks expertise and experience across three dimensions: technical expertise, geography and clients, and business processes.

In an effort to yield data about people that is both comprehensive and trustworthy, SkillFinder's personal expertise profiles are culled from three sources:

‣ The tool scrapes data from institutional records, such as human resources data about projects and credentials. This provides a source of authenticated, "official" data.

‣ It also asks people to fill out free-form narratives and apply short tags that endeavour to give additional colour and flavour to descriptions of what people know.

‣ It asks colleagues to endorse their peers and bosses and adopts recommender and badging techniques common to such platforms in the private sector. These third-party endorsements, à la LinkedIn, corroborate self-assessments.

One of the challenges of building such an expert network is to organise information using a vernacular familiar to the institutional culture. The World Bank already possessed a shared vocabulary comprised of a taxonomy of development topics at the core of the Bank's mission on one hand, and a taxonomy of business functions common throughout most organisations on the other. It was able to build on these dictionaries to ensure that expertise would be more searchable using terms common within the Bank.

Accelerating the Search For Knowledge With A Help Desk

To accelerate Skill Finder's adoption and enhance its usefulness, the Water Practice within the World Bank is testing its ability to target and pinpoint expertise by launching a help desk called "Ask Water" to help people formulate questions, find experts and get answers. Saura explains that the goal of the helpdesk is to "enable a dialogue between knowledge and implementation by connecting operational teams to what the organisation knows." It makes it easier to access a "complex system of distributed experts, databases, search and analytical tools," including SkillFinder.

Staff can email the helpdesk with a specific question, such as "What is our approach to tackling urban groundwater challenges in megacities?" One of the six helpdesk person-nel use SkillFinder to reach out to internal and external experts, search databases and reply within a few days to a week. Interesting answers are also shared with the group on their internal platform. The help desk has received and processed around 30 requests in its first three months of operation, with good feedback.

A helpdesk is a mediator, returning concrete, quick and targeted answers to technical ques-tions. As Saura explains, "the helpdesk activates networks, promotes experts, build a more cohesive practice while testing the extent—and the limits, of our current knowledge."

Figure – Employees reaching out to the Help Desk get an answer based on a network of knowledge sources such as databases, experts, and the talent network SkillFinder

RECOMMENDATION 5

The NHS should build a single-skill data science expert network modeled on the New York City's Mayor's Office Volunteer Language Bank or DataJustice.US

Model: New York City's Mayor's Office Volunteer Language Bank

Creating a comprehensive expert network of employee skills and abilities is a significant knowledge management undertaking for a large organisation. Tracking one skill at a time would enable the NHS to gain experience of building such a network and simultaneously create an incentive incentive for people to participate. One example of tracking a single skill is the New York City Mayor's Office Volunteer Language Bank, an expert network that matches the supply of multilingual civil servants to the demand for translation services. Recruitment is done through bi-annual appeals to City workers. Volunteers are self-selected and do not have to be certified, although their work is rated afterward.[37]

---

37   *We would like to thank Azadeh Khalili, Executive Director of Citywide Language Access Initiatives, and Kenneth Lo, Language Access and ESOL Initiatives, both of the Mayor's Office of Immigrant Affairs, for their assistance.*

Model: The GovLab/Justice Management Institute's Data Justice Platform

Data Justice (datajustice.us) is a website to foster peer-to-peer learning among criminal justice practitioners and policymakers and helps officials get fast and comprehensive answers to their questions about how to make better use of data to reduce incarceration and crime.

On the website, criminal justice practitioners can search for colleagues with relevant experience, ask and answer questions, and track their own knowledge of innovative ways of using data at every stage of the criminal justice process. The DataJustice.US platform launched in January 2017.

Why are these interesting for the NHS?

By focusing on a single skill widely distributed among the workforce, the Volunteer Language Bank and DataJustice do not not necessitate formal changes to workplace organisation and connect officials across jurisdictions to elicit translation and ata science skills, respectively.

As such, these may be an instructive example for how the NHS might develop and deploy a directory, for example, of data science skills. Much as one would like to have more "end to end" data scientists, comments Dr. John Varlow, Director of Information Analysis at NHS Digital, the reality is that one person is good at formulating queries and another knows a lot about terminology and a third knows more about programming or relational databases. Technology could help to pinpoint those skills.

Volunteer Language Bank

Language services—translation and interpretation—are a matter of urgency in New York City, where some 25 per cent of residents (1.8 million people) have limited English proficiency. Public health, public safety, and child welfare demand communication with all residents in a language they can understand. In addition, language access to City services is a right under civil rights law, and the City has been sued because people with poor English comprehension were not receiving needed services. In 2008, Mayor Michael Bloomberg signed Executive Order 120, which required agencies that offer direct public services to provide meaningful language access. Through the Volunteer Language Bank (VLB), multilingual volunteers from among City workers offer a valuable supplement to professional interpreters and translators. While the Volunteer Language Bank has existed in some form since 1989, the launch of the Citywide VLB database on the City's intranet in July 2010 provided a new tool to harness the language skills of City employees.

Figure – NYC Volunteer Language Bank homepage

New York City public schools and the 311-service that accepts calls from residents who are reporting problems in the city have long provided interpretation services. The 311 operators have tools to identify the language of the caller. Within three minutes, they can get an interpreter on the phone. Schools and 311 operators still rely exclusively on professional interpreters. Numerous other agencies, however, use the VLB to remedy the shortfall in available translators, especially for small jobs.

The City posts requests to recruit volunteers twice a year. Currently, more than 1200 volunteers, who speak approximately 70 languages, have joined and worked a total of more than 3500 hours helping to interpret and to translate or review documents and to provide rapid responses in specific situations when there is a need.

After obtaining permission from their supervisors, workers visit the City's intranet to sign up as a volunteer. In turn, agency staff members can post and manage requests for translation or interpretation assistance. If there is a document to translate, the agency posts it on the intranet with a request for specific languages. If a person contacts the agency with a specific request, the agency summarises the situation and language needed on the intranet.

The database relays requests to volunteers with the appropriate language skills. Volunteers can then respond and sign up to do a translation. Agencies informally rate volunteers after their interventions. Medical and legal interpretation are typically not provided through volunteers, because these involve technical or complex language and require additional training.

For several years, the City had a certification program to test the language proficiency of volunteers, but this process was expensive, on the order of $400 to $500 to certify one person. Thus, it was defunded three years ago.

Figure – Creating a Request for Translation

What motivates City workers to volunteer for this on top of their regular jobs? Usually, it's a desire to help fellow immigrants. The City issues customer service awards each year for volunteers, noting those who provided a particularly large number of hours translating or got good ratings from the agency.

The success of the Volunteer Language Bank shows that government workers can be motivated to contribute extra effort, particularly when they personally identify with the people they are helping. The New York City experience is also an example of the hidden talents that exist throughout government, and suggests how it can be brought to the surface and harnessed. It shows that fast, highly-responsive services can be offered using modest network technology and a volunteer database. Nonetheless because the City bears a legal responsibility to ensure that key services are delivered in a language citizens can understand, the City is scaling back its use of the Volunteer Language Bank in order to emphasise the use of professional services.

RECOMMENDATION 6

The NHS should use two-stage prize-backed challenges like the Harvard Catalyst Diabetes Challenge to augment the supply of internal data science capacity with external know-how

Model: Harvard Catalyst Diabetes Challenge

In 2010, Harvard Catalyst undertook an experiment to open up how universities generate research questions. Catalyst sponsored a $30,000 prize-backed challenge to come up with research topics that might be promising. Harvard uses challenges as part of a wider program to improve research and innovation and not as a one-off contest.

Some questions in health and social care may cross-disciplinary boundaries as well. For example, although epidemiologists might have gathered the data about incidence of in-hospital infection rates, this does not give physicians a monopoly on good ideas for how to reduce those infection rates. An architect, a public health policy professional, a psychologist, or an HR manager working in a large company all might have experience with creating incentives for washing hands.

Why is this interesting for the NHS?

One strategy for managers within the NHS to connect with such diverse expertise quickly is through the use of a prize-backed challenge, in particular a two-stage challenge like

the Harvard Catalyst Diabetes Challenge. Not only did the incentive of a prize help to elicit outside expertise quickly, but the two-stage design was well suited both to attract diverse ideas and thinking from a wide array of sources and to result in implementable proposals. The Health Quality Information Partnership, which was highlighted in Story 1 of this report for its difficulty in turning data into actionable interventions, might use challenges to get public input on how to do so.

Harvard Catalyst

Harvard Catalyst is a "pan university clinical translational science center" situated at Harvard Medical School that explores how to rethink all aspects of the scientific research process to help "high-risk, high impact problems in human health-related research." Catalyst takes each aspect of the research process and designs a strategy for opening it up to more information and input from a wider audience.

The goal is to bring in fresh ideas and novel perspectives. Typically, an academic decides on the direction for his or her lab. In an effort to generate new ideas from unlikely sources for promising approaches to fighting Type 1 Diabetes prior to investing research funding, The Catalyst challenge was unique because it did not ask people to come up with answers, as is typically the case in crowdsourcing projects. Rather, contributors supplied the questions. This enabled people to suggest ideas whether or not they had the resources to solve the problem they proposed.

After six weeks, 150 solid research hypotheses were submitted, encompassing a broad range of approaches from different disciplines. The creators of the Challenge analysed the subject matter of the submissions and found that they were "quite different from what existed in the literature and from the existing body of ideas under investigation within the Type 1 Diabetes research community."[38] The Leona Helmsley Trust put up $1 million in grant funding at Harvard to encourage scientists to create experiments based on these newly-generated research questions.

In addition to normal advertising of the grant opportunity, Harvard Catalyst used Profiles to identify researchers whose record indicated that they might be particularly well suited to submit proposals. The Profiles system takes the PubMed-listed publications for all Harvard Medical School faculties and creates a database of expertise based on a classification of their published papers. The topics of the experimental design proposals were matched to classifications within Profiles to facilitate searching. "The intention was to move beyond the established diabetes research community and discover researchers who had done work related to specific themes present in the new research hypotheses but not necessarily in diabetes."

In the end, the matching algorithm yielded over 1,000 scientists who potentially had the knowledge needed to create research proposals for these new hypotheses. "As ex-

---

38   Eva Guinan, Kevin J. Boudreau and Karim R. Lakhani, *Experiments in Open Innovation at Harvard Medical School, MIT Sloan Management Review 54 (2013), 45, 46.*

pected, some were accomplished diabetes researchers, but many others had expertise that diverged markedly from diabetes investigation." Harvard Catalyst then emailed these faculty members and announced the new funding opportunity. The team introduced a control at this point by randomly selecting half of the targeted researchers to give additional names as potential collaborators, identified by Profiles as being complementary, and suggested that they work together. As in the public sector, this kind of targeted outreach is not common in biomedical research.

The outreach resulted in thirty-one Harvard faculty-led teams competing for funding, of which twenty-three had been identified using Profiles. Of these, fourteen had no significant prior involvement in Type 1 diabetes research. In the end, seven proposals were funded with five of the lead investigators having not done any prior work in Type 1 diabetes.

It is too soon to know the impact of this experiment on sufferers of the disease. Typical timelines for medical innovations translating from bench to bedside are fifteen to thirty years. But the experiment did yield some exciting results: "The core insight driving Harvard Catalyst's experiments was that all stages of the previously narrow and fully integrated innovation system—from hypothesis generation to idea selection to execution—can be disaggregated, separated and opened to outside input. By opening participation to non-traditional actors, Harvard Catalyst achieved its objectives of bringing truly novel perspectives, ideas and people into an established area of research."

RECOMMENDATION 7

The NHS should use more foundation funded fellowships modelled on the Presidential Innovation Fellowship to augment the supply of internal data science capacity with external know-how

Model: Presidential Innovation Fellowship

In addition to prize-backed challenges, short term fellowships are becoming a more common way to get "outside eyeballs" working on hard problems.

By taking advantage of existing legal authority to bring in people from universities and non-profits, often funded by foundations, governments are accelerating the practice of hiring academics on shorter stints.

Figure – Presidential Innovation Fellow logo

Why is this interesting for the NHS?

These vehicles complement traditional hiring authority as a mechanism to bring in more talent from outside, including from the private sector. Although such temporary assignments—known as *secondments*—have long existed, these Fellowships are unique because the time to hire is unusually quick and they are focused on solving a specific problem.

The federal government in Mexico as well as the United States have created a new pathway — The Presidential Innovation Fellowship (PIF) program — to bring in more technologically skilled people from the private and university sector.

Launched in 2012, PIF is a competitive program that connects innovators from the business, non-profit, and academic sectors with government departments. Together, they work to produce innovative, short term projects to improve government efficiency. The program has evolved from one that parachutes new people into the White House to one that pairs innovators with civil servants to help implement change. Cabinet departments and agencies first compete to have their projects chosen, and then PIFs are selected in response to an agency's defined need.

What is especially unique and significant is that in exchange for getting the help of a PIF, public managers have first to write a strong business case for the Fellow, linking the position to a problem statement. By articulating the problem to be solved, instead of merely writing out the position, the job becomes focused on producing key deliverables and achieving value quickly.

The Hack the Pentagon Program (2016) is another a pilot program to attract talent on a short term basis. The program invites hackers to try and compromise the cybersecurity of the U.S. Department of Defence. It is designed to attract the best talent from the private sector. It is similar to security improvement strategies implemented by private companies.

The NHS should create an Open Data Learning Hub like the Commerce Data Usability Project for data scientists

Model: The Commerce Data Usability Project

The United States Commerce Department's Data Usability Project, which is a set of tutorials aimed at people outside the Department, guides interested researchers in how to work with the data to create useful research and products. This project might be instructive of new ways to generate more value with relatively little investment.

Why is this interesting for the NHS?

England is swimming in a complex data lake that includes hundreds of healthcare datasets, including patient-level and other data from the NHS Digital, Department of Health, Public Health England, Healthcare Quality Improvement Partnership, Quality Care Commission and many more. In addition to statistical reports based on patient-level data, much of the population level and institution data is published as open and downloadable raw data on England's national data catalog. The challenge, however, for each body within the health system is having the resources to translate data, once collected, into actionable recommendations to improve performance and realise value. The shift to action depends on getting a wider swath of people to use this data. To accelerate the gathering of practical value, the Department of Commerce has gone beyond the mere opening of its data and created tools to teach people how to use its data.

The Commerce Data Usability Project

The U.S. Department of Commerce collects, processes and disseminates data on a range of issues, including data on the economy, the environment, and technology. The U.S. Department of Commerce has become known as 'America's Data Agency' thanks to its tens of thousands of datasets, including satellite imagery, material standards, demographic surveys, and weather data.

But having a host of open and accessible data is not enough to ensure that the data furthers the agency's priorities. To accelerate its own ability to make this data useful, the U.S. Department of Commerce launched the Commerce Data Service, a public startup to help the Department's 12 Bureaus create and develop projects rapidly to advance the Department's mission. The Service is a project of Commerce's new Chief Data Scientist Jeff Chen, whose growing staff of analysts will soon number 34.

Organised around a new shared services model, the Data Service tackles core infrastructure issues around the Commerce Department and helps enrich human capital.
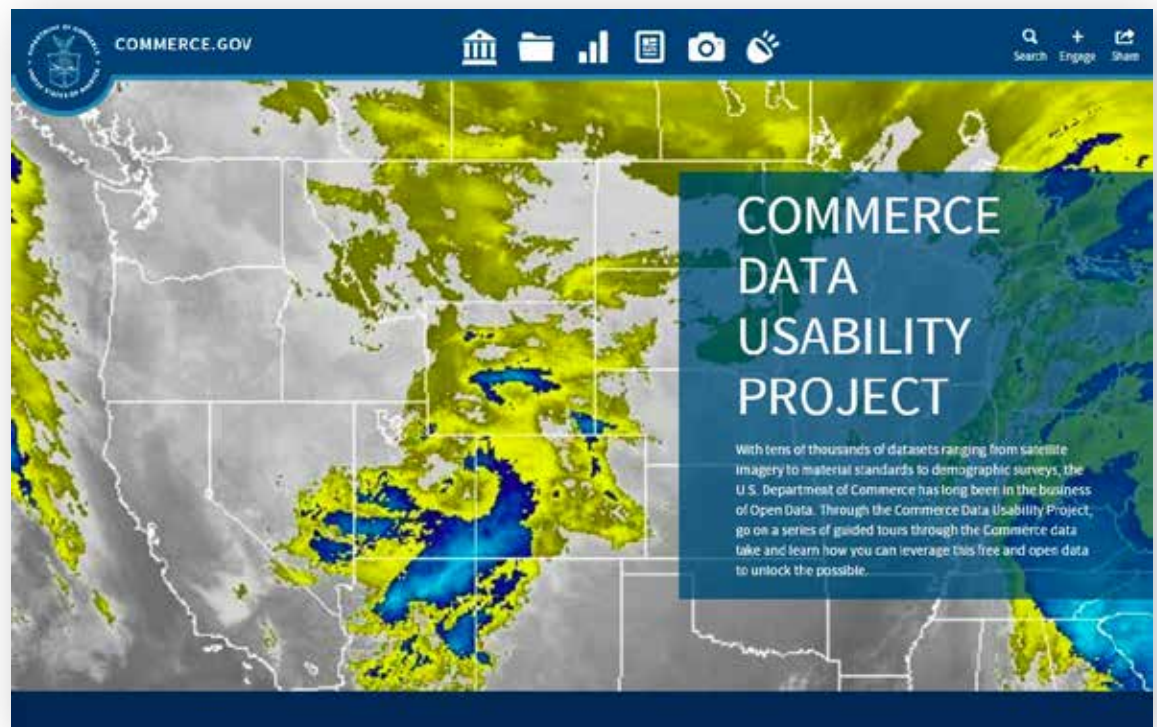

Figure – Commerce Data Usability Project Website

The Data Usability project includes a set of guided tours concerning how to turn the Department's raw data into valuable knowledge. The four initial video tutorials available leverage weather, cybersecurity, satellite and demographic datasets. They show how to:

‣ Analyse the risk of hail damage using NOAA's Severe Weather Data Inventory.

‣ Identify patterns in cyber security vulnerabilities using NIST's National Vulnerability Database.

▸ Approximate population density through nighttime satellite data with NOAA's Visible Infrared Imaging Radiometer Suite.

▸ Assist organisations to target communities to serve using Census' American Community Survey

"Whether you're a student, a developer or an entrepreneur interested in Commerce data, our tutorials can get you off to a running start by walking you through each phase of the data discovery process, from initial data touch through preprocessing, analysis and visualization," says Commerce Department Chief Data Officer Jeffrey Chen.

Each toolkit includes the original data and step-by-step instructions for using the data to conduct sample analyses and create visualisation from the data. The Department hopes that the kits will help to build more capacity around the use of open government data among those outside of government and to get more people interested in using the data to do research, test hypotheses and create valuable apps. Private sector companies, including MapBox, Earth Genome, Microsoft, and Zillow, have committed to contribute tutorials as well, showing how those companies use Commerce Department data.

The NHS should establish a process to explore the creation of Data Collaboratives

Model: DataCollaboratives.Org

A recurring theme in our research is the premise that data analytics is only as good as the accessible data.

Those we interviewed indicated that a lot of data could be valuable for gaining new insights and thereby achieving operational, outcome or strategic improvements, but stressed that such data is widely dispersed and collected by various public and private actors not used to sharing or legally required to share the data. For instance, within the current fragmented health data ecology, there is no way to document and follow patients across and within health settings including walk-in, acute, etc., limiting both performance and predictive analysis of what has worked and why.

Why is this interesting for NHS?

To leverage the potential of data to improve health and health care, NHS England will want to create and accelerate the use of "data collaboratives." The GovLab uses *the term data collaborative* to refer to a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors—including private companies, research institutions, and government agencies—exchange data to help solve public problems. For instance, within the US in 2015, the Obama Administration unveiled the *Precision Medicine Initiative*, a new model of patient-powered research and engagement that aims to build a knowledge base for individualised medicine. Specifically, the Initiative builds partnerships that enable it to include data from a number of

existing research cohorts, patient groups, and private-sector labs to expand our understanding of cancer genomics and eventually allow for treatments tailored to individual patients' genetic profiles.

To make data collaboratives more common across the NHS England, we recommend a process with key NHS data partners to explore different options by starting to answer the following questions:

▸ What different types of data sharing already exist within and outside NHS England, and what are the pros and cons of each type? What kind of data collaborative is best suited for each kind of data use or demand?

▸ What sorts of incentives and value propositions lead businesses and non-profits to decide to share their data assets with health data analysts? What factors must data owners weigh in making the decision to share?

▸ What are the risks of sharing data compared to the benefits, and how should data providers mitigate these risks? In particular, what are the risks to privacy including that of unauthorised disclosures, and what specific steps, technical and otherwise.can be taken to minimise those risks? What data governance frameworks are needed to leverage the potential and mitigate the risks of data collaboratives?

▸ What sorts of data sharing solutions—APIs, data pools, or research partnerships, for example—are best suited to maximising the public value of private data? Some companies donate data as a form of philanthropy, while others *have shared corporate data* through prizes and challenges, trusted intermediaries, intelligence products and other mechanisms. What are the respective roles—the relative benefits and disadvantages—of each of these mechanisms for data sharing?
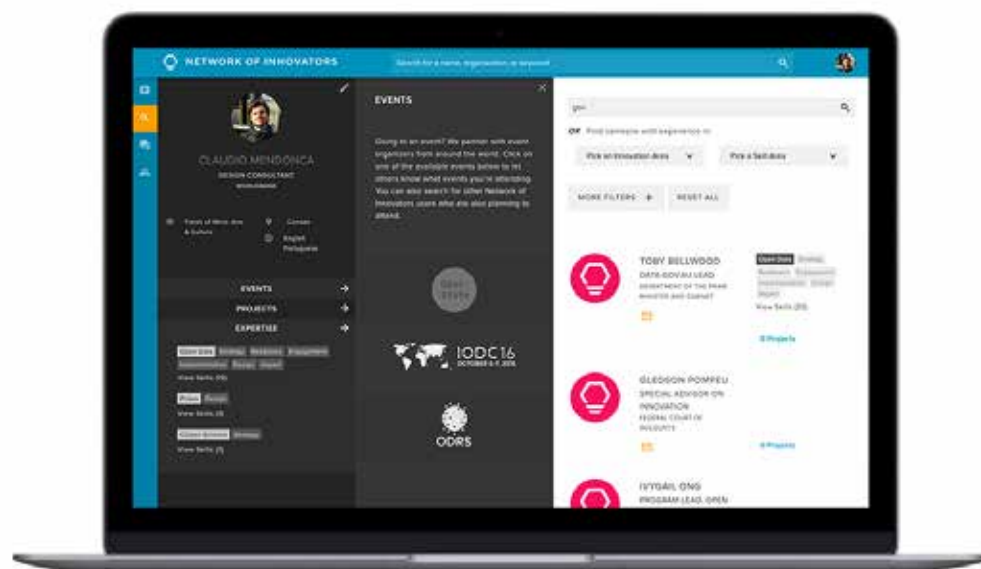
Figure—Network of Innovators

The NHS should connect UK university know how systematically to the NHS using a public-facing expert network

Model: Network of Innovators

An online 'Brain Trust,' filled with smart individuals working across government, universities, and the private sector, could provide data, facts, opinions, advice, and insights about pressing healthcare questions of the day.

Why is this interesting for the NHS?

Imagine if, on data analytic issues, the NHS, together with the Department of Health, NHS Digital, and other components, had access to a network of leading data scientists, health policy and clinic experts acting as a kind of knowledge network or talent bank? Instead of having to rely on a small group of stakeholders for advice, the NHS could use this Brain Trust to expand its expertise outside its institutional borders. A pilot deployment would make it possible to determine empirically whether technology can bring expertise in from the outside efficiently.

Network of Innovators

The NHS needs both a secure physical infrastructure to manage its data and the "conversational" infrastructure to enable the organisation to tap the best know-how on an ongoing basis to undertake the difficult work of opening up its data and delivering on the promise of more diverse and effective approaches to delivering care.

Tapping intelligence and expertise locally and globally needs to become a hallmark of the NHS and is a key ingredient to enabling the world to benefit from the first nation to open up the system's healthcare data. New tools—what GovLab calls technologies of expertise—are making it possible to make more searchable what people know and to match the supply of citizen expertise to the demand for it in government.

The NHS will want to take advantage of the technology of expertise to create a "public data science brain trust"—an online expert network to enable NHS decisionmakers to talk "to the geeks" on an ongoing basis. An expert network combines a directory of people and what they know with communications tools that facilitate questions and discussion.

The GovLab worked with the NHS Data Policy team over the last year to design and proto-type such a Network, described in detail in Appendix 3. Such a tool could make it possible to know who knows available data and its uses across the public sector, and to communi-cate with them as well as with self-selected participants from universities and industry.

This open source tool, designed by government officials for government officials, should be piloted among the existing community of analysts across the Cabinet Office, includ-ing DH as well as the rest of NHS England and NHS Digital and HQIP (under 1000 peo-ple) to test the value proposition of using technology to get answers to questions more quickly. The Network will capture participants' data analytic (and other) know-how as well as types of expertise (credentialed, skilled, experiential) in a searchable directory.

In parallel, 250 academics should be asked to participate in the pilot. Leadership at Im-perial College and University of London in the UK and New York University's Tandon School of Engineering have already agreed to be a part of such a test.

A pilot deployment would make it possible to determine empirically whether such an expert network with Q&A and discussion capability could help:

▸ Managers and decisionmakers to get answers regarding how to use data analytic techniques to solve problems

▸ Analysts to learn, from one another and from those outside the public sector, which techniques to use and how

▸ Create the incentive for academics and others to answer questions and contribute their know how for public good

Such a test should be run as a controlled trial over a three-month period with impacts measured against working in closed teams. Testing with real users would allow for the identification of desired features and practices before any major investment is made either in enterprise-wide technology or in changing how people work.

APPENDIX 1
INTERVIEWEES

Dr Martin Bardsley, Senior Fellow at Nuffield Trust and Health Foundation

Ravi Baghirathan, Deputy Director, *Strategy Group*

Emily Bird, Policy Lead, *NHS England, New Towns*

Thomas Brain, Enterprise Lab, *Imperial College*

Trevor Butterworth, Sense about Science

Martin Caunt, Senior Analytical Lead, *NHS England, Head of National Operational Research in the Operational Research and Evaluation unit*

Dr Alok Das, Senior Scientist for Innovation Design and Rapid Response Team Lead, *AFRL*

Dr Mark Davies, Mede Analytics

Dr Marc Davies, Director of Informatics and Digital Strategy, *Department of Health*

Emma Doyle, Head of Data Policy

Jane Dwelly, Director, *NHS Expo*

Elizabeth (Liz) Dyer, Strategy Analyst, *NHS England New Towns*

Julian Flowers, Consultant in Public Health and Director of the Eastern Knowledge and Intelligence Team in Public Health England

Susannah Fox, Chief Technology Officer, *Department of Health and Human Services*

Professor Paul French, Vice Dean (Research), *Faculty of Natural Sciences, Imperial College London*

Dr Noha Gaber, Director, *Office of Internal Communications, Office of Public Affairs, Administrator's Office, US Environmental Protection Agency (gaber.noha@epa.gov)*

Tracey Gyateng, Data Lab Project Manager, *New Philanthropy Capital*

Jeff Hammerbacher, Cloudera

Timo Hannay, Digital Science (former)

Cathy Hassell, Deputy Director, *Quality Programmes, NHS England, manager of the NHS Technology Enabled Care Services (TECS) programme*

Loraine Hawkins, Chief Analyst, *NHS England*

Jessica N. Hernandez, MA, *MS, Team Leader, Innovation and Technology Solutions (ITS), Division of Administrative and Laboratory Services, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration*

Sir Jeremy Heywood, Cabinet Secretary

Professor Danny Keenan, Medical Director, *Healthcare Quality Improvement Partnership*

Tim Kelsey, National Director of Patients and Information (former), *NHS England*

Sir Bruce Edward Keogh, KBE, *FRCS, FRCP, Chief Medical Officer, NHS England*

Azadeh Khalili (Azi), Executive Director, *Citywide Language Access Initiatives, Mayor's Office of Immigrant Affairs*

Dr Geraint Lewis, Chief Data Officer, *NHS England*

Kenneth Lo, Language Access and ESOL Initiatives, *Mayor's Office of Immigrant Affairs*

Paul Maltby, Director of Data, *Cabinet Office*

Amen Ra Mashariki, Head of Data Analytics, *New York City*

Rupert McNeil, Chief People Officer, *Cabinet Office*

Professor John Newton, Chief Knowledge Officer Public Health England and Deputy Chair of the National Information BoardHolly Norman, *Strategy Advisor, NHS England, New Towns*

Ronan O'Connor, Regional Director, *Patients and Information, NHS England*

Tony O'Connor CBE, Senior Analytical Strategist, *Department of Health*

Daniel Ray, Director of Data Science, *NHS Digital*

Nichole Rosamilia, Technical Information Specialist, *Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration*

Anthony Rudd CBE, National Clinical Director for Stroke NHS England

Pascal Saura, Senior Knowledge Management Officer at The World Bank

Professor Eric Schadt, Carl Icahn Institute, *Mt. Sinai Hospital*

Sir Nigel Shadbolt, Open Data Institute, *Oxford University*

Adam Steventon, Director of Data Analytics, *Imperial Health Foundation*

Geraldine Strathdee, National Clinical Lead, *Mental Health Intelligence Network, Consultant Psychiatrist, Visiting Professor, Integrated Mental Health Program UCL Partners, Former National Clinical Director for Mental Health, NHS England*

Charles Tallack, Head of NHS Operational Research and Evaluation

Roger Taylor, author of God Bless the NHS

Prof. John Taysom, University College London

Prof. Philip Treleaven, Department of Computer Science, *University College London*

John Varlow, Director of Information Analysis, *Health and Social Care Information Centre*

Hester Wadge, Policy Fellow, *Director of the Sowerby eHealth Forum*

Professor Sophia Yaliraki, Imperial College

Dr. Zeynep Engin, Center for Science and Policy, *Cambridge University*

SMARTER HEALTH

BOOSTING ANALYTICAL CAPACITY AT NHS

GOVLAB

thegovlab.org