GOVLAB

THE OPEN DATA ERA
IN HEALTH AND SOCIAL CARE

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

Stefaan Verhulst
Beth Simone Noveck
Robyn Caplan
Kristy Brown
Claudia Paz

NHS England

EXECUTIVE SUMMARY

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

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

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

▸ **Accountability:** The use of data to hold healthcare organizations and providers accountable for treatment outcomes.

▸ **Choice:** Providing open data to help patients make informed choices from among the healthcare options available to them.

▸ **Efficiency:** Improving the efficiency and cost-effectiveness of healthcare delivery.

▸ **Outcomes:** Improving treatment outcomes by using open data to make the results of different treatments, healthcare organizations, and providers' work more transparent.

▸ **Patient satisfaction and customer service:** Using open data to educate patients and their families and make healthcare institutions more responsive.

▸ **Economic growth and innovation:** Using open data to fuel new healthcare companies and initiatives, and to spur innovation in the broader economy.

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

Overall, this draft blueprint from the GovLab describes the need for careful assessment of healthcare open data initiatives and outlines an approach that can help in such an assessment. The approach helps both in assessing the opportunities of open data, and also in considering potential risks and challenges. Though there has been significant discussion about the concerns and risks emerging from open data, there has been little research into its potential economic and social benefits. In today's

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

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

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

▸ **Build an Open Data Learning Capacity and Culture** within the NHS. This can include a number of steps to institutionalize action and learning about open data, such as creating a corps of volunteers with technical knowledge, setting up an Open Health Data Academy, soliciting calls for questions from the academic community, and posing challenges and offering prizes for the best uses of open health data.

▸ **Engage the public in defining metrics** of success with regard to open data. In addition to ensuring that direct stakeholders (e.g., patients and providers) are included in decisions about open data, the NHS may also want to consider allowing citizens to participate – for example, though a citizens' open health data panel.

▸ **Develop a common assessment framework** with partners in other sectors and countries to measure the impact of open health data, using meetings, online communication, and additional tools.

▸ **Stay flexible and adaptive in measuring impact**. As the use of open data evolves and becomes more dispersed, the NHS will need to find a balance between a centralized measurement function (which sets measurement standards, aggregates data across programmes, and monitors data quality) and a more diffused structure that empowers users at different levels of the organization.

▸ **Share what is learned (including failures) with everyone**. By sharing what is learned (including mistakes and failures) on a regular basis, the NHS can enable distributed oversight and peer review mechanisms to ensure the quality of its research and enhance the way open data is used.

▸ **Build a research network and expert network** to build capacity quickly and broker debate around methodology and findings. Outside expertise, gathered through meetings, online input, or an advisory board, will be critical in determining the best ways to assess impact.

▸ **Develop an open health data ecology map,** possibly using crowdsourcing, with a dictionary of all open health data sets used along with the variety of uses and users.

▸ **Publish, integrate and fine-tune the open data conceptual framework** for feedback and improvement, using expert advisors and input from stakeholders.

▸ **Engage citizens and practitioners in shaping the open health data programme.** The NHS should engage experts in the field on a regular basis to determine which datasets have the highest priority for them; what new datasets should be released as open data; and which open data collections are particularly easy or difficult to use.

We are releasing this as a draft in order to encourage discussion and additional insights from interested readers. We welcome comments and suggestions, and look forward to incorporating them into a subsequent version of the report.