High quality datasets:
Our vision

NHS England INFORMATION READER BOX

Directorate		
Medical	Operations	**Patients and Information**
Nursing	Policy	Commissioning Development
Finance	Human Resources	

Publications Gateway Reference:	00745
Document Purpose	Guidance
Document Name	High quality datasets: Our vision
Author	Chief Data Officer
Publication Date	January 2014
Target Audience	CCG Clinical Leaders, CCG Chief Officers, NHS England Regional Directors
Additional Circulation List	CCG Clinical Leaders, CCG Chief Officers, Medical Directors, NHS England Regional Directors, Communications Leads
Description	This is NHS England's vision for high quality data. Aimed at commissioners it sets out NHS England's ambition to bring the NHS data services up to modern standards and calls on commissioners to think about their strategies for collecting and using data. Plans and the next steps for commissioners are provided.
Cross Reference	N/A
Superseded Docs (if applicable)	N/A
Action Required	N/A
Timing / Deadlines (if applicable)	**N/A**
Contact Details for further information	Chief Data Officer P & I 7th Floor Quarry House England.CDO@nhs.net

Document Status

This is a controlled document. Whilst this document may be printed, the electronic version posted on the intranet is the controlled copy. Any printed copies of this document are not controlled. As a controlled document, this document should not be saved onto local or network drives but should always be accessed from the intranet

VISION FOR HIGH QUALITY DATASETS

Foreword

The NHS belongs to the people, and NHS England is entrusted with ensuring that there is high quality care for all, now and for future generations. At a local level, clinical commissioning groups (CCGs) are responsible for securing high quality care for their population alongside NHS England's area teams. To do so, clinical commissioners must be fully informed about (1) local health needs and (2) the quality, efficiency, and equity of the care that their population is receiving.

When we published our planning guidance *Everyone counts: Planning for Patients 2013/14* in December 2012, we promised to produce advice for CCGs on the characteristics of the datasets that they will need in order to be fully informed about these two elements of health intelligence. In return, we asked each CCG to identify its own strategy in the light of our advice. This document sets out our vision for good data and is primarily aimed at supporting CCGs as they develop their strategies for making better use of data. We hope that it will also be helpful more generally to everyone with a stake in improving the use of data and information within the health service.

At the heart of our vision is a comprehensive, pseudonymous dataset that will link each person's interactions with the health and care service. Known as the *Care Episodes Service* (CES), the dataset is being developed as part of the "care.data" programme, which will create a modern data service for the NHS. Because it will span all care settings, CES will enable commissioners to consider their population's health needs and experiences in the round. CCGs and the organisations that support them will be able to use the collated data to generate cutting-edge visualisations and to feed a range of analytical, predictive, and evaluative tools. The data-driven insights generated by these images and applications should then guide commissioners, clinicians and patients to make better, more considered decisions and choices. We encourage CCGs to work with the wider health economy to maximise the use of data for a better patient experience and improved outcomes at all stages of care.

The use of data and informatics within the health and care system is at a crucial point of change. NHS England has embarked on an unprecedented programme to promote transparency in the health service. By exposing high-quality information about unwarranted variations in the care being delivered by the NHS, we believe we can empower citizens and patients to play a far more active role in the design of local health services. Together, these twin levers of transparency and participation will result in higher quality, more cost-effective, and more compassionate health care for us all.

Geraint H. Lewis
Chief Data Officer

Tim Kelsey
National Director for Patients and Information

Vision

Compared with only a few years ago, the public now expects to be offered a choice of high quality, high value services in every sector of the economy. If younger generations are to treasure the NHS as fondly as the post-war generations, then healthcare services will need to become as responsive to their needs and as consistent in their quality as the best supermarkets and airlines. Customer service feedback, insight, and data analytics have been key drivers of improvement in these industries. However, in healthcare, these quality improvement levers have been relatively underutilised. Now, as we face the combined pressures of an ageing population, the rising prevalence of chronic disease, and budgetary constraints, it has become vital for us to deploy every quality improvement device available whilst ensuring there are robust safeguards in place to protect people's privacy and confidentiality.

There is a growing evidence base that greater transparency encourages clinicians and commissioners to learn from their peers, and supports patients to make more meaningful, informed choices. Therefore, at the core of NHS England's strategy for securing high quality care for all, now and for future generations, we want to see a cultural shift towards greater transparency and more participation. We are embarking on a sustained and systematic drive first to publish higher quality and more comprehensive data in more accessible formats, and second to help citizens, patients, and clinicians use this information to expose unwarranted variations, to call for change where it is needed, and to improve the design and performance of local services.

The bottom line is that the NHS must eliminate unwarranted variations in the quality, equity, and cost-effectiveness of the care that we commission and provide. We can no longer afford unnecessary duplications in care and other wasteful inefficiencies. Instead, we must record, collate, analyse, and publish better information, so that the most effective and efficient treatments and services are highlighted and allowed to propagate through a process of innovation, rapid feedback, adaptation, and dissemination. The financial and demographic realities facing the NHS are now forcing all stakeholders to take data and information far more seriously. If we are to close the quality gap and the financial gap now facing the health service, then as commissioners we must become far more sophisticated users of data on behalf of the populations we serve.

How data will transform our health service
Underpinning NHS England's strategy of using transparency and participation to secure high quality care for all is the use of good, pertinent data. We recognise six key benefits from using good data as described in Figure 1.

Figure 1: Benefits of good data

Such data can broadly be considered in three

classes: *Routine administrative data*, *user data*, and *my data*. Routine administrative data refers to the analysis of large, linked datasets drawn from multiple sources. They can be used to identify trends and patterns, draw comparisons, predict future events and outcomes, and evaluate services. User data, also known as citizen-generated data, are created when the public or the users of a service are engaged and generate feedback. Finally, my data involves returning a person's own information back to them so that they can put it to work in their own interests.

Routine administrative data

In today's world, successful organisations rely on managers making data-driven decisions. Leading organisation in the public, private, and third sectors are using data in increasingly innovative and imaginative ways. They use data to map out and understand their current processes and operating environment, to predict future challenges and opportunities, and to deliver and assure consistently high quality outcomes. Indeed, the integration and analysis of multiple information sources has become a core activity for a whole gamut of professions and organisations, ranging from manufacturers and retailers to investors, weather forecasters, and local government departments.

Of all of the sectors in the economy, healthcare is one of the leading generators of administrative data. During each episode of a patient's care, huge amounts of information may be generated, ranging from demographics to symptoms, to observations and investigations, through to diagnoses, treatments, procedures and outcomes. Increasingly, this information is being recorded and stored electronically rather than on paper; however only a tiny fraction of this information is currently fed into central flows of data where the information is collated and used for wider patient benefits. Administrative data should be seen as a core strategic asset and should be at the heart of all

decision-making. To keep pace with other sectors, NHS commissioners must become avid, sophisticated users of data, drawing on the full spectrum of qualitative, quantitative, and predictive approaches.

One of the unique strengths of our being a national health service is the completeness of our routine administrative data and the existence of a single, universal identifier: the NHS number. Already, the hospital episodes statistics (HES) dataset is the envy of the world because it contains information about every hospital admission that has occurred since 1989, covering the entire population of England and spanning every hospital provider nationwide. Researchers have used HES to publish thousands of peer-reviewed research studies – despite the fact that the information contained within HES is currently limited to demographic, diagnostic and procedural information. Over the coming years, we will transform HES into the Care Episodes Service (CES), which will not only include a far richer hospital dataset but will also be expanded to include all other settings of care.

Any shopper will be familiar with how loyalty cards and online platforms offer them tailored deals and recommendations. To create this customisation, retailers need a detailed understanding of the

> "In healthcare, for example, big data levers can boost efficiency by reducing system-wide costs linked to under-treatment and overtreatment and by reducing errors and duplication in treatment. These levers will also improve the quality of care and patient outcomes." McKinsey Global Institute, *The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation* - Jan 2013.

customer's shopping patterns and those of thousands of similar customers. The underlying data systems allow retailers not only to manage their stocks and supply chain with unwavering efficiency, but also to engage each customer as an individual. In this way, routine administrative data allows retailers

to meet the needs and preferences of their customers far more closely than the generic arrangements that came before. Ultimately, these organisations are using routine administrative data to achieve better outcomes by better satisfying each customer as an individual.

The potential role of routine administrative data in healthcare is greater than in almost any other sector. Intelligence-based maps and models can expose inefficiencies and under-performance. For example, tens of millions of pounds could be saved simply by exposing variations in the use of generic versus branded medicines (see Figure 2). As richer data become more available and transparent, so this type of analysis can be repeated with other classes of medication and medical devices, as well as with clinical procedures and other health and care services. Indeed, a better understanding of the flow of patients through different services based on more granular data will help commissioners understand how different types of patients experience different health outcomes and access different types of care,

They can then optimise the nature, location, and specification of the preventive and reactive services they commission for their population.

More importantly, as well as identifying areas where outcomes are abnormally good or bad, by conducting *segmentation analyses* on routine administrative data, it can sometimes be possible to identify the root causes of abnormality, based on complex overlaps of treatments, conditions, and behaviours. By studying real-time feedback from patients alongside health outcomes data, analysts can help ensure that patients are protected from unacceptable standards of care and that examples of best practice are replicated and emulated across the NHS. Segmentation can be performed according to any cluster, such as geography, demographics, clinical details, or pathways of care, and can be used to detect trends and outliers in disease distribution, severity, and mortality as well as the quality, equity, and efficiency of the care provided.

Another strength of analysing routine administrative data is the ability to integrate unconventional data

Figure 2: Percentage of proprietary statin prescribing by CCG Sep 2011- May 2012.
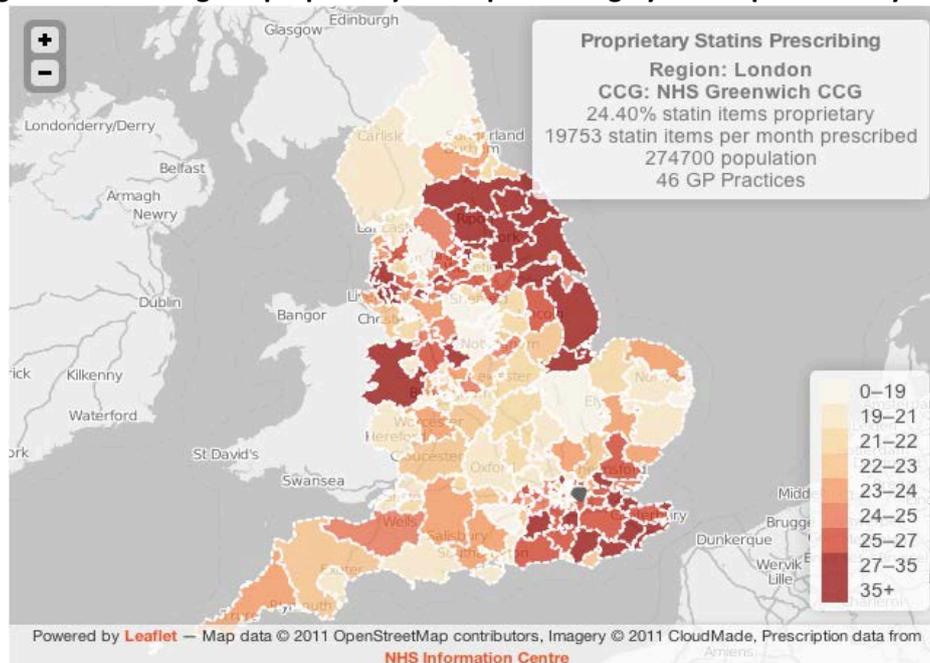
Source: Prescribing Analytics

sources. By linking to social media feeds, analysts can derive a whole new set of insights relating to the health needs of different subpopulations and their use and experiences of health and care services. For example, they can use linked data to measure rates and patterns of vaccinations, and then design and target resources aimed at increasing the uptake of vaccinations in the most at-risk and difficult-to-reach groups. Innovative use of tools such as Google Analytics, Care Connect, and NHS Choices, combined with Twitter and Facebook feeds, overlaid with pseudonymous clinical and health services data can help detect potential outbreaks as they emerge, and to target information campaigns and resources where they will have the most impact.

This approach is not so dissimilar to innovative policing practices now occurring in some American cities such as Chicago. Here, dispatchers send patrols to predicted crime hotspots, based on patterns and conditions that were previously associated with higher incident rates. Police analysts then use algorithms to predict subsequent clusters of crime, in a manner analogous to earthquake aftershocks. For example, following reports of a spate of car break-ins in one neighbourhood, those neighbouring areas that are now at higher risk of further burglary can be predicted and targeted with additional police presence.

In healthcare, routine administrative data are similarly being used to predict and prevent adverse outcomes. Some leading U.S. accountable care organisations, for example, run banks of predictive models and impactibility models on their datasets in order to determine which patients are at high risk of adverse outcomes, and which individuals are most likely to benefit from preventive interventions, respectively. By integrating multiple sources of information, these organisations are learning more about the safety, efficiency, and equity of the care received by their population. And they can use a stratified approach to preventing *triple fail* events,

such as the inappropriate initiation of haemodialysis, which are simultaneously (1) markers of potential low quality, (2) are costly to the health care system, and (3) represent a poor patient experience.

User data

Over the last 20 years or so, the Internet has utterly transformed our way of life. It is always on; it is infinitely personalised; and for many of us, it is in our pockets almost wherever we are. But what truly makes the Internet so powerful is its ability to transfer knowledge in all three directions, namely to, from, and between users. Excitingly, this democratising power of the Internet is still only in its infancy. Known as Web 2.0, this current phase of development is beginning to use the collective power of people to catalyse improvement in all walks of life and to stimulate growth across all sectors of the economy. Rather than simply flowing into our homes, businesses, and schools, information is increasingly ebbing back to organisations carrying the voices, opinions, and suggestions of the crowd.

This counter-flux of public-generated information is known as user data or citizen-generated data, and it may be either *passive* or *participatory* in nature. We all generate user data passively through our day-to-day interactions with the Internet. Take for example, the terms we look up using search engines such as Google and Bing; the web pages we visit; and the purchases we make. Or we can enable technology to generate passive user data on our behalf, such as when citizens in Boston install an app that uses the accelerometer in their smartphone to detect potholes as they drive over them, and automatically report the location to city authorities. But we can also generate user data more actively through likes, comments, reviews, complaints, tweets, posts and surveys. This type of participatory user data can reveal what services are working well or where improvements are needed. Both types of user data are powerful indicators of public opinion and can give voice to the opinions of marginalised citizens

who might previously not have been listened to.

Importantly, we can stimulate people into providing more user data by publishing more o*pen data* (i.e., by increasing the availability and usability of information about a service). For example, the City of Chicago's website has a "snow portal" that shows in real-time the location and activity of the city's snowploughs. It also allows citizens to report problems, share shovels and snow-blowers, and to adopt a stretch of pavement to clear. As *The Economist* put it, "Traditionalists may not be keen to brag on Facebook about how many pavements they have cleared, but plenty of others are—thus helping to spread the city's message that everyone has to do their bit."

User data allows any one of us to become a surveyor, auditor, or reviewer. Local government initiatives are beginning to use public complaints to hold contract providers to account. For example, many local government websites post complaints about public services that then map directly to the corrective actions taken by the contractor. Within healthcare, websites such as *Google Flutrends* use search terms related to 'flu symptoms to predict epidemic locations and trends across the globe. And within the NHS, the Friends and Family test is helping us listen to our patients and staff on an unprecedented scale. Over the coming months, it will begin providing a wealth of information about to the quality and experience of local services.

My data
The third category of data that we commend to commissioners is my data. There is growing evidence that giving people access to the personal data that organisations hold not only improves the quality of data as we review and correct any mistakes, but also that we make better decisions when we are engaged and informed. For example, the data that a phone company holds about my mobile use can help me choose a better tariff. In personal financing, giving me greater access to my past spending habits can

help me make better buying and saving choices. Likewise, providing me with a breakdown of the nutritional content of my shopping basket can help me make healthier choices. The Department of Business, Innovation and Skills (BIS) has already embarked on the *midata* project along with leaders in the energy, communications, and banking sectors to promote the use of personal data in these and other new ways.

Within healthcare, the *Blue Button* initiative is allowing patients in the United States to download their personal health information. This information can be used for a whole range of purposes, from showing it to their family and carers, to sharing with other healthcare providers, and analysing it against best practice guidelines. The data can be downloaded either directly from the electronic medical record system of their provider, or from a central database spanning multiple providers. While the original idea was to "create a large, prominent button that would represent data liquidity and access", the focus for the second generation of the Blue Button, known as Blue Button+, is to ensure that the data are both human-readable and machine-readable.

NHS England is currently working with the Department of Health, Public Health England, and others to develop a single, online portal for patients and the public. This portal will bring together information and online services currently fragmented across NHS Choices, NHS Direct online, NHS 111 online and HealthSpace. In addition, GP practices are now incentivised to enable patients to interact with their data, starting with the ability to book appointments and order repeat prescriptions online. As part of the care.data programme, patients will be able to download their entire CES history of encounters with all health and social care providers nationwide.

A call to action for commissioners

These are challenging times for commissioners. The Health and Social Care Act 2012 led not only to a whole scale reorganisation of the commissioning landscape, but to equally extensive changes in information governance, with a shift towards the use of pseudonymous datasets as the norm. As these changes begin to settle down, we would urge commissioners to see routine administrative data, user data, and my data as three of the greatest resources at their disposal.

We would like the web pages associated with this document to become an on-going resource for commissioners.

The web pages can be found at:
http://www.england.nhs.uk/resources/resources-for-ccgs/good-data-strat/

Please contribute any examples of best practice you would like us to consider for inclusion on the web page to england.cdo@nhs.net. We are particularly keen to share best practice about the collation and use of good datasets, and how they have led to demonstrable improvements in care. Equally, we want to hear more from CCGs about what guidance and support you need from us to develop your data infrastructure.

Next steps

Over the coming weeks, a working group composed of colleagues from NHS England and CCGs will co-produce an example data strategy for commissioners, which other CCGs will then be free to use and adapt for their local circumstances. This working group will report by 30 May 2014.

Once the working group has published its example data strategy, CCGs will be asked to work with their local stakeholders to adapt it to fit their local context.

CCGs should submit their data strategy to their Area Team by 1 September 2014, so that the agreed strategies can be published by 31 October 2014. These strategies will be 'living documents', which will be kept under review each year and amended and updated as necessary.