

Open Source Data

2.1 Public domain healthcare data

The UK government has generally been pursuing a policy of opening up access to data, indeed open data and data transparency was at the core of the 2012 Health and social care act. As a consequence of some high profile data security breaches in the last few years, the public via the press, have been voicing concerns around data privacy. None of these data breaches has been in the NHS but the NHS have become more risk averse. The result has been a slowing down in the release of new health data into the public domain.

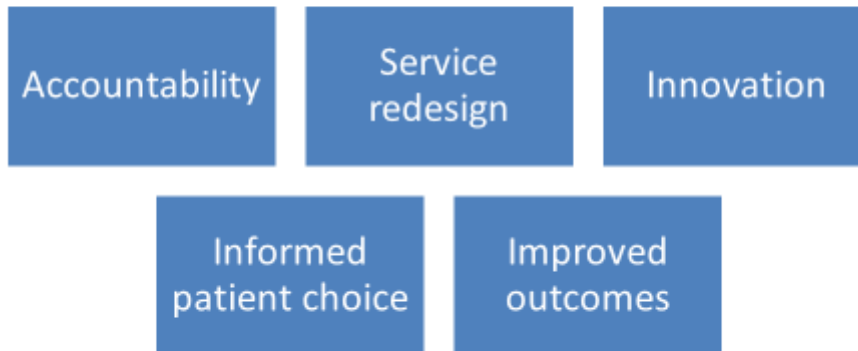
NHS Digital (previously known as the HSCIC) is the organisation that has responsibility for data within the NHS. Its 2016 business plan states that its objectives include

- Ensure that every citizen's data is protected
- Improving health and care in England by putting technology, data and information to work in the interests of patients, clinicians, commissioners, analysts and researchers in health and social care
- Managing the collection, storage, processing and publication of national health and care information, as directed by the Secretary of State and NHS England and
- Acting as the national source of indicators

Interestingly none of its objectives mention of making data *publicly* available although a lot of their data is. NHS Digital's focus now appears to be inwards, helping the health and social care system, and away from being a gateway enabling wider access to, and scrutiny of, healthcare data.

There is no question that the opening of healthcare data to scrutiny does provide the opportunity for patient benefit and service improvement. Patient benefit and service improvement has always been a core message of Pharma and they have been keen to engage with and help the NHS in these areas. But pharma must be mindful that when they use NHS data, or any open data, that they adhere to the terms of use that will be stated in the data licence.

Benefits of open health data



2.2 What is out there?

Open Health Data:

- Currently open data is published under the Open Government Licence, and is available through the NHS Digital and the Data.gov.uk site.

Restricted Health Data:

- Data that potentially carries personal or proprietary information, which is anonymized through the removal of personally identifying information. Access requires approval by the NHS and the terms of the agreement will specify and restrict the purposes for which it may be used.

Published Health Data:

- Data that has been published in a particular format by the NHS or other governmental entities. This data is generally prepared for presentation in an interactive, .doc, or PDF format. The original source data files are usually not included in the release.

The challenge for researchers looking at the UK market is that the NHS is fragmented into 4 autonomous regions: England, Scotland, Wales and Northern Ireland. Each region is responsible for its own data disclosure and each is at different stages of development.

2.3 Right to use?

Most of the public domain data is available for use under the open data licence. Restricted data will usually have specific licences terms that must be respected. Always check the terms

before using any data commercially, even summary data copied from a publicly available web site; and always reference it.

2.4 Key data sets

Primary care data

- **General practice prescribing:**

This is an excellent and detailed data source. It provides detailed prescribing data by practice, by month. The data includes prescriptions written by GPs and other non-medical prescribers (such as nurses and pharmacists) who are attached to GP practices.

 - The products prescribed are available at form/strength level and organised using the BNF (British National Formulary) classification system.
 - The metrics are: The total number of items prescribed and dispensed, the total net ingredient cost, the total actual cost & the total quantity are all available.
 - As of April 2016, monthly data is available for England, Wales, Northern Ireland and Scotland.
- **QOF (Quality and Outcomes Framework):**

QOF was designed as a tool to encourage practices to perform better. Measurement of particular indications showed them how they were performing and by achieving targets practices could gain extra income. The major wins have largely been achieved so QOF is now more of a monitoring tool than an incentive scheme. There are good reports on the NHS Digital web site as well as the underlying data. These data could be used to assist with targeting, segmentation, and service redesign.

 - QOF can tell you the overall size of a GP practice as well as the number of patients registered with any of 22 conditions (Asthma, Cancer, Diabetes, Dementia, Hypertension, Obesity, & Smoking to name a few).
 - For each of these 22 conditions performance metrics are included like 'The number of patients with diabetes in whom the last IFCC-HbA1c is 59 mmol/mol in the preceding 15 months.

Secondary care data

- **Hospital Episode Statistics (HES):** This is a (monthly) data set of every patient episode in an English hospital. It is derived from the data prepared by hospitals to gain reimbursement for the services they provide to patients. Key limitations are that the data set does not have information on the drugs prescribed during the episode. A&E and Outpatient data is less detailed than inpatient data.
- **Accessing HES:** Companies can apply to NHS Digital to gain access to patient level HES data but this is an increasingly difficult process. Whilst gaining a commercial benefit from the data is not specifically precluded, it must be secondary. A consequence is that few if any pharmaceutical companies have direct access to HES and this is unlikely to change. Intermediary companies have been able to retain access and they can provide insights to Pharma so long as these fall within their conditions of use.

- The conditions of use are set by NHS Digital in conjunction with an external panel "IGARD" the Independent Group Advising on the Release of Data. The overarching principle is that use of HES data must be for the primary benefit of the "health and social care system". Because the data is patient based (although anonymised) the conditions of use are strictly controlled and restrictive.

There are many other data sets, mostly annual, available on the NHS Digital website. Here are some useful URLs:

<https://digital.nhs.uk/article/191/Find-data-and-publications>

<http://content.digital.nhs.uk/searchcatalogue>

2.5 Key published information

Primary care

Information exists on line for every (English) Clinical Commissioning Group (CCG). This will detail their strategic and operational plans and their performance against their targets - Health Indicator frameworks. Similar information will exist in the regions.

Secondary care

Each hospital trust and acute hospital will have a website. This will publish information on their strategy and plans. However there is no template for these sites nor benchmarks for the quantity and quality of information provided.

FOI requests

Additional information can often be obtained using FOI (freedom of information) requests. These take time to come back and given the number of FOIs that organisations are dealing with at any one time the response will typically answer the letter rather than the spirit of the request.

In the pipeline?

There was an initiative to link by patient primary and secondary data sets together: Care.Data. Public concerns over the creation of large-scale healthcare datasets that could be used outside the NHS has effectively killed this initiative.

2.6 Getting started

When looking to find out information about a particular healthcare issue in the UK the first step should be to review what is available in the public domain; after all it's free! The NHS Digital web site is a good place to start.

The next places to look will be the web sites of the organisations of interest, for example the CCGs and Hospital trusts. A number of agencies will have expertise on the information available from these sites and may even have consolidated it in some form and providing added value services based on the data.

Once you have built up a basic understanding, the next step is to dig deeper using your own expertise and resources or engage with a partner (agency or individual) to help you. Whilst working with a partner or agency will incur a cost, the advantage is that they will have worked with the data and should know it well enough to advise you on its strengths and weaknesses. Some agencies will have established products, services and in-house experts that could save you a lot of time and effort.

Simon Brander

Founder Director

CSL (CompuFile Systems Ltd)

Phone: +44 (0)1483 528303

Mob: +44 (0)7710 037489

E-mail: simonb@CSL-UK.com

Website: www.nhs-analytics.com