Primary Health Care SG

Primary Health Care Specialist Group (PHCSG) of the British Computer Society

Clinical Computing Special Interest Group (CLICSIG) meeting on the NHS Federated Data Platform (FDP) and other health data repositories

Date: Saturday 13th May 2023

Venue: ISH Venues, Regent's Park, London

Attendees: This CLICSIG meeting was a hybrid event attended by 14 BCS/FCI members or invited guests with an interest in clinical/health informatics, data protection, health data repositories or information governance. Roles of attendees ranged from current/retired General Practitioners, ICS Chief Clinical Information Officers, employees of national organisations involved in large data collections or organisations campaigning for confidentiality and consent in health, members of advisory groups related to health data use/access and an information governance consultant.

Attendees were present to contribute their own personal views and experiences and not to represent their respective employing organisations.

Reason for the meeting

During the panel session of the 2022 PHCSG annual conference in Harrogate, a discussion was held regarding the difficulties in extracting health data for research purposes and the associated problems regarding data access and governance models. As part of that discussion, the content of this CLICSIG meeting was proposed in order to allow continued detailed debate.

The aims of this CLICSIG meeting were to raise concerns plus offer advice from the group regarding:

- the current intention(s) of the proposed NHS England Federated Data Platform for direct patient care and secondary uses
- the problems people are facing, who want to use health data for legitimate and beneficial reasons
- the overall approach to information governance relating to the use of NHS health data

This meeting was held under Chatham House rule. No statements within this report have been attributed to any individual person or organisation.

The meeting aimed to cover the benefits of (legitimate) access to health data repositories along with exploration of some of the problems (technical or governance related) encountered by those attempting to use health data for beneficial research or direct patient care and offer some recommendations.

Below is a summary of the range of **data platforms that currently exist** (or are proposed to exist in the near future) and are referenced within this document. This will help to ensure clear understanding of the distinction between each one and their intended purpose(s):

- **NHS England Federated Data Platform (FDP)**
  NHS England intends to procure a federated data platform, described as "an ecosystem of technologies and services to be implemented across the NHS in England". The stated primary purpose is direct patient care, but the intention also exists for secondary use of that data. No specific use case for research and development exists. *See more detail on the FDP in the next section.*
- **Trusted Research Environments (TREs)**
  A secure data repository that can be remotely accessed by (validated) researchers. Patient level data cannot be extracted. Queries can be defined and run, and aggregate results obtained to answer particular research questions. Focus is very much on research and not direct patient care. The standards for TREs are owned by the Office of National Statistics (ONS).
  TREs are country based and are not UK wide. Each TRE operates in a silo. There are no common standards. Re-identification of patients rarely possible.
- **Secure Data Environments (SDEs)**
  The NHS English National SDE is the Data store for NHS Digital TRE: NHS Digital is now part of NHS England
  Examples of key TREs in the UK = 4 locality-based NHS-owned SDEs within London, Wessex, Greater Manchester and the Thames Valley
- **Local Shared Care Records**
  Locality based (eg. Integrated Care Systems(ICS)) shared or integrated care records that vary greatly across England in terms of platform provider, approach and governance. It is intended that the FDP should be able to incorporate or complement existing data repositories at locality level.

Summary of Discussion

NHSE Federated Data Platform (phase 1 budget: £480m)

Procurement process

In May 2023, the Faculty of Clinical Informatics published a positional paper on the FDP which contained pertinent background information regarding the proposed data platform (this was not recapped during the meeting). The meeting attendees agreed that it was particularly difficult to find information regarding the FDP and its intention prior to the meeting. The official stated intention of the FDP, at the time of the meeting, was to serve as a replacement for the COVID-19 data store which was created to monitor trends during the pandemic and plan resource effectively (eg bed capacity or available ventilators).

It was acknowledged that the lack of available information regarding the FDP, particularly around the procurement process, is not helping to create an environment of trust. The group agreed that the questions raised by FCI (and others) must be answered in order to help to reassure and prevent scepticism from NHS employees and patients.

FDP structure

Tenancies

(a separated data area within the FDP only available to the respective data controller)

The FDP will not be mandated and therefore use by NHS Trusts and Integrated Care Boards (ICB) is optional. Data contained within the FDP will be de-identified for the national tenancy but *can* include identifiable data within the local tenancies (subject to appropriate governance arrangements being put in place and an established legal basis for access).

Each acute trust will have their own tenancy and will determine the content and access governance processes. Multiple trusts within an ICB can choose to link up to the ICB tenancy. The ICB tenancy can subsequently link up to the national tenancy but not all data will flow between them all; it will be selective. It is not clear whether this data flow will be two way or singular or who the data controller is in each case.

It was noted that there is no explicit link between the FDP and TREs and the FDP has no published use case for research.

The FDP is not about collection of data. It is about sharing and use of that data via a "privacy by design" principle, allowing ICBs to determine their own methods of use and scope of content. It will use cloud-based software.

Scope and Data Sources

Centralised collection of **GP data** is not currently in the scope of the national FDP and is not referenced in any of the use cases. GP data is available to ICSs who may "choose" to include it in their tenancy.

The FDP will initially be based upon secondary care data. It is likely that the purposes to which the data will be put, will be significantly different to those for which it was collected, which may lead to misleading conclusions.

The primary care record is so rich, it could be argued that it is more beneficial to start there and then flow secondary care data in later. If the long-term intention is to include primary care data in subsequent phases, the FDP structure should be designed to accommodate this at the very outset.

ICBs will have the flexibility to choose to ingest primary care data into their tenancy in the future. If that is the case, the group felt that the sooner they start thinking about the best approach (regarding IG and patient engagement) the better. In this instance, data will flow to the local repository only. It will not flow to the national data view (as per published plans at least).

Existing local health data repositories

Current approaches to locality-based health data repositories vary greatly with some areas very advanced and others with little access to data and varying degrees of integration of data sources eg. social care. This makes a unified approach difficult and for some, the ability to use data for planning or secondary uses is limited.

The FDP is an ambitious programme that will take a long time to establish fully. COVID-19 showed the benefit and need for readily accessible data. Few are in doubt about the potential benefits, however, previous attempts by the NHS to build a comprehensive, secure health data environment with appropriate governance alongside transparent communication to patients has not been successfully realised to date (eg. care.data and GPDPR). Historically, this led to significant loss of public trust shown by high levels of patient opt-out and lessons from the past should be noted in order to avoid a similar outcome for FDP.

If the FDP exists for direct care and secondary uses and the TREs exist for research and development, we can already see multiple systems existing in silos. It is unlikely that explicit definitions of research, planning, commissioning, direct patient care exist and quite often use cases include more than one of these. Add to this locality based shared care records (as currently the FDP has no data source available that could replicate the summary care record) and we have multiple systems co-existing, not a single system.

Information Governance

The FCI paper includes a reminder of the 8 Caldicott principles with particular attention drawn to principles 7 and 8 regarding duty to share information and keeping patients fully informed. The paper also includes a list of concerns regarding FDP and questions to help define the purpose, scope and governance requirements for the FDP.

Whilst the layered approach to FDP sounds flexible and could ameliorate appropriate access, it could equally result in 42 different approaches to the same platform by each ICB, particularly in relation to information governance. Standards of IG provision and understanding of legal guidance currently varies greatly between localities (with some being very poor, particularly understanding of *common law duty of confidentiality*). High quality IG training should be mandated for staff within these key roles.

NHS England states that a **Data Protection Impact Assessment (DPIA)** has been produced for the FDP procurement but as yet, this has not been published (Oct 23). It was asserted that templated IG documentation would be shared with ICBs and trusts to assist with standardisation (though many may choose to use their own) and there will be a national team to help localities ensure compliance and consistency. Nevertheless, some of this process may prove to be costly and complicated for ICBs. Publication of the DPIAs in good time might lead to improved confidence in the process.

Concern was expressed at the meeting about those localities that have invested large amounts of money in **local data repositories for shared care records.** The FDP should aim to ensure compatibility with whichever systems already exist, where desired by the locality. It might be anticipated that over time there will be a natural move away from local solutions towards the FDP, which may be encouraged by funding streams from the centre.

Vendor agnostic solutions

There was overall support for the idea of sharing and use of health data at a national level (for appropriate purposes). The importance of having a vendor agnostic solution with common standards, rather than reliance on vendor specific solutions, was felt to be important to ensure flexibility and value for money, particularly taking into account changing uses and technology.

Concerns

Systems and data to support direct patient care are very different to those for secondary uses such as planning and analysis. There is significant concern that conflating these uses may end up with systems which are not fit for either purpose and NHS and system support staff without the right clinical and informatics skills.

There is a real concern that the intention of the FDP is not clear but appears to be too big and too ambitious in its approach. Focus should be on secondary uses (such as the COVID-19 data store, which it is intended to replace) and once achieved to a satisfactory standard, progression might be made towards patient care applications.

For patient care applications, patient re-identification is essential which is not stated to be within the scope of Phase 1, however there is little published information on potential subsequent phases and long-term ambitions. The required level and granularity of data needed will vary enormously depending upon the direct patient care proposed and may not be available.

The key themes in any data platform are query dissemination, data linkage and data collection yet there is no mention of these in the procurement documentation.

There is a need for clinical steer to help clarify any questions users are trying to ask of the data. Without this skill input, the data platform and analysts alone may struggle to achieve anything of real clinical value or anything over and above what can already be achieved via a TRE such as OpenSAFELY.

There is currently very little awareness within the general public regarding the FDP or the ability to opt out of sharing your data with the FDP[1], with a notable absence of national communication. Information around opt out models generally are confusing and not well understood, even by those working in the NHS. Many patients believe that by opting out of data sharing for secondary uses, they are opting out of all data sharing outside of their own direct patient care, which is not the case.

It is understood that public engagement about the FDP has commenced via engagement with a patient participation group. Communications are planned to increase once a specific supplier has been confirmed. It is understood that the FDP will need to maintain the national data opt out model. It is important to note that patients cannot opt out of a data system (eg. such as a specific supplier) but they can dissent to a data flow (or the purpose of that data flow) though it was acknowledged that even the national data opt out has exemptions (eg. Faster Data Flows[2]).

[1] Since the meeting NHS England has said in their FAQ there will be no opt out for patients

[2] *NHS England's Pilot Faster Data Programme aims to create daily collections of patient data from acute care settings (the Providers)*

Opt outs for secondary uses are less impactful than opt outs that directly impact patient care. This highlights the risks of having a product or system that is attempting to provide both direct care and secondary data uses, without a clearly defined approach to opt out and dissent.

It isn't clear whether those patients who have previously selected a type 1 opt-out, or a National Data Opt Out for research and planning, will need to opt out again for FDP. ICBs will have to consider how to handle opt outs when merging different data sources with different consent models, where patients may have given consent for the use of different data at different times for different purposes. Lack of clarity in the extend and availability of any opt out, will not increase trust in the system. Our recommendation is that opt out should be a simple, transparent, and well-advertised process. Patients will not opt out if they have trust in the system.

IG discussion (general health data use)

The National Data Guardian paper (December 2022) on how to enable better public benefit evaluations when planning to use, or allow access to, data collected during the delivery of care for planning, research, and innovation projects, was highly recommended for the deployment of the FDP.

It was stressed that de-identified/pseudonymised data is NOT anonymised data. Data anonymisation can be considered like a spectrum with pseudo data being closer to identifiable than anonymised. The group acknowledged the good work in this area previously done by Ross Anderson (https://www.cl.cam.ac.uk/~rja14/#Med)

Concern was raised about a new draft *Data Protection and Digital Information Bill** being debated in parliament which aims to reform UK data protection regime following Brexit and the changes proposed which:

- would give data controllers within organisation the discretion to decide when personal data can be classified as anonymous,
- amends the definition of "scientific research" to allow access to personal data and
- introduces a new type of consent that allows permission from an individual to allow use of their data for scientific research can be assumed to apply to further projects that were unknown at the original time of consent.

Update:** ***This bill has since been replaced by Data Protection and Digital Information (No. 2) Bill *(https://bills.parliament.uk/bills/3430)*

Independent review panels

The group agreed that an independent group or board should have regular oversight of the collection and release of data flowing via FDP. Previous review panels for similar NHS data projects were referenced and are documented below:

- **IGARD** (which ceased on 31 January 2023) was the advisory board to the NHS Digital board and was formed to (independently) make general recommendations or observations regarding dissemination of data. Historically, GPES was approved as it had an independent advisory group which considered the collection as well as the release of information. The replacement (IGARD) only considered data releases (not collections). https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/independent-group-advising-on-the-release-of-data

- The interim advisory group for data (**AGD**) provides guidance on the release of data and consider all requests for the dissemination of confidential information. NHS England is establishing a new data advisory group to provide this in line with statutory guidance issued by the Secretary of State for Health and Social Care which has been published in draft.

https://digital.nhs.uk/services/data-services-for-commissioners/governance

The previous members of IGARD have continued, at the request of NHS England, to meet and provide advice on data access matters. They have been asked to form an interim data advisory group alongside a representative of the Caldicott Guardian, the Data Protection Officer, and the Data and Analytics function. A representative of the Senior Information Risk Owner will also attend meetings.

There is a concern that the AGD is no longer a fully independent group as it is understood that NHS England now has panel members (who were previously only observers). There was a concern that the independence of the panel could be open to change by the organisation whose procedures are being reviewed. As a result, it is recommended that that IGARD moves from NHS England to the National Data Guardian. This is a sensible place for any guidance to come from. All data uses (from NHS England) should then be passed through this truly independent group.

It was noted that the **DARS** (Data Access Request Service) is severely under-resourced and required further funding. This service offers clinicians, researchers, and commissioners the data required to help improve NHS services. The DARS team also manages the Data Sharing Agreement (DSA) applications.

It was agreed by the group that it would be beneficial to have a common understanding of what is considered “good information governance practice” and a set of resulting mutually agreed standards. All data sharing activity should be seen to be consensual, safe and transparent (this also means non-controversial). Patients should understand how their data is used and understand all their choices.

To demonstrate trustworthiness, the group recommended that NHS England should show the public what is happening to their data as part of the pilots for the Faster Data Flows project. It was felt that there was a difference between seeking trust and being considered trustworthy. Opting for soft, low-key product launches is not the best approach even if it does satisfy the *minimum* that is required.

Data Quality

There is often an assumption from users regarding the comprehensiveness, quality, accuracy, and reliability of available data. The group stressed that quality checks ***must*** be built into the foundation of any data platform and that investment is made into the improvement of record quality.

Interoperability is also a major consideration. The context and provenance of data needs to be preserved in order to retain meaning and ensure clinical safety.

List of recommendations from the Primary Health Care Specialist Group (PHCSG) regarding the Federated Data Platform

- The questions raised within the Faculty of Clinical Informatics positional paper should be answered to help to reassure and prevent scepticism from NHS employees and patients.

- More detailed information regarding the FDP including the procurement process, planned subsequent phases and long-term ambitions must be made available. This is imperative to the creation of an environment of trust.

- All DPIAs associated with the FDP should be published and in good time. This might lead to improved confidence in the process.

- If the long-term intention is to include primary care data (at some point) at any level, then the FDP structure should be designed to accommodate this data at the very outset. In addition, ICSs should start thinking about the best approach regarding IG and patient engagement as soon as possible.

- Standards of Information Governance provision and understanding of legal guidance varies greatly between local NHS organisations. High quality IG training should be mandated for staff within key roles relating to the FDP.

- For those localities who have invested large amounts of money in local data repositories for shared care records, NHS England should aim to ensure compatibility between the FDP and whichever systems already exist (where desired by the locality).

- A vendor agnostic solution with common standards is considered preferable to reliance on vendor specific solutions, to ensure flexibility and value for money.

- For those analysing or querying the data held within the FDP, it is strongly recommended that there is clinical input and guidance to help clarify any questions asked of the data. Analysts without clinical or domain expertise may struggle to produce anything of clinical value from any data platform.

- There needs to be a significant increase in communication regarding the FDP to the general public. Furthermore, there needs to be a simple, transparent, and well-advertised opt-out process. Information regarding opt out should be well understood by all NHS staff.

- ICBs should be clear on how to handle opt outs when merging different data sources that have different consent models or processing agreements in place.

- The recommendations made within the National Data Guardian paper (December 2022) on how to enable better public benefit evaluations when they are planning to use, or allow access to, data collected during the delivery of care for planning, research, and innovation projects, are highly recommended for consideration in the deployment of the FDP.

- In order to maintain independence, the group recommends that that IGARD moves from NHS England and over to the National Data Guardian. This is also a sensible place for guidance to come from. All data uses (from NHS England) should then be passed through this group including those of the FDP.

- To demonstrate trustworthiness using an existing example, the group recommended that NHS England should show the public what is happening to their data as part of the current pilots for the Faster Data Flows project.

- Data quality or validation checks must be built into the foundation of any data platform and investment should be made into the improvement of record quality.

- The context and provenance of shared data needs to be preserved in order to retain meaning and ensure safety. Interoperability of data is a major consideration.