The post-COVID landscape

COVID-19 has changed the NHS in fundamental ways; from the levels of PTSD in our nurses and doctors, to the configuration of health systems, to the expectations of health information systems from staff and patients alike.

There are a number of new information systems, several new bodies, as well as some new(ish) ‘partnerships’ and contractors. Quite a number of these have been assembled in haste in a crisis under emergency powers – not necessarily the ideal conditions under which to redesign a national health and care system. (If indeed any real ‘design’ effort is being put into social care...)

The data controllers

Under the COPI Notices, the Secretary of State (DHSC) has taken extraordinary powers, and given some to NHS England.

Service control, hence data controllership, of NHS patients’ data – as well as local authority and variously-contracted social care service users’ data – is by its very nature widely distributed. Data controllers being a person or body which, “alone or jointly with others, determines the *purposes* and *means* of the processing of personal data” – i.e. a matter of *fact*, not of (contractual) assignment or assertion. In several instances during the pandemic, however, data controllership has become unhelpfully muddied.

If we take a look at some of the high level constituent parts covered in ‘The data flows of COVID-19’ – only some of which most of the general public will be aware – then where does the data end up, and who is in control?

- **Test and Trace**, including the **NHSx exposure notification app** – DHSC is data controller; NHSx is neither a statutory nor formally constituted public body.
- **Vaccination systems** – data controllers vary along the data pathways.
- **Public Health England**, as it ‘merges’ into the ~~National Institute for Health Protection~~ **UK Health Security Agency**, along with the **Joint Biosecurity Centre**, Test and Trace, etc. – DHSC is the data controller, but executive agencies may themselves act as data controllers.
- **NHS COVID-19 Data Store** and **‘NHS Foundry’** – NHS England is the data controller; though Palantir clearly determines “the manner in which any personal data are processed” well beyond the competence of NHSE/x.
- **NHSx** itself (which is, as noted, not a data controller) and related initiatives such as the ‘Transformation Taskforce Unit’, and the ‘Data Alliance Partnership’ announced on page 17 of ‘Busting Bureaucracy’ which brings together *“key bodies such as the Care Quality Commission, NHS Business Services Authority, PHE(?) and NICE”* – unclear which, if any, of these apart from **CQC** are data controllers in their own right, or which other NHS bodies are involved.
- Projects such as the **National COVID-19 Chest Imaging Database (NCCID)**, NHSx’s **‘AI Lab’** and **‘skunk works’**, and **Project Oasis** (i.e. symptom reporting via commercial apps),

etc. – data controllers unclear; the DPIA for NCCID states the controllers are *“NHS England joint with DH, and also joint with NHS Scotland for Scottish data only”* but seems to ignore the data controllership of the contributing hospitals.

- **NHS Digital** data extracted for **secondary uses**, e.g. SUS/HES from hospitals, GPES data for COVID-19 pandemic planning and research, the Mental Health Services Data Set and various Adult Social Care collections – NHS Digital is the data controller, at least until it disseminates the data...
- **NHS Digital** data systems supporting **direct care**, e.g. Personal Demographics Service, National Record Locator, GP2GP, the Summary Care Record / SCRa, etc. – NHS Digital is the data controller.
- **OpenSAFELY** – under the COPI Notices, NHS England and GPs are data controllers of the records OpenSAFELY processes; outside the COPI Notices, NHS England is not.
- **Genomic data**, *to the extent that viral sequences are linked to individuals' personal data* – data controller(s) are unclear. The COG-UK website suggests it may be the Wellcome Sanger Institute, but (given the lack of information and no DPIA) it may equally be Genomics England Ltd and/or UKRI...

As the COVID-19 pandemic passes, and as the NHS stands down from emergency incident levels, the COPI Notices will expire. At that point, if not before, those crisis-driven data flows which remain useful will need to be transferred to a new or different legal basis.

COVID-19 has been a unique situation – not least because the entirety of the NHS, the Government, the research sector, commercial interests, civil society, and the population at large have all been pushing in the same direction at the same time to achieve the same thing. In more normal times, it can be difficult to get, e.g. NHS England and just *two* different hospitals to even want to push in the same direction at the same time – let alone NHS England and *every* hospital.

Quite aside from data sharing – much of which would have been lawful anyway, and all of which should have still respected DPA 2018 – what the COPI powers did in practice was to *remove the barriers to coordination*, as everyone was already aligned.

Once the “Regional or National” Incident is no longer in operation and the pandemic has receded, officials in health and care – as well as across the rest of government – should be prepared to account for everything they have done. Those who failed to meet their legal obligations, or who continue not to meet them, will likely face consequences. The regulator may have been lenient during the pandemic; preserving public trust is a much higher bar.

To take one example, the COPI Notices, under which vast quantities of patient data has been processed:

- Section 60 of the NHS Act 2001, under which the Health Service (Control of Patient Information) Regulations 2002 were issued, was repealed. The statutory basis for the COPI Regulations, and therefore the COPI Notices, now resides in Section 251 of the National Health Service Act 2006, of which clause (7) states: *“Regulations under this section **may not make provision for** or in connection with **the processing of prescribed patient information** in a manner **inconsistent with any provision of the data protection legislation**”* – that legislation being the Data Protection Act 2018. In other words, all processing of special category personal data (i.e. individual-level health data) must still have complied fully with UK GDPR.

GDPR requires that all processing of personal data is "lawful, **fair, and transparent**". Thus those NHS bodies which have failed to be transparent throughout the pandemic about what they are doing with citizens' health data – fairly and openly explaining what data is being processed for which purposes, who else is being given access, and what they are doing with it – will have breached a core Data Protection Principle, and will most likely have been (or will still be) breaking the law.

- NHS Digital has maintained its Data Access Request Service (DARS) throughout the pandemic. Indeed, it speeded up and improved professional oversight of some uses of data, while consistently publishing minutes, approvals and release registers – as well as DPIAs, Data Provision Notices and extensive transparency information, e.g. on GPES Data for Pandemic Planning and Research

- By comparison, NHS England had to be threatened with legal action before it would publish contracts it had handed to companies like Palantir and Faculty Science. It has failed to publish a DPIA for major programmes, and has never once published a data release register for its COVID-19 Data Store. The incomplete and sketchy information it does provide on the Data Store provides little insight into what is being done with the medical data of 55 million people (and more besides).

A clear distinction has emerged amongst NHS bodies; those which have demonstrated themselves capable of being trustworthy data controllers, and those which have not. All of them have been under unprecedented pressure, so there is no excuse. And those bodies which have failed to meet their most basic legal obligations like being lawful, fair and transparent must change both practice and culture if they expect public trust and confidence.

The existing moving parts

Quite aside from the large number of contracts let during the pandemic, NHS data / digital **procurement** was already undergoing a major 'revamp' – beginning with GP IT Futures, which replaced GPSoC ('GP System of Choice') in January 2020, following the creation and piloting of a Digital Care Services Buying Catalogue for data services for primary care. (N.B. NHS Digital also offers a 'procurement pipeline' for potential suppliers.)

The roll-out of this "pan-NHS **procurement platform**" may not have gone as originally planned, but the future scope is for it to expand to cover all care settings.

- The first six approved suppliers on GP IT Futures in December 2020 were for (COVID) appointment booking solutions, which some practices have been buying unwisely.

- The Crown Commercial Service has also set up a new £800 million 'Digital Capability for Health' agreement, which is set to run for 4 years; twelve consultancy suppliers have been listed, but other companies (e.g. Faculty) have been able to 'piggyback' themselves in.

These frameworks and 'lots' within them may launch with a wide range of suppliers, each of which has had to go through some sort of approval process, but the extent to which this diversity – and critically, *interoperability* between every solution – can be maintained remains to be seen. As with

GP IT systems over the past decades, once a few players gain significant market share for a particular type or types of functionality, it can be more difficult for smaller organisations or new entrants to compete.

How will this look in practice? The emerging NHS ‘stack’...

Potentially massive efficiency savings *could* be available were, e.g. machine learning approaches applied to functions like health and care transportation or NHS energy use, but – setting aside such vital operational functions, and focusing instead on the predominant obsession with patients’ health data – in broad-brush terms, several competitive arenas are already emerging:

- **Cloud** – is already beginning to look like a ‘two horse’ race in UK health between Microsoft Azure and Amazon AWS, with Google Cloud and other providers like IBM, e.g. offering deep discounts in an attempt to buy their way in. What is vital to note is that each of these cloud service providers also have their own health data ambitions:
    - While Microsoft finally shuttered Healthvault in 2019, it launched Microsoft Cloud for Healthcare last year – including ‘Microsoft Health Bots’, which healthcare organisations can use to build and deploy AI-powered ‘virtual health assistants’ and chatbots. Far from simply flowing data across systems and through FHIR APIs, this involves sophisticated processing of individual-level patients’ data and the generation of tools and IPR that may be exploited/exploitable elsewhere. Cui bono?
    - Amazon also has long-held ambitions in health, and not just in the NHS. It was, e.g. the basis of a prototype and existing ‘shadow’ health record for DWP, and was allowed to write its own deal for NHS content (and the NHS logo) when Matt Hancock got all excited about Alexa. As one would expect, AWS offers a wide range of ‘healthcare solutions’ across infrastructure, applications, analytics and AI, user experience – offering many ‘building blocks’ that third parties incorporate into their products and services, thus tying the end customer to Amazon’s platform.
    - Google’s cloud offering is a well-funded third place, but looks like it will keep losing money until it doesn’t. Google’s track record with the NHS, and in health generally, has been far from spotless – notably the Google DeepMind and Royal Free scandal, and then Google Health reabsorbing DeepMind despite promises it never would. Also Project Nightingale in the US – highlighting just how easy it is for a single clause in an agreement (*“Exploring artificial intelligence / machine learning applications that will have the potential to support improvements in clinical quality and effectiveness, patient safety, and advocacy on behalf of vulnerable populations, as well as increase consumer and provider satisfaction”*) to shift the relationship from straightforward, well-delimited data processing, to open-ended data controllership and R&D. In such negotiations, when the NHS sends doctors, Google sends $500-an-hour lawyers and trained negotiators.

  The issues aren’t exclusively about the choice of infrastructure, and the extent to which each provider will always try to lock the sector into its own platform. Rather, it is necessary to take account of these companies’ strategic intentions – hence clauses in contracts that turn the company from being a data processor into data controller, with the right to, e.g. ‘improve’ its own software using identifiable patient data flowing across its platform.

Such secondary uses are nothing less than direct commercial exploitation of patients' data, presented as an 'efficiency benefit' to NHS organisations as customers, who pay three times: for the services they buy on contract, with the patients' data they hand over to companies, and in each and every future price hike for 'new and improved' services; none of these promised health AI bots and insights will come for free!

- **Electronic Health Records (EHR)** – who will be the preferred suppliers at Trust / LHCR / ICS level? While a flurry of procurement frameworks suggests another attempt at a competitive market, in reality most GP practices still only buy one of three IT systems – either EMIS web (AWS) , TPP SystmOne (AWS) or INPS / Cegedim Vision (Azure) – to manage their patients' records, and most hospitals buy one of these:

    - System C / Graphnet (Azure);
    - Cerner (AWS);
    - Epic (AWS also possibly Azure, but not Google);
    - Orion Health (AWS)

  (The market situation in social care is broadly similar: three systems – LiquidLogic (part of System C / Graphnet), OLM and Servelec – dominate the market, with a handful of other players such as Careworks, Azeus, Civica and TPP offering other solutions.)

  From a 'stack lock-in' perspective, the cloud partnerships listed above are not insignificant. For example, while FHIR (Fast Healthcare Interoperability Resources) has become a standard for health data exchange, and all of the above suppliers support it, as the dominant players battle it out in the marketplace will different 'flavours' of FHIR emerge, with "optimisations" for one cloud ecosystem versus another? Similar has happened in tech, many times before...

  N.B. The NHS in Scotland and Wales and a couple of Trusts in England are exploring openEHR which, while notionally deployable on AWS and Azure clouds – and Google's too – appears more problematic to deploy on Azure.

  It is worth noting that the 'home grown' player in the mix above (System C / Graphnet) appears to be out on a limb with Microsoft. NHS England uses Azure cloud, e.g. for the COVID-19 Data Store, and NHS Digital uses them all – not least for resilience. But the real prize that everyone is playing for is...

- **The 'Shared Care Record'** – which has been DHSC and NHS England's long-term ambition since long before the Target Architecture and 'Data Lake', through the legislative proposals for the Long-Term Plan, and now in the Health and Care Bill 'Integration and Innovation' white paper. This may be somewhat difficult to pull off, if everyone is already locked into non-interoperable EHRs with different schemas...

  ...though that might work out very well for Palantir.

We go into more detail on ShCR in other documents linked from our separate Shared Care Records piece.

Consensual, safe, and transparent?

Our driving concern remains that *every* use of patients' data is consensual, safe and transparent. In practice, this essentially boils down to a number of practical measures and processes:

- National Data Opt-out (NDOP), Type-1 objections, SCR opt-outs
- Safe Settings, Trusted Research Environments (TREs), OpenSAFELY
- Data Usage Reports, Analysis and Inputs Reporting, public communications

...along with the governance, oversight and accountability that goes with them. These are probably easiest to demonstrate in a context that has received greater attention over the years than some may have liked.

medConfidential publishes a scorecard that we will update, and apply to a bunch of new bodies as the new legislation becomes clearer.

GP data

In May 2020, NHS Digital issued a Data Provision Notice for a new collection of **GPES Data for *Pandemic* Planning and Research (GDPPR)**. GPES, the General Practice Extraction Service, has for that decade collected information from GP practice systems for a wide range of purposes; the majority are to do with payments for GP practices, but some are for other uses like research.

This GPES collection for COVID-19 purposes was/is clearly valuable, and arguably necessary – but, as the DPIA shows, GDPPR extracted legally-protected codes, and included details such as people's medications and their strengths. And some of the onward disseminations ignored people's National Data Opt-outs, which is explicitly promoted to patients on the basis that it will *exclude* their data from being used in "planning and research".

Officials may – and no doubt will – offer justifications for each of these moves, just as DHSC and NHS England / NHSx keep justifying their failure to deliver on transparency promises made back in March 2020 (and before). But perhaps these bodies should be required to deliver on their existing promises before expecting patients and frontline staff to trust them to make any more?

A sort of independent advisory group – akin to GPES IAG, the review body that first spotted care.data's intended extraction of GP data in 2013, which was abolished in 2015 – was (re)instated during the pandemic. This new 'Professional Advisory Group' is now part of NHS Digital's Data Access Request Service (DARS) for the COVID-19 extract at least, but it is unclear what the BMA and RCGP are doing about all of the still-undelivered promises of transparency.

What basis is there for trust, other than NHS bodies, DHSC and the rest of Government keeping their word?

Consensual?

A significant amount of patients' GP data across England has now been extracted, respecting people's Type-1 opt outs but not always respecting their National Data Opt-outs. This data includes:

- diagnoses and findings

- medications and other prescribed items
- investigations, tests and results
- treatments and outcomes
- vaccinations and immunisations

each data item extracted from each person's GP record in each one of the above categories being linked with the following personal information:

- NHS Number
- Postcode
- Address
- Surname and Forename
- Sex
- Ethnicity
- Date of Birth
- Date of Death

NHS Digital claims to have done patient communications, but whatever NHS bodies cite and claim happened at the time, little evidence seems to remain and long term effectiveness is therefore doubtable. Depending upon what they said, there will be measures of success.

In a similar vein, earlier in the pandemic, reflecting normal practice, NHS 111 told patients that the data they gave to the NHS COVID-19 data store respected patients' general disssent, and then quietly decided that they didn't. The problem is not that either of these decisions were necessarily wrong; it is that NHS 111 changed its mind and kept quiet about it, hoping that no-one noticed. (We did.)

Safe?

NHS Digital does now have, within its Data Access Environment (DAE), a Trusted Research Environment (TRE) which it says operates on the Five Safes model. And with regard to the GPES Data for Pandemic Planning and Research (GDPRR) extraction that began in May 2020, the Data Protection Impact Assessment (DPIA) explicitly states:

> ***All requests*** *for data held by NHS Digital,* ***including GDPRR data****, for aggregate, identifiable and pseudonymised will be accessed through the Data Access Environment (DAE) within DPS,* ***excepting*** *those cases where DAE cannot support the requirements of that customer at that time and/or where data must be disseminated to the customer based on specific need.*

So how did that promise hold up? What data, if any, has been processed within the TRE only? And how much (new) data was still disseminated? **Was "excepting" the exception, or the norm?**

The latter two questions are easy to answer: NHS Digital's monthly data release registers show every dissemination of GDPPR data since the very first to have been a data release, not controlled access via TRE:

- July 2020 (all links are to Excel workbooks) - University of Oxford
- August 2020 - University of Oxford & HDR UK ongoing
- September 2020 - Imperial College one-off, add NHS England ongoing
- October 2020 - UK Biobank one-off & **103** "Identifiable, Sensitive, Ongoing" flows to ONS

- November 2020 - add Genomics England, ONS drops to 6 ongoing flows
- December 2020 - University of Sheffield one-off, ONS drops to 1 ongoing flow
- January 2021 - add University of Cambridge ongoing
- February 2021 - add **4** PHE flows, University of Manchester & **46** CCG one-off releases

If NHS Digital did not (could not?) keep its “TRE-first” promises for the massive extractions it did of patients’ GP data in 2020-21, then – given it shows no sign of remedying the situation – why did it make it? And how should patients or the profession trust any further promises?

The GP collection was not the only COVID-related extraction of data. On 17 March 2020, both the Secretary of State and NHS England issued Directions to NHS Digital; Sec State issued the **COVID-19 Public Health Directions 2020**, which were amended two weeks later, and NHS England issued the **COVID-19 public health NHS England Directions 2020**, also amended two weeks later.

During the course of the pandemic, and quite aside from the COPI Notices issued to NHS Digital, these two Directions provided the legal basis for NHS Digital to create a wide range of new ‘information systems’, for which it issued corresponding Data Provision Notices:

- Adult Social Care Management System Coronavirus (COVID-19) Status Collection
- Child Protection Information Sharing (CP-IS) for COVID-19 Child Health Safeguarding
- COVID-19 At-Risk (Clinically Extremely Vulnerable) Patients Data
- COVID-19 Situation Reports
- COVID-19 Unpaid Carers Data
- COVID-19 Vaccine Data
- COVID Oximetry @home Data
- Electronic Prescribing & Medicines Administration (EPMA) Data
- Hospital Electronic Prescribing & Medicines Administration (HEPMA) Data - Scotland
- National Cardiac Audit Programme (NICOR) Data
- National Vascular Registry (NVR) Data
- PHE Second Generation Surveillance System (SGSS) & COVID-19 Hospitalisations in England Surveillance System (CHESS) Data
- Shielded Patient List - Cancer and Rare Diseases Data

As the data release registers for this period show, NHS Digital is still providing its customers with copies of linked individual-level pseudonymised data which it describes as “Anonymised - ICO code compliant”, referring to the Code of Practice on Anonymisation issued by the ICO in November 2012, *six years* before GDPR, which has now been withdrawn:

> *“It used to be the case that if the patient data planning had been anonymised in line with the ICO’s code of practice on anonymisation it was no longer subject to the requirements of the common law duty of confidentiality, and ethical review from the HRA was not required. However,* ***this Code of Practice has been withdrawn*** *pending an update and we await further guidance.”* – DHSC’s Guide (no longer a ‘code of conduct’) to good practice for digital and data-driven health technologies, January 2021. DHSC later changed that blog post to claim the Code is still “useful”, but useful is not necessarily lawful.

Dame Fiona Caldicott said there should be “No surprises” for patients, and Dame Onora O’Neill suggests three tests for whether someone is trustworthy: Are they competent? Are they honest? Are they reliable? With regard to ‘safe’, NHS Digital has yet to show it is.

Other rapidly stood up programmes like Test and Trace have shown they cannot stay within the law, much less train thousands of call centre workers how to do medical-grade confidentiality when handling millions of people's details. As the statutory safe haven for patients' data in England, NHS Digital must not only **do** much better; **it must *show* that it has.**

Transparent?

With no transparency comes no accountability, or trust!

Institutions that wish to be trustworthy, and to be seen as trustworthy, need to provide details of the projects that used data. You need to tell people what projects were done, why, and who did them.

Privacy notices are not enough. Issuing pages and pages of legalese and NHS jargon, focusing narrowly on technical pathways to legitimise what's being done, may meet the letter of the law on 'transparency publication' but will never command full professional or public confidence while the words and actions ignore both the spirit and the full corpus of law.

Public communication is a chronic failing of central NHS bodies, which – not routinely providing care to patients – have no direct relationship with the population at large. This must change.

For programmes involving patient-level data, transparency must be **end-to-end**. This begins with clear communication to professionals and the public, describing in plain language what is intended, and including the choices people have and the means to exercise those choices. It requires more detailed, comprehensive explanations of exactly what data is being processed (extracted, treated, stored, linked, excerpted, disseminated) and the conditions of and for such processing. Evidence of the information governance and oversight applied to data processing must be published regularly, especially regarding applications / approvals to access the data. And people should know how **their** data is used.

Issuing massive monthly spreadsheets, listing hundreds of data releases – some of which rely on a withdrawn Code of Practice – may be the necessary minimum for historical accountability, but fair and transparent patient communication requires plain language web pages in addition to these densely-detailed spreadsheets. It is also the case that to claim "robust" contractual relationships with customers who breach both your contract *and* the law without meaningful sanction is unwise.

medConfidential welcomes the creation of a BMA / RCGP 'Professional Advisory Group', which plugs a hole that has existed for half a decade, and which should allow the DARS process to be properly applied to data applications across a range of clinical specialities, with the Cancer Registry going first. As with the missteps which led to the Cancer Registry being moved from PHE into NHS Digital, safe access and full transparency are both necessary – and it is far better to be proactive than reactive after failures, such as giving cancer patient information to a 'causes of cancer' study run by a tobacco company…

What next?

The COVID-19 pandemic is not yet over, and may not be completely over for some time. NHS England's COVID data contract with Palantir runs until December 2022, which is the earliest point at which pandemic vigilance could be stood down.

While there has been a more "permissive approach to data sharing for care during the coronavirus response", as described in DCMS's National Data Strategy, this 'permissiveness' has come about through the temporary alignment of Governmental, public bodies', commercial and public interests in a time of global crisis and, notably, by the imposition of extraordinary statutory measures that would face public outcry and widespread challenge in more normal times. It has not created a shift in public or professional attitudes as yet.

Is data 'sharing' really 'permissive' when it is required or mandated by law in a public health emergency?

Around health data, as we refer to above, these statutory measures are not solely those in the Coronavirus Act and hundreds of COVID Regulations, nor even the COPI Notices. Both the Secretary of State and NHS England issued COVID Public Health Directions which have been used to legitimise unprecedented data extractions, including significant quantities of patients' GP data – which when previously attempted, outside pandemic conditions, has been highly controversial (cf. care.data).

With impending NHS legislation, as laid out in the recently-published white paper, and another top-down reorganisation of the NHS into Integrated Care Systems – with consequent issues around shared care records, and interoperability across administrative boundaries and the systems and practices of health and social care – before any new or lasting data programmes are begun there must be a full, transparent accounting for the decisions, actions and outcomes of the pandemic. We cover this more in our consideration of Shared Care Records.

Genuine appreciation for all that staff have done must not be allowed to feed political and bureaucratic data delusions and dashboard envy, nor be twisted into "Well, it worked in the pandemic" half-truths used to justify yet more tech solutionism.

As ever, it is the *right* lessons that must be learned!

medConfidential