The data flows of COVID-19

Drawing on published documents, media reports and various individual and crowdsourced efforts listed in 'Other references' below, what **do** we know about the (health and social care) data flows of COVID-19? We all know there's far more than just dodgy [PPE contracts](#), [Serco](#), [Faculty AI](#) and [Palantir](#) to be concerned about – but few seem to appreciate the full extent and very few the detail of what has been going on.

N.B. This document covers what has been going on with health data in England in 2020, with occasional references to what has been happening in the other nations.

Legal basis

Firstly, in terms of the lawfulness of data flows, what changed?

Much of what has been done with data during the pandemic has been enabled by the **[COPI Notices](#)**, first issued in March 2020, extended in July 2020, and extended again in February 2021. At the time of writing, the COPI Notices will expire on 30 September 2021.

What is perhaps less well known is that on 30 January 2020, a "**Level 4 National Incident**" was declared by NHS England. This was six weeks before the Government was seen to act.

In a crisis, the NHS operates at one of four different incident levels – described on [page 11 of the National Incident Response Plan](#) – that is decided according to the severity and impact of the incident on the health service. At Level 4, NHS England takes control of the overall coordination of the NHS's response. What this means in practice is that, in data terms, **NHS England took operational control of care providing bodies such as hospitals, and so became a joint data controller of all of the data flowing through and around the NHS at the very beginning of the pandemic**.

(The Level 4 incident has not, however, persisted throughout the pandemic. It was de-escalated to Level 3 on 1 August 2020, and was [re-escalated to Level 4 on 5 November 2020](#).

This de-escalation (under which control returned to NHS England's regional structures, rather than being under NHS England's *single national* control) did not materially alter what NHS England did with tens of millions of people's data during those three months but, as the pandemic subsides, NHS England's data controllership and its and other bodies' lawful bases for processing much of the data they have done during COVID will expire.

Meanwhile, and contrary to [the impression the regulator gave](#), the Data Protection Act 2018 and UK GDPR have *still applied* to all processing of personal data during the pandemic.

The statutory basis for the COPI Notices is [Regulation 3(4)](#) of the Health Service Control of Patient Information Regulations 2002, but [Regulation 7](#) makes it clear that confidential patient information must not be processed *"more than is necessary to achieve the purposes for which he is permitted to process that information under these Regulations"*, that such data should be de-identified (as far as is practical), that only authorised persons (health professionals or those with an equivalent duty

of confidentiality) have access, and that *"appropriate technical and organisational measures are taken to prevent unauthorised processing"*.

Furthermore, the original basis in primary legislation for the 2002 COPI Regulations (and thus the COPI Notices) is no longer the Health and Social Care Act 2001. The relevant Section 60 of that Act was repealed and replaced by [Section 251 of the National Health Service Act 2006](#), which at 251(7) clearly states that:

> *(7) 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]**.*

The data protection legislation being:

> *(13) In this section—*
>
> *["the data protection legislation" has the same meaning as in the **Data Protection Act 2018** (see section 3 of that Act);]*

Thus, while the ICO may have adjusted its [regulatory approach](#) during the pandemic, breaches of DPA2018 / GDPR are still breaches – and organisations and bodies which fail(ed) to follow the law and the [Data Protection Principles](#) within it may yet face consequences. The ICO investigates almost all incidents in retrospect, not in advance.

There was an assumption that people working on the COVID response were working in the interests of public health, not personal profit – as more information emerges over time, any project which was "inconsistent" with the COVID response can be prosecuted as a Data Protection breach.

(Not forgetting, of course, the data-specific elements of all of the Coronavirus / COVID [legislation](#) and [statutory instruments](#).)

Data programmes and initiatives

A number of high profile new programmes were established, all of them justified by the COVID pandemic, many of which involve the processing of significant amounts of special category (i.e. health-related) personal data and/or confidential patient information. What follows is a brief outline of each one, including issues arising and questions that remain to be answered on the data aspects of each initiative:

1) DHSC 'Test and Trace'

One of the more controversial programmes of the pandemic, "NHS Test and Trace" (T&T) has an [overarching privacy notice](#), last updated on 11 January 2021, which refers to other privacy notices that have been published at various points during the pandemic, e.g.

- [Coronavirus (COVID-19) Testing: privacy information](#)
- [NHS Test and Trace Contact Tracing privacy notice](#) (on a PHE domain)
- [NHS COVID-19 app: privacy notice](#)

DHSC declares itself to be the data controller in the overarching privacy notice:

> *The Department of Health and Social Care ('DHSC') has commissioned the NHS Test and Trace programme on behalf of the UK government and will be the data controller for the purposes of data protection legislation.*

But DHSC still has **not published its Data Protection Impact Assessment** for the Test and Trace programme as a whole, despite a [legal challenge by Open Rights Group](#) in June/July 2020.

DHSC does now publish [weekly statistics on Test and Trace](#), with a two week lag, and the Health Foundation produces a fortnightly [Test and Trace performance tracker](#).

1a) Testing – the data issues around testing are mainly to do with where data is being shared / copied, including whether positive test results always make it to where they should (e.g. to patients themselves, to their GP record, and to other organisations – both national *and* local). The testing programme, which involves a number of commercial providers as well as NHS and PHE organisations, collects a significant amount of personal information which may include:

- first and last name
- date of birth
- gender
- ethnicity
- landline and mobile phone number
- email address
- home and delivery address, including postcode
- vehicle registration number
- National Insurance number
- NHS number, if known
- details and date of onset of symptoms
- vaccination status
- your NHS login account identifier (if you access our services using your NHS login details)
- recent travel history (e.g. whether you travelled overseas in the last 14 days and the country you spent most time in)
- employment details
- other household members' details, if you are ordering or booking a test for them as well
- whether you tested positive more than 90 days ago

There was some early confusion as to which body would be retaining Test and Trace data, and the period for which it was to be retained, e.g. '[Public Health England will keep personal data of people with coronavirus for 20 years](#)'. This period has subsequently been amended and is now anything between 14 days and up to 8 years "for patient record" (10 years for "legal complaints") according to DHSC's [overarching privacy notice](#).

There has also been confusion around the 'reporting lines' within the programme, with the responsibility for different types of testing being spread across the four Pillars, and periodic outbreaks of controversy around the [reporting methodology](#) for testing data and specific failures in the handling of it, e.g. '[Excel: Why using Microsoft's tool caused Covid-19 results to be lost](#)'.

Locally, there have been concerns around walk-through and drive-through testing facilities, including instances of people being sent other people's (negative) test results via SMS.

By early 2021, many local authorities now have '**Local Tracing Partnerships**' that first emerged in the summer of 2020: "100" LTPs by October, "200" by November – with another 100 "in the process of being implemented" at that point.

1b) Contact Tracing – even more contentious has been the *Tracing* side of Test and Trace. Not least because of the involvement of Serco, Sital, et al. and the early sidelining of local public health teams. Contact tracing is a long-established public health protocol, but standing up an entirely new system, staffed with minimally-trained and often non-medically-experienced new recruits has been problematic. Not to mention expensive.

There were early concerns at reports that some contact tracers were sharing (and possibly filling in) personal data of 'contacts' on Facebook and WhatsApp – and questions as to whether all of the personal data being collected was actually being used, or whether some 'types' of contact were being 'triaged' and lots of information was simply being ignored. This raised further questions of where and on what systems all of the contact tracing-related information is being stored, and the controls put around it – especially given a new hire while *technically* (contractually) sharing the same duty of confidentiality as a registered medical professional is clearly not going to appreciate or manifest it in the same way.

More recently, it has emerged that contact tracers working on Sitel's Test and Trace operation were told to use their personal e-mail addresses to share information about cases. While this would appear to be a clear breach of data protection law, if unreported by Sitel to DHSC, it would also be a significant breach of contract.

Another major concern has been DHSC's decision to allow **police access** to self isolation-related information, a matter on which we sent an FOI request to DHSC for the MoU (which was refused on the grounds that the MoU was "intended for future publication" – though, despite the law being changed in the interim, it still hasn't been published) and we wrote to DHSC (and others) directly on 24 October 2020.

The Local Testing Partnerships (see above) appear to be one route via which the police are being passed details of people who are non-responsive to either T&T or LTP contact tracers, an issue which blew up again in January 2021 when the Government updated the Self-Isolation Regulations.

(Though its contribution to actual *contact tracing* appears to be minimal, despite its much-vaunted 'exposure notification' function, there is also the NHS COVID-19 app – which we cover in the NHSx section below.)

2) NHS Digital

Throughout the pandemic – and mostly unnoticed, as things continued to work – NHS Digital has been maintaining and updating critical systems and providing data and programme support.

- In March 2020, NHS Digital added **COVID codes** for statuses such as low, moderate or high risk of developing complications, and events such as a positive diagnosis of COVID-19 to the NHS Spine.

- NHS Digital also added [additional information](#) including people's significant medical history (past and present), reasons for medications, care plan information and immunisations, and enabled wider access (via [SCRa](#)) to patients' **Summary Care Records** – continuing to respect existing opt-outs. What will be done with all this additional information after the pandemic remains an open question, with potential implications for the development of the 'Shared Care Record', but [patients' individual rights in relation to their SCR](#) remain clear.

- NHS Digital also manages the **[Shielded Patient List](#)** (after an [initially](#) [rocky](#) [start](#) under NHS England) and guides its [ongoing use](#) – as well as a [host of other COVID-19 initiatives](#), including:
  - Using the **[QCovid](#)** risk prediction model[1] to support the [NHS coronavirus response](#). The model was commissioned by the Chief Medical Officer via NERVTAG [in June 2020](#); the study was led by Professor Julia Hippisley-Cox, and the research protocol and papers are [published here](#). A QCovid score is derived from a range of factors such as age, sex, ethnicity and existing medical conditions and is used to predict risk of death or hospitalisation from COVID-19.

    QCovid was approved by NHS England on [16 February 2021](#), and NHS Digital began using it to identify additional people to be added to the Shielded Patient List (SPL). NHSD has published a [full list of the factors](#) used by the model, in the approximate order of their impact on coronavirus outcomes, but the original model itself has not been published

  - Quite separately from the daily statistics on the [GOV.UK Coronavirus Dashboard](#), NHS Digital publishes a number of other **data dashboards**, such as [Coronavirus in your area](#), open data about the [Shielded Patient List](#), [NHS Pathways](#) for COVID-19, as well as more 'standard' dashboards such as Activity in NHS Hospitals and [NHS e-Referrals](#).

  - NHS Digital stood up various services for patients (e.g. for [getting an isolation note](#) and, with NHS England, [COVID Oximetry @home](#)) and health and care staff (e.g. [Emergency Department Digital Integration](#) and support for the Remote Clinical Desktop) as well as maintaining a single COVID-19 ['Governance Hub'](#) for all of these.

- Indeed, the statutory Safe Haven for England, i.e. NHS Digital, not only maintained but *improved* its Information Governance – effectively reinstating BMA and RCGP scrutiny of some applications for data, via a 'Professional Advisory Group' not dissimilar to the independent advisory group which first spotted care.data in 2013. NHS Digital also sped up its processes by increasing the frequency of IGARD meetings, and maintained transparency – regularly publishing [minutes and approvals](#), as well as a [monthly data Release Register](#) and separate [COVID-19 release register](#) throughout the pandemic – proving that it is far easier to stay within the law if you have a culture of transparency, and have embedded the governance and processes necessary for trust in all of your operations.

- During the pandemic NHS Digital has made several COVID-19 related data extractions, most notably of **[GPES data](#) [for COVID-19 pandemic planning and research](#)**, and of various **Adult Social Care** collections (covered in more detail 'Social Care' section below).

---

[1] For which none of the codes have been published for wider review, with other questions about missing data: [pulsetoday.co.uk/news/coronavirus/over-400000-patients-given-inflated-covid-risk-scores-due-to-missing-data/](#)

- We picked up what was (at the time) being called the 'tactical'* GPES extract in May 2020; analysing the affected codes and keeping notes and a **timeline** as we engaged with NHS Digital.

    > * *with the clear intent that a much broader (non-COVID) 'strategic' extract was to follow...*

- We were assured that all uses of GPES (i.e. GP) data would be via NHS Digital's Trusted Research Environment (TRE) service for England

- **Type-1 opt outs** *were* respected (i.e. that data *wasn't* extracted), but **National Data Opt-Outs** were only considered "on a case by case basis" and may not have been applied "depending on the specific purposes for which the data is to be used".

3) NHS England's 'COVID-19 Data Store'

In October 2019, useMYdata publicly asked why NHSX told their members there was a public release register, when there was no public release register in October 2019, and as of March 2021, there still isn't.

The Data Store (which NHSE/I/x initially referred to internally as "PPDS"; a "Privacy Preserving Data Store") first came to public attention in April 2020, when confidential documents used by Palantir, Faculty and NHSX officials to plan and develop the COVID-19 Data Store were exposed via an unrestricted portal. Civil society followed up with an open letter to Matt Hancock in May and, at the beginning of June, openDemocracy and Foxglove forced the publication of the four initial contracts with **Faculty**, Google, Microsoft and **Palantir**.

The **COVID-19 Data Store timeline** covers, i.e. the Data Store DPIA, privacy notice and Data Store "Reference Library" – that were published at the same time. Further contracts were issued in subsequent months:

- **Palantir:** NHSE contract 1 (Mar - Jun 2020), DHSC contract 1 (Jun - Sep 2020), NHSE contract 2 (Jun - Oct 2020), NHSE contract 3 (Dec 2020 - Dec 2022)

- **Faculty Science:** NHSE contract 1 (Feb - Aug 2020), contract 2 (Apr - Sep 2020) & contract 3 (Mar - Aug 2021, see "NHSx AI Lab" below)

In brief, the **COVID-19 Data Store**, of which NHS England is data controller, collects masses of data – some of which is operational data, a great deal of which is **personal data** – from a wide variety of sources; the 'Reference Library' of data sets within the Data Store is incomplete, and no comprehensive list has ever been published.

The personal data is de-identified / pseudonymised and then stored in 'Data Marts' in a Microsoft **Azure** 'Bastion' controlled by NHS England. The Data Store DPIA reveals, however, that Arden and GEM CSU has the pseudonymisation key (i.e. the ability to re-identify pseudonymised data) so that it can link PHE data with other datasets:

> *Pseudonymisation will be undertaken as per NHSE/I processes. Once the PHE data has been pseudonymised, the key will be provided to AGEM Data Services for Commissioners Regional Offices DSCRO in order that they can then de-identify any other required datasets for linkage purposes.*

Pseudonymised individual-level data, aggregated data and operational data is then extracted to an **Amazon Web Services** data store in London, on which an instance of **Palantir's** Foundry is running. In the early stages of the pandemic, Faculty built a number of 'dashboards' to help inform decision-makers and these and the analytic capabilities of Palantir Foundry are used by an unknown number of people and organisations for a range of unknown purposes. The purposes we do know of are those laid out in those various contracts:

1) **'Self-Service' Data Integration and Analytics Capability** – allowing users to compile their own custom reports for sharing and export; *"integrating activity data sources, modelling activity data in an object-based data model (ontology) and performing analytics and reporting tasks to support operational decision-making"*;

2) **Dashboard to manage resources used by projects using Foundry** – including users *other than NHSE itself* such as "Local Systems", i.e. NHS bodies;

3) **Strategic Decision-Makers' Dashboard** – which (according to the contract) explicitly includes *coordinating the national response to EU Exit*! If it relates to personal data, such functionality would breach the terms of the COPI Notices, which explicitly only cover COVID-19 purposes;

4) **Recovery of Critical Services tool** – which provides *"the ability for the Buyer to transition this Tool for general business-as-usual monitoring"*, again raising questions about whether these are COVID-19 purposes, not to mention NHS England's long-term ambitions;

5) **Early Warning System** – developed out of a bunch of *ad hoc* modelling done in the early stages of the pandemic; EWS was not even mentioned in the original contract with Palantir, though some of the models now in EWS do appear in the first Faculty contract. In essence, EWS is a three-week forecasting tool for several critical NHS capacities, e.g. ICU beds, ventilators, oxygen – but not all, e.g. care homes;

6) **Supply Management Capability** – for managing the inventory of PPE, ICU consumables and ICU equipment; also a Sales & Operational Planning tool for managing things like the PPE supply chain which has experienced significant controversy, including damning reports from the [National Audit Office](#) in November/December 2020 and from the [Public Accounts Committee](#) in February 2021;

7) **Immunisation and Vaccination Management Capability** – does demand modelling based on (near real-time) data collected under COPI powers and, being plugged into the suppliers' and wholesalers' systems, can be used to order or allocate supplies of the vaccines;

8) **Workforce Analytics Capability** – understood to be for internal NHS England use (the redacted "Data Classes" in the £23m contract are fields of employee information), real-time NHS-wide workforce surveillance has been one of NHSE's key ambitions for years.

9) **Adult Social Care Dashboard Capability** – somewhat strangely, lasting only 6 months. This dashboard integrates test data, deaths data, care home reference data, infection prevention controls questionnaire, and capacity and workforce data.

10) **Integrated Planning Tool** – rather vaguely *"to support planning between national, regional and local systems"*, which is another item on NHSE's long-term wishlist – and one where the inclusion of the phrase *"and business-as-usual activity"* raises questions of whether this

purpose is COPI-compliant, and whether NHS England has in fact let contracts it intends to persist long after the pandemic.

NHS England claims to be running a "single Front Door process" to manage applications for access (or to add data) to the Data Store, but – in stark contrast to NHS Digital – it has failed to publish a list and details of all applicants, their applications, its own consideration of those applications, plus any approved applications, and the details of the data access that was provided.

For an aggregation of so much sensitive health information to lack any meaningful transparency for so long is extraordinary. NHS England states that applications are "considered on a case-by-case basis", according to criteria which include:

- **The purpose** – "which must be for COVID-19 purposes"
- **The type and amount of data requested** – "any request for data will need to be justified e.g. if requesting record level data, you will need to explain why"
- **Transparency** – "how you are being transparent about the data requested for COVID-19"
- **The legal basis** – "the legal basis which supports the applicant to process the data"

But, even after *ten months*, NHS England appears incapable of providing any of this information to the public. (Either that, or there have been no applicants – or no approved uses...)

4) Genomics

The frequent sequencing of positive samples (of the virus not the human) has been crucial during the pandemic. It has enabled the surveillance and analysis of COVID-19 outbreaks and transmission, as well as identifying, mapping and understanding new variants of the virus.

This *pathogen* genomic work was initiated by the COVID-19 Genomics UK (**COG-UK**) consortium in April 2020; COG-UK is run by the National Institute for Health Research (NIHR), the Wellcome Sanger Institute and UK Research and Innovation (UKRI). It received further funding from DHSC's 'Testing Innovation Fund' in November 2020 and lists its funders and partners.

- **Lineage B.1.1.7** – also known as **20I/501Y.V1,** SARS-CoV-2 Variant of Concern 202012/01 (**VOC-202012/01**), and commonly as the **UK** or **Kent variant** – was detected in November 2020 by the COG-UK consortium.

- The use of genomic sequencing has significantly increased since the introduction of **surge testing** and, as of January 2021, around *45% of all SARS-CoV-2 sequences uploaded to public databases anywhere in the world* originate from COG-UK.

- Though no regular statistics are published, reports indicate that COG-UK had sequenced more than 250,000 genomes by February 2021 – with the reported percentage of PCR tests that are (also) sequenced varying over time, from 0.3% to 9%.

- The funding figures most quoted are the £20 million COG-UK received in its initial stages and the £12.2 million from DHSC in November 2020; the ongoing cost of all this genomic sequencing may, however, be significantly greater.

- COG UK and Genomics England Ltd also participate in **GenOMICC**, an international research study supported by Sepsis Research (FEAT), the Intensive Care Society, the

Wellcome Trust, and the Medical Research Council (MRC). GenOMICC is active in Australia, Canada, China, Hong Kong, Ireland, Norway, the UK, the USA, and Vietnam.

5) NHSx 'special projects'

Responsible for delivering Matt Hancock's 'Tech Vision', NHSx is neither statutory nor a body corporate; it is a joint working arrangement between DHSC and NHS England, established in April 2019 and formally launched in July 2019 and due to be subsumed into NHS England in 2021/2022.

NHSx is charged with driving forward digital transformation in the NHS – a job it has struggled to do, according to both the Public Accounts Committee's Inquiry and the National Audit Office's report on 'Digital Transformation in the NHS' in May 2020.

medConfidential first wrote about the different likely types of app in 'Apps for the next pandemic' on 3 April 2020, and followed the early development of the NHSx COVID app (amongst other initiatives) in subsequent posts on 9 April 2020, 17 April 2020 and 24 April 2020.

During the pandemic, NHSx has been responsible for a number of initiatives:

5a) NHS COVID-19 App

Sometimes known as the NHSx 'contact tracing' app, although its core novelty is in fact an attempt to determine people's exposure to other phones running the app, the development of the NHS COVID-19 app generated significant outcry in the early months of the pandemic.

By 29 April it became obvious that 'GCHQ and NHSX's contact tracing app' was going to have to incorporate the Google-Apple API, a fact which became even more abundantly clear after a lackluster pilot of the NHSx app on the Isle of Wight in May.

NHSX Covid-19 app contracts found by searching for "NHSX" AND "Contact Tracing" OR "Covid-19 App" on https://www.contractsfinder.service.gov.uk/:

- Zuhlke Engineering Ltd – **£6,774,432**
    - #3 £974,592 / #4 £1,074,240 / (*#8 £3,938,000 - withdrawn 5 May)* / #8 £4,725,600
- VMWARE UK LIMITED – **£3,025,739** award
- Go Pivotal (UK) Limited – **£1,822,210**
    - #1 £500,243 / #2 £1,321,967
- Accenture – **£869,266**
    - #13 £560,587 / #13a £308,679
- Mason Advisory Limited – **£266,400** award
- NatCen Social Research – **£235,000** award
- Eggplant – **£213,600** award
- HeleCloud Limited – **£162,768** award
- NCC Services Limited – **£132,375**
    - #6 £67,050 / #11 £76,500

- Cloudflare Inc – **£78,504** [award](#)

- Nine23 Ltd – **£51,600** [award](#)

- Britain Thinks – **£30,000** [award](#)

In May, Zühlke Engineering was commissioned to rewrite the app incorporating the GApple API and, despite [grumbles from NHSx](#) and an [attempt by Matt Hancock to blame Apple](#), the new COVID-19 app was trialled on the Isle of Wight and in parts of London during August and finally launched on [24 September 2020](#) – the app's [Ethics Board having been scrapped](#) in the meanwhile (in July).

The developers of COVID-19 app have, to give them their due, been far more open and transparent about what they were doing than many of the other pandemic initiatives. While as part of the Test and Trace programme, DHSC is still the data controller for the app, NHSx has published:

- [NHS COVID-19 app: privacy notice](#) plus [additional annexes](#);

- [NHS COVID-19 app: data protection impact assessment](#);

- [NHS COVID-19 App System documentation](#) on GitHub;

- as well as [documentation and code](#) for the 'Beta' version.

And, since February 2021 – when it was claimed that, [according to research](#), "[600,000 cases have been prevented by the app](#)" – NHSx has published **weekly statistics** [on the NHS COVID-19 app](#), which show that, e.g. alerts from the venue 'check in' functionality were [barely used](#) by Test and Trace, despite tens of millions of citizens doing as the Government requested and scanning a code at locations they visited.

The [currently definitive book](#) has 400 pages on the pandemic response; it covered the effects of the app in less than half a page. Which is about fair.

5b) NHSx 'AI Lab'

Not so much a lab as a [skunkworks](#), NHSx's "artificial intelligence laboratory" and an attending [£250 million budget](#) were first announced in [August 2019](#) – just a month after NHSx itself [officially opened for business](#).

- A [single paragraph blog post](#) on 13 March 2020 announced the AI Lab's "strategic partnership" with a controversial company with close connections to Dominic Cummings and Number 10.

  - ASI Data Science, which had rebranded itself **Faculty Science** in 2019 after their part in [Cambridge Analytica scandal](#), was originally contracted to help set up the AI Lab – [this pre-COVID contract](#) was varied to focus on COVID-19-related projects in February 2020, and was reportedly varied again so that Faculty did not receive Intellectual Property Rights in any work it did for the NHS during the pandemic.

- Another company contracted as a "strategic partner" to the NHSx AI Lab was [FutureGov](#), which seems to have been mainly responsible for helping to administer the £140 million 'AI in Health and Care Award', run by the [NHS Accelerated Access Collaborative](#) in association with the [National Institute for Health Research (NIHR)](#)... and NHSx, of course.

- For an initiative that purports to tell others how to do "AI ethics", NHSx AI Lab's partnership with "AI mercenaries" Faculty Science is not the only issue; Lord Agnew owned £90k of shares in Faculty when its second £2.3 million contract was awarded. (Whether he still owned tham at the time the third £500,000 contract was awarded is unclear.)

- Aside from **NCCID**,* the Warner brothers' extraordinary access to Number 10, and maybe developing some of the underlying models for the Data Store dashboards, it is unclear what contribution Faculty has actually made in the pandemic. Some actual, rather than promised, transparency might help...

- Notable by its absence is **Google DeepMind**, which might conceivably and usefully have, e.g. retrained its existing pneumonia model. That real AI experts (Demis Hassibis attended some SAGE meetings; Mustafa Suleyman talked sense into Number 10) appear to think AI has limited utility in the pandemic raises further questions about the use of unvalidated models to drive policy and public health decisions.

*The **National COVID-19 Chest Imaging Database** **(NCCID)** is an initiative by 20 NHS Hospital Trusts using an existing clinical system, the Image Exchange Portal (IEP), to 'warehouse' X-Ray, CT and MRI images from hospital COVID-19 patients in an Amazon Web Services S3 'bucket'. NCCID collaborators are currently running about a dozen AI research projects on the imaging data of just under 8,000 patients.

A number of pre-pandemic AI imaging initiatives, e.g NCIMI (Oxford) and 4 others, received £50 million funding back in November 2018. While some of these "centres of excellence" received 'top up' funding during the pandemic, not all of them did. NCIMI, which had for example been developing AI validation – a project that has since been handed to private contractor, Faculty Science – was passed over.

Despite the fact that Faculty's AI validation contract does not complete until August 2021, in January 2021, NHS Shared Business Services Ltd (NHS SBS) published – with very little fanfare – a massive new £360 million AI, imaging and equipment framework.

With such massive amounts of public money being pumped into "AI" initiatives, and with such a small number of favoured players involved, it is absolutely essential for patient safety and public trust that NHSx (which is not even a statutory body, much less a regulator) develops proper governance, independent oversight and full transparency.

5c) Project OASIS

A programme that could be described as "Moar dataaaa! with added MoD", **Project OASIS** was begun early in the pandemic to take copies of data collected by third party "Symptom Tracking" apps, chatbots and websites, run by academics and commercial entities ("Data Collectors"), and then to process and analyse this data on an "Intelligent Automation Platform" provided by AltViz.

- jHub – essentially, the MoD's 'venture capital' wing – processes what is indisputably (at least some) identifiable data: *"jHub will **receive and review** the data, removing **any information which may inadvertently identify users**"* and passes it on to NHS England / NHSx. There is no mention of Project OASIS in the NHS Data Store 'Reference Library'

and it is unclear what, if any, governance has been applied to millions of people's health data.

- Despite NHSx's assertion to the contrary, the Data Collectors did *not* all meet "their privacy policy and data protection obligations", as this thread from May 2020 illustrates.

- A further check in March 2021 shows that of the eight Project OASIS partners, one no longer exists, another has had its account suspended, and that several still make no mention of even the NHS or NHSx – much less jHub or the MoD, even where they list all of their other data processors in their privacy policies.

- The largest Project OASIS partner is the **Zoe Symptom Tracker** which is not named, but described by both NHSx and jHub as the "C-19 COVID Symptom Study provided by the BREATHE Health Data Research Hub for Respiratory Health, in partnership with its trusted research environment, the SAIL Databank" (in Wales). While Zoe does say it shares health information with the NHS, it also shares it with several UK and US universities, as well as universities in Italy and Sweden.

- If Project OASIS is intended to be a model or blueprint for hoovering up health data from commercial apps into the NHS, then it is fundamentally flawed – and likely unlawful outside and, to the extent the commercial partners continue to fail to comply with DPA 2018, *under* the COPI Notices.

6) JBC, PHE, ~~NIHP~~ UKHSA and other DHSC manoeuvrings

Exercise Cygnus exposed the UK's pandemic unpreparedness in 2016; the pandemic exposed just how little preparation had actually been done since then. While preparedness was a collective responsibility of the Government and the UK's institutions, the ways in which the Department of Health and Social Care and its various arm's length bodies have performed is worthy of note:

- In terms of contact tracing capacity at the start of the pandemic, SAGE minutes from 18 February 2020 stated: *"Currently **Public Health England** can cope with five new cases a week (requiring isolation of 800 contacts). Modelling suggests this capacity could be increased to 50 new cases a week (8,000 contact isolations) but this assumption needs to be stress tested with PHE operational colleagues"* and also,*"When there is sustained transmission in the UK, contact tracing will no longer be useful"* – which probably informed PHE's shift in focus from contact tracing within the population at large to looking at outbreaks in contained spaces (e.g. care homes and prisons) in March 2020.

- The **Joint Biosecurity Centre** – noting that it took months for the first version of this page to even appear – announced in May 2020, was initially set up by senior Home Office counter-terrorism official and son of Douglas Hurd, Tom Hurd, (presumably) along the lines of MI5's Joint Terrorism Analysis Centre. Hurd was replaced by "senior spy" Clare Gardiner, head of cyber resilience and strategy at the National Cyber Security Centre, less than a month later.

- **JBC** was intended to "bring together data analysis and epidemiological expertise to ensure Covid-19 outbreaks are detected and brought under control quickly". In evidence to Parliament, Dr Gardiner confirmed that JBC gets data from Test and Trace and PHE and provides "insight and analysis at a national and local level". JBC was originally supposed to

be responsible for "setting the new COVID-19 Alert level", but it now only "advises" (and there are new Alert levels).

- On 17 July 2020, the CMO told the House of Lords Science and Technology Committee that the budget for the Joint Biosecurity Centre was £9 billion – based, he said, on "this fact that we really have not invested in health protection over the last several years".

- This was followed in August 2020 by the extraordinary decision to 'scrap' Public Health England and to replace it with a new 'National Institute for Health Protection' (NIHP) that would subsume parts of PHE, the JBC, and Test and Trace. In March 2021, NIHP was then renamed the UK Health Security Authority and placed under the leadership of the then Deputy CMO.

- To undertake such a massive re-organisation of Public Health in the midst of a pandemic was seen by many as foolhardy and unduly disruptive. Questions still remain as to who will pick up the many other non-pandemic Public Health functions that PHE served; plans which DHSC has said, in its White Paper in February 2021, will be published "in due course" – covering a wide range of issues which medConfidential publicly refers to as "offline harms".

- To list every Matt Hancockup of the pandemic would be unfair; it is clear the Secretary of State for Health and Social Care is not solely responsible for the deaths of 126,000+ of our fellow citizens, and for serious lasting harms affecting many thousands more. But what is certainly true is that he has presided over serious mistakes and missteps in his area(s) of accountability, and pushed ahead with ideological policy agendas, while pursuing his 'app-happy', data- and tech-fanboy enthusiasms that have resulted in little of genuine, lasting (public) health value.

7) COVID-19 vaccination programme (England)

NHS England has overall responsibility for England's COVID-19 vaccination programme, which involves a complex set of interacting systems and players, including:

- **National Immunisation Management Service (NIMS)** – the 'System of Record' for COVID-19 vaccinations – is delivered by NHS South, Central and West CSU for NHS England; NHS Digital* supports. (DPIA now published.)
- **National Immunisation and Vaccination System (NIVS)** delivered by Arden & GEM CSU for NHS England; NHS Digital* supports.
- NHS England Regional Local Offices commission vaccination services from GP practices via the **Calculating Quality Reporting Service (CQRS)**
- **National Booking Service for COVID-19 vaccination** built by NHS Digital* on NHS.UK
- **National Surveillance System for Pharmacovigilance** run by DHSC and the Vaccines Task Force

> *Many people publicly claim credit, meanwhile NHS Digital quietly provides, e.g. all of this tech and data support for **vaccination delivery** (not forgetting 'flu as well)...*

Then, of course, there is NHS England's own COVID-19 Data Store **Immunisation & Vaccination Management Capability (I&V Capability)**, built and maintained by Palantir, which:

> *"...brings together Buyer's [i.e. NHS England's] demand models with Buyer's data sources including demographic data, patients by cohort, supply data, and vaccination events data from GPs. I&V Capability connects into supported supplier and wholesaler systems and can be used for ordering of vaccine. The I&V Capability will also be used to support the allocation of vaccines across England."*

Much as PHE chose to do with test results earlier in the pandemic, it appears that the National Immunisation Vaccination System (NIVS) – the one for health and care staff, school age children and pregnant women – uses **Excel / CSV files** for bulk upload of data to the system.

Although still under review at the time of writing, debate around the introduction and use of **COVID-Status Certification** or '**vaccine passports**' rages on. medConfidential first predicted the emergence of '**immunity certificates**' at the beginning of April 2020 in 'Apps for the next pandemic', going on to post open questions, thoughts and suggestions on 'Identity and Immunity' – as well as responding (twice) to Ada Lovelace Institute's rapid review, and highlighting what the Government was actually doing (and funding) despite Ministers' denials.

8) 'Bit players'

The pandemic has seen thousands of private companies provide products and services to the health and care sector, along with charities, voluntary organisations and public bodies. We do not go into the cronyism of some procurements in this document, as that is well covered elsewhere.

The National Audit Office has, however, conducted a number of investigations into Government's COVID-19 response – including on procurement, and Test and Trace – which some Parliamentary Committees have followed up on, notably the Public Accounts Committee and the Parliamentary Public Administration and Constitutional Affairs Committee, on data handling and data sharing during the pandemic, and longer-term issues around public trust.

8a) The big SIs

Generally, the 'Systems Integrators' of choice for Government are the 'Big Four' accountancy firms, i.e. Deloitte, Ernst & Young, KPMG and PwC – plus the 'Big Three' management consultancies, i.e. McKinsey, Boston Consulting Group and Bain. The Government was already spending nearly half a billion pounds a year on consultancy by 2019-20; during the pandemic its spending on consultants has increased by a further £175 million.

It should be said that whoever does it, e.g. setting up and staffing hundreds of drive-through and walk-through testing centres – and ensuring all the systems work and interoperate properly – involves significant cost and management. Whether such money should (or could) have been spent building capacity and resilience in Local Authorities rather than paying £1000-a-day consultants to do it, while they were put up in local hotels, remains an unanswered question. (Given the past few decades' experience of public sector IT, it is by no means certain this *could* have been achieved at the necessary speed and level of technical integration. But that is a wider issue deserving of later consideration.)

The SIs which have profited the most from pandemic data consultancy appear to be:

- **Deloitte**

  - Awarded [honorary contracts](#) by NHS England "where *direct access to data* [in the Data Store] is required to support NHS requirements";
  - Significant involvement in Test and Trace, including ["half" of Project Moonshot](#);
  - Government [faces legal action over its award of a £145 million contract](#) to Deloitte "to support testing", just as NHS Digital hands is a further [£51 million](#);
  - The number of [consultants](#) provided is equivalent to a small government department (Accessing precisely what data? Generating insights and IPR for who?)

- **McKinsey**

  - [Honorary contracts](#) with NHS England "where direct access to data is required to support NHS requirements";
  - At least 23 contracts awarded, to a value of £18.6 million – including a £4 million deal to provide "testing support" and half a million to come up with a ["vision, purpose and narrative" for Test and Trace](#)...

- **PwC, KPMG, BCG, etc.**

  - Price Waterhouse Cooper got £28 million of contracts, and KPMG got £23 million;
  - Boston Consulting Group, whose staff were reported to have been paid [£7,000 per day](#) to support T&T, got 14 contracts worth £19.7 million;
  - Quite aside from "following the science", the Government has been awarding large numbers of ['small', rapid-turnaround consultancy projects](#) throughout the pandemic – what exactly was their advice? And did it help?
  - Outside the NHS and DHSc, these companies also provided many consultants to Number 10 and Cabinet Office, as well as other Departments as part of the 'Whole of Government Response' to the pandemic.

All this after the Government (prior to the pandemic) commissioned three separate reports from members of the 'Big Seven' on "the value of NHS data". Thus far, **[Ernst & Young](#)**['s is the only one to surface](#) but even based on these crude estimates, and with post-Brexit trade deals proceeding, the sharks have begun circling...

8b) The Commissioning Support Units (CSUs)

The CSUs are looking more and more like NHS England's 'commercial wing', and a way for it to advance potentially unpopular aspects of its Long Term Plans. Consolidated down from the 25 CSUs that were established in 2013 to just five in 2021, the Units – alongside Capita, Optum and a consortium called eMBED Health which are accredited to provide CCGs with paid support.

- **Arden & GEM CSU** and senior NHS England officials had been in discussion with **Palantir** long before the pandemic; conversations [over watermelon cocktails](#) [had been going on for months](#). AGEM has delivered a number of COVID-related initiatives, including the COVID-19 Patient Notification System ([CPNS](#)) to help providers accurately record patient deaths and the National Immunisation Vaccination System ([NIVS](#)) capturing vaccination data on health and social care workers and priority groups of patients, as well as doing data modelling and some limited trials of a remote healthcare service. **AGEM CSU also administers the "single front door" application process for access to (copies of?) data in NHS England's COVID-19 Data Store** – though it makes no mention of this on its

website, nor has it published any of the applications, considerations, approvals or
disseminations of data from the Data Store throughout the entire pandemic. (Unlike NHS
Digital, which has.)

● **South, Central and West CSU** provides a COVID-19 Clinical Assessment Service for
hospitals in its area, does data collection and modelling, and for COVID-19 says it has done
"120 projects, both large and small". SCW CSU also runs the National Immunisation
Management Service (NIMS), the 'call and recall' process and System of Record for
COVID-19 vaccinations, using data provided by NHS Digital - which SCW CSU then
passes to **Palantir**.

8c) Local Authorities, Health Protection Teams, Local Resilience Forums, etc.

The transfer of public health powers and duties from the NHS to local government and PHE was
one of the most significant changes in the 2012 Act, but arrangements for local response to an
emergencywere in place long before that – as laid out in Part 1 of the Civil Contingencies Act 2004,
"Local Arrangements for Civil Protection". (Part 2 of CCA is to do with actual emergency powers,
which the Government decided not to deploy – instead it used powers under Public Health statutes
which have less Parliamentary oversight. CCA may not apply due to the existence of those other
powers, but given the flexibility the Government has shown to the law, that is not an
insurmountable barrier, especially with new primary legislation)

Arrangements for local Health Protection *pre-COVID* looked roughly like this...
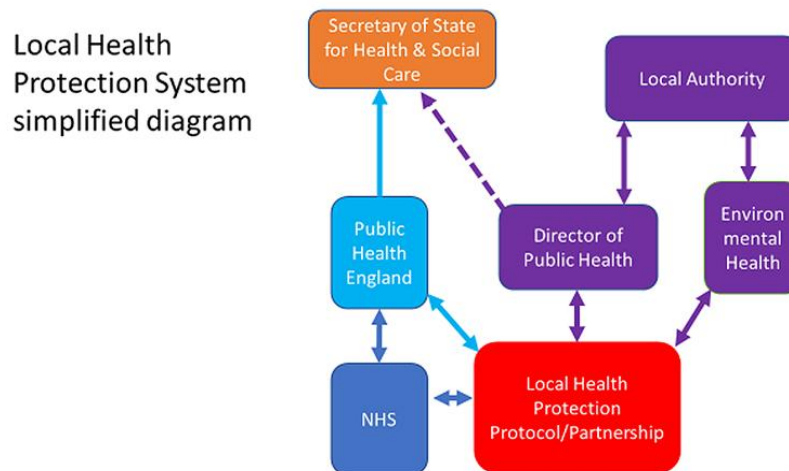
...but the introduction of two massive new bodies (Test and Trace and the £9 billion Joint
Biosecurity Centre) significantly altered the landscape. And the data flows, thus:

Key Organisational Elements
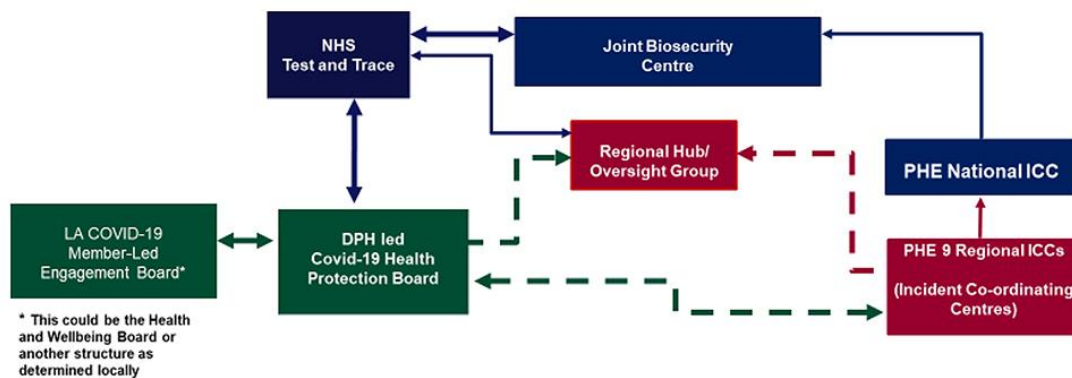
In the early stages of the pandemic – and for quite a while after that – there was clearly a desire to flow data from local systems to the centre, but no so much for the centre to pass information down to local level. And there were some notable missteps:

- Local authority environmental health officers were not even asked to help out when PHE's contact tracing capacity was overwhelmed in March 2020; PHE had just 300 staff working on contact tracing, while councils had over 5,000 EHOs. (When Test and Trace launched on 28 May 2020, it had recruited over 21,000 contact tracers.)

- Police access – specifically the local authority / police authority system(s) used for, e.g. sectioning people under the Mental Health Act

9) Social Care

As became clear in the first wave of the pandemic, social care is very much the 'poor cousin' of the Health and Social Care family. In terms of digital maturity, while privately-provided parts of the sector may have adequate information management systems, the data capability of local authority-provided care is far more variable – and in some cases during COVID has proven to be nearly non-existent. And though many of the larger private providers may *have* IT systems, it can be difficult to get them to *release* even necessary data (e.g. bed occupancy) when the providers consider such information to be 'commercially confidential'.

- Future Care Capital's Social Care Data Finder survey of publicly available social care datasets in 2020 showed that, of the authorities sampled, while one-third had some "useable social care data", another third had a "platform available but no social care data", and the remaining third had "no platform or open data" at all.

- An existing NHS Digital programme to provide care homes with NHSmail (begun in 2018, which had brought 1,000 care homes online by late 2019) was accelerated through 2020 to get a further 10,000+ care homes online, and had "13,828 sites with 36,215 NHS Mail user accounts" by 15 January 2021 – each with basic connections via NHS Mail and Microsoft Teams, some with Microsoft Office 365 integration as well. These expansions, plus other hardware initiatives, may provide for the first time the possibility of maintaining 'minimum viable' data flows across (public) social care – and beyond that, a meaningful data strategy for social care. (We also note, that in the midst of the pandemic, NHSmail also migrated to

Microsoft Hosted Office365, weeks before they would have been completely hacked. Good timing.)

- NHS Digital made a number of Adult Social Care data collections and, crucially, issued its v1.2 beta Adult Social Care collections **Data Dictionary** in December 2020; defining what data is necessary to deliver, monitor and improve care is a far more complex problem than merely applying one of the existing medical data dictionaries, like SNOMED CT.

    - Adult Social Care management system status collection (Aug 2020)
    - Scheduled **ASC collections for 2021** (Sep 2020)
    - Adult Social Care **Analytics Hub** (Microsoft Power BI)

- Meanwhile, the pre-existing Digital Social Care Pathfinders Programme 2019-21 struggled to find many more than a dozen partners – and NHSx (in a pandemic?!) has been doing case studies on AI in social care which basically replicate predictive analytics. More recently, on 26 February 2021, NHSx has issued a tender for a £10-15 million **Dynamic Purchasing System (DPS) for the provision of digital solutions** to support care planning and recording by adult social care providers.

- While the Government talks in very broad terms of 'transforming social care', it has published no meaningful strategy. Those reading between the lines may, however, infer some of the likely directions of travel:

    - **Levelling Up Our Communities** report, September 2020
        - Prime Minister's response, other responses
    - **The Care Commitment** report, February 2021
        - Article on Demos website, plus video

This summary of the data flows of COVID-19 is, of course, an evolving document. It is at best a high-level 'snapshot' from an outside perspective, looking back from a certain point in time. We fully expect further evidence to emerge, and do not assume that we have been able to identify more than a fraction of the many health data initiatives undertaken during the pandemic.

If you have any further information regarding any of the projects and programmes we cover – or on any we have missed – then please get in touch.

medConfidential, March 2021
coordinator@medconfidential.org

Other references

- Selected media coverage (Jan 2020 - present)

- '**The power of data in a pandemic**' (NHSE/X, 28 Mar 2020)

- **Timeline:** UK government, coronavirus and data (Mar - Jul 2020)

- [How not to use tech in a pandemic - a **timeline** of the UK's response](#) (Mar - Jun 2020)

- All The Citizens: **[COVID-19 public sector contracts](#)** (Mar 2020 - present)

- [Faculty Science **contracts**](#) across the whole of Government (Mar 2020 - present)

- '**[At the bottom of the valley between two peaks](#)**' (medConfidential, 28 June 2020)

- medConfidential [paper to Ming Tang](#) on upcoming 'NHS Bill' (Dec 2020)

- '**[Data integration – driving improvements in patient care](#)**' (NHSE/X, 18 Dec 2020)