Preserving patient confidentiality - technical, cultural & political issues

For BCS Herts meeting, Nov 17th 2016

Ian Herbert, FBCS
Director, S I Herbert & Associates Ltd
Committee Member, BCS Primary Healthcare Specialist Group
stuartianherbert@gmail.com

References to FSfA. e.g. FSfA 7.5 refer to sections of "Fair Shares for All" published by the PHCSG in 2012

My scope today

I'm concerned with the sharing of the vast amount of **patient identifiable data (PID)** that the NHS holds. Most of this is collected by clinicians or their support staff during and for the provision of NHS care.

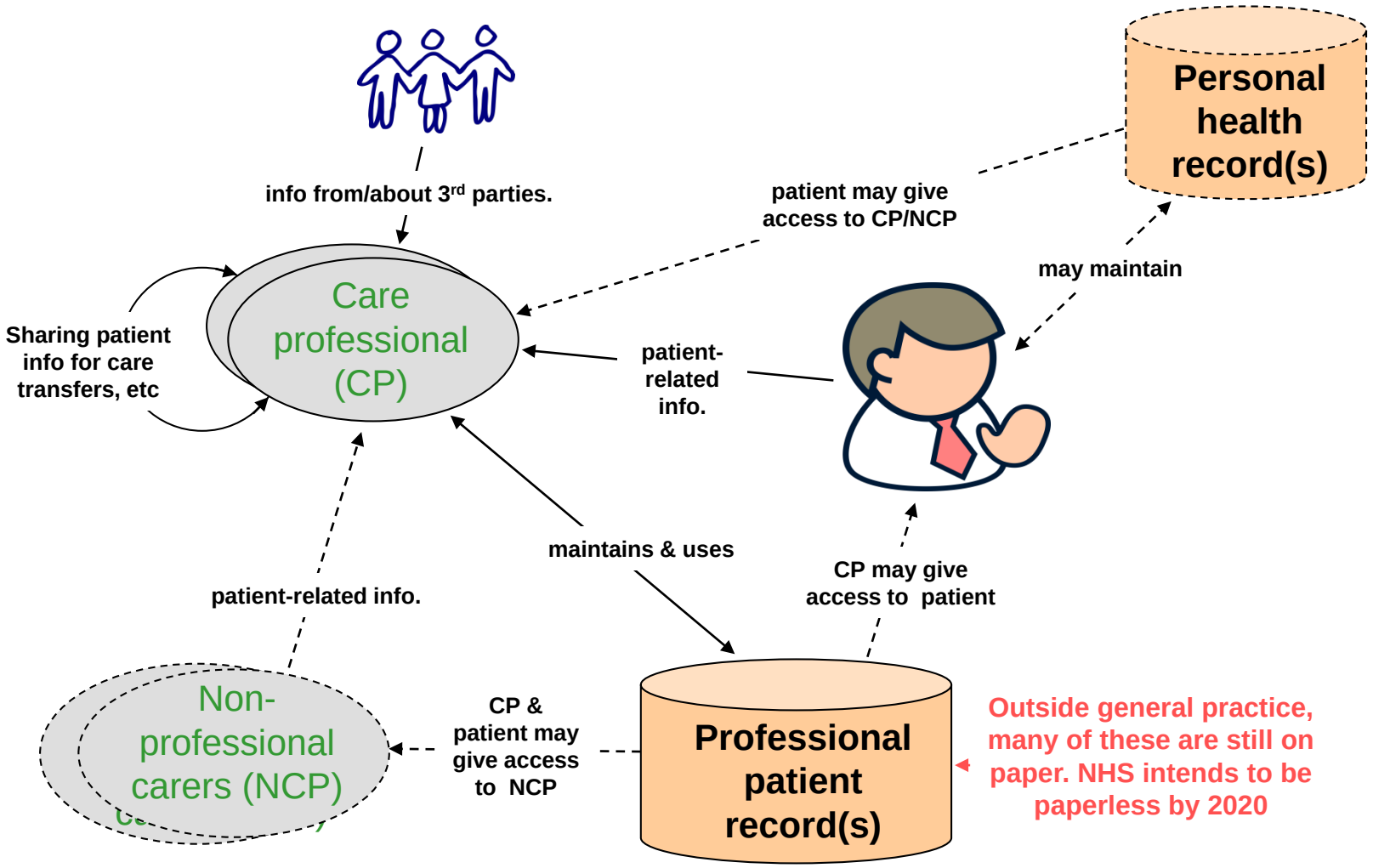
A person's data is identifiable if it:

1. identifies the person, e.g. contains NHS number or name & address, and/or
2. can do so when combined with other information in the possession of, or likely to come into the possession of, the person holding it. Just removing identifier(s), or replacing them with pseudonym(s) is generally not enough to de-identify rich patient data.

Clinicians & others working with patient data have a common law duty of confidentiality towards their patients. This limits their sharing of PID, and is the **major bulwark** protecting patient privacy. The processing of identifiable data is also governed by the Data Processing Act 1998 (e.g. fair processing reqs, & the right to object to processing) and implicitly by Human Rights Act 1998.

Non identifiable patient data may be freely shared for any purpose. It should not be re-identified without authority, but this is not (yet) a criminal offence □

Identifiable data collected for care



Sharing identifiable patient data for care

1. Clinicians should share data with other professionals caring for the patient - I would add “relevant” before “data”. Seeking care implies patient consent to sharing data for personal care .
2. But a clinician should respect patient objections to sharing particular data for care with another care professional / carer / relative, unless he or she considers:
   a. It’s essential for the provision of safe care (which must be explained to the patient),
   b. its concealment would seriously harm someone else, or is justified in the public interest. Where possible, the patient should be informed.
3. Clinicians / organisations may use patient data collected during their provision of care for local clinical audit & management, including sharing it with those also bound by a duty of confidentiality
4. There are other situations where a clinician may / must share patient data with others not caring for the patient, e.g. detection of gunshot wounds, a court order
5. A clinician should seek patient consent before sharing their data with others &/or for purposes (e.g. for insurance reports or benefit claims) that don’t have a statutory basis □

So far, so good …

What we have shown so far marries ethics & the law nicely together. Clinicians should (& almost all do) use PID **as their patients would expect**. The General Medical Council gives comprehensive guidance on how clinicians should use PID.

For secondary purposes (other than local clinical audit & management), a clinician caring for a patient should only share a patient's identifiable data:

a. with the consent of the patient (or a legal proxy),
b. if statute requires or permits it (see DPA 1998 schedule 3))
c. where there is a legal basis to do so (e.g. a court order)
d. where the public interest overrides the duty of confidentiality to the patient & the public interest in a confidential health service

Route b & c require a means of overriding the common law if not already included □

Use for secondary purposes

1. Problems start when patient identifiable data (PID) is sought for secondary (2ndry) purposes & it's not possible / considered too onerous to obtain consent.
2. Common reasons given for needing PID are to:
   a. identify patients so that they can be asked for consent (rather than asking clinicians to do it)
   b. enable linkage of data from different sources, such as different care providers, Registrar of Births and & Deaths. etc. NHS number is popular for linking data from NHS sources. Name, address / postcode &/or dob are usually only needed to link NHS data to external data
   c. Obtain partial identifiers, e.g. gender, date of birth &/or postcode, when they are themselves research parameters
3. Demand for data for 2ndry uses is booming, the digital data mountain is growing exponentially, electronic access is becoming much easier & the hype about big data is intense. HMG has caught the fever & expects significant benefits from 2ndry use of NHS patient data (qv PM's speech of Dec 5 2011, FSfA 7.5).
4. The demand for easier access to PID started c.2000, when the DPA 1998 started to bite, & researchers were in the van.
5. Here's what happened next. Blue date = action improving data protection, light blue = documents suggesting improvements. Red and pink show the converse, and black is for neutral items. The colouring reflects my opinions □

History of patient info. governance - 1

1998	**Data Protection Act (DPA 1998) comes into force**.
2002	**NHS Act 2002** legalises Cancer Registries' collection of identifiable patient data & spawns Patient Information Advisory Group (PIAG) to vet applications to use identifiable patient data for secondary purposes & approve them if sufficiently in the public interest.
2006	Academy of Medical Sciences *Personal data for public good: using health information in medical research*
2008/7	Cabinet Office commissions *Data Sharing Review* from the ICO and Mark Walport, Director of the Wellcome Trust. Moots idea of SoS-issued data sharing orders for 2ndry uses of PID.
2008/10	NHS Act 2008 spawns **National Information Governance Board** to maintain NHS policy on use of identifiable data & advise on unconsented applications for 2ndry uses. PIAG becomes NIGB Ethics & Confidentiality Committee (ECC)
2009/6	Wellcome Trust publish *Towards consensus for best practice Use of patient records from General Practice for Research*
2009/9	**Coroners & Justice Bill** proposes to implement *Data Sharing Review*, but relevant clauses withdrawn after vigorous opposition from the BCS, Privacy International and the BMA
2011/1	Medical Research Council publish *UK e-health records research capacity and capability*
2011/1	Academy of Medical Sciences publish *A new pathway for the regulation and governance of health research* urging adoption of the recommendations of the 2008 Data Sharing Review

History of patient info. Governance - 2

2011/1	**NHS Bill 2011** to abolish NIGB & give NHSE and SoS power to issue Directions to extract identifiable patient data irrespective of common law duty of confidentiality. Flouts promises given to patients in preceding White Papers – "nothing about us without us", etc, etc..
2012/3	**NHS Health & SC Act 2012**. Very little changed from Bill, & IG elements of it badly worded
2012/5	GPES Independent Advisory Group founded to vet applications for GP patient data.
2012/5	2nd Caldicott Review started
2012/12	NHSE plan to collect identifiable patient data regularly from GPs – "care.data"
2013/1	Not-for –profit action group Medconfidential founded in response to care.data initiative
2013/3	2nd Caldicott Review published
2013/3	ECC becomes HRA Confidentiality Advisory Group (CAG)
2013/4	**SoS offers patient data sharing opt outs**
2013/9	HMG response to Caldicott 2 published
2014/5	**Care Act 2014** says HSCIC can only disseminate data for health & social care or health promotion(?)
2014/6	Partridge Report finds significant IG failings in data sharing by HSCIC (now NHS Digital)
2014/10	National Data Guardian (NDG - Fiona Caldicott) appointed
2015/5	GPES Independent Advisory Group abolished
2016/6	NDG's report Security, Consent & Opt outs published: Medconfidential reply .
2016/9	Consultation on Security, Consent & Opt outs ends. It started on 20160707. PHCSG responds □

History of patient info. governance - 3

"Fair Shares for All" describes the story up to & including the the HaSC Bill 2011, & gives references for the cited documents. Highlights of what followed are shown below.

HaSC Act 2012 deliberately fragmented governance of NHS patient data. It is now the concern of:

a. Health Research Authority (HRA) - managed by research interests

b. HRA CAG - assesses proposals to use PID for 2ndry purposes without consent (**other than by a HaSC Act 2012 Direction**). Has significant representation of patients & patient data controllers

c. CQC - appreciable user of PID. Its National Information Governance Committee was abolished in 2015

d. NHS Digital - responsible for the IG Code of Practice & the largest warehouse of NHS PID. Its Data Access Advisory Group considers requests for non-identifiable sensitive patient data that NHS Digital holds.

e. National Data Guardian - whose statutory position is still unclear

f. And the (largely hidden) hand of HMG / DoH.

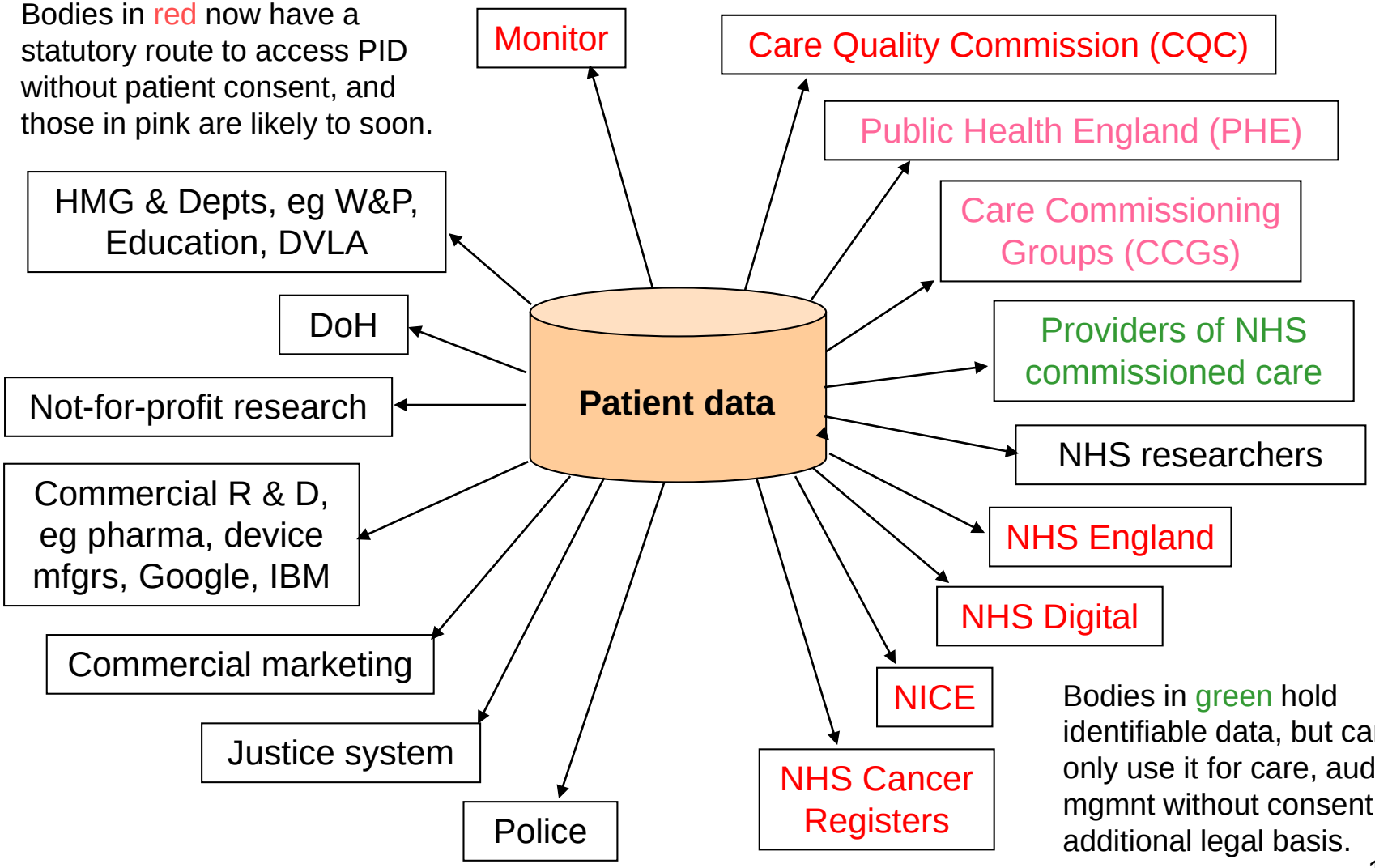
The HaSC Act also dishonours undertakings given in the NHS Care Record Guarantee, and (probably) the right to a private life enshrined in the Human Rights Act 1998.

SoS/DoH has taken over 3 years to define & implement the SoS opt-outs, and then only under pressure from GPs, BMA & Medconfidential, whose draft legislation to do so was rejected by HMG.

A 'son of care.data' – standard SCCI 2090-2222 - plans to collect identifiable GP patient data from 2017 onwards to assess the impact of GP Healthchecks on patient health □

The end result

Bodies in red now have a statutory route to access PID without patient consent, and those in pink are likely to soon.



Bodies in green hold identifiable data, but can only use it for care, audit & mgmnt without consent or an additional legal basis.

Back to basics - the “least” principal

All users of identifiable patient data should abide by the ‘least’ principle, as the DPA 1998 suggests. They should:

1. use the least data (fewest data subjects & the least data per subject)
2. use it in the least identifiable form
3. hold the data for the shortest period of time
4. restrict access to the least number of authorised people
5. copy it the least number of times (ideally none)

necessary for their purposes.

The ideal from the IG point of view would be to collect the relevant data from its original sources as and when it is required, where all sources are always available on line, or better still use it in situ. It is not often feasible yet □

What would patients like?

Patient wishes and concerns about the use of their identifiable data for secondary purposes are well researched, & FSfA 4 summarises the results of 17 studies. Most patients are unaware of how widely their identifiable and de-identified data is used and by whom.

In general patients are happy to contribute data for worthy secondary uses. They are less happy to share it with commercial companies. Patients regard seeking their consent as a mark of respect, On balance, clinicians prefer that their patient data for 2ndry purposes is used with patient consent, or if not, with the consent of a body such as CAG. 30% of patients are happy with the latter too.

But where rich identifiable data is being warehoused for uses and users that are largely unknown, an opt in is generally accepted as required (and more acceptable to data subjects) than an opt out (altho' there are concerns about how meaningful consent is in such circumstances). The BMA agrees, and large projects such as Biobank and the Million Women study operate successfully on this basis. Researchers fear that opting in will lead to significantly biased data, but have not presented significant evidence of this. Few 2ndry purposes need total population data .

Patient wishes have largely been ignored by researchers and DoH / HMG - until the balloon goes up (as it did with care.data in 2013). That is why a much larger % of English patients than Scots opted out of their respective Summary Care Record, and 1.2 million English patients have registered objections to sharing their identifiable data for other purposes as a result of the care.data fiasco □

Data warehouses

Warehouses holding petabytes of data are feasible & offer many advantages – collect, clean and link data once for many purposes, and add to them incrementally, potentially for ever. That most GP records are now in silos run by their system suppliers makes them easier to build (although GPs are still the data controllers). Centrally linking and cleansing data will reduce (but not abolish) the need for providing identifiable data to end users. But linking data implies collecting identifiable data, and retaining it if further linking is likely. Pseudonymised data may well provide linkable data for 'one off' collections, but is not suited for incremental collection over time from large numbers of sources (as was proposed by care.data).

Warehouses enable analyses that we could only dream of 10 years ago, and will be attractive to a wide variety of would-be users. But they break "least" principles 1,2 and 3.

To maintain public trust and to have any hope of being considered a 'safe haven', warehouse collection & dissemination of data and its security must be governed by a body with a majority of patients, patient data controllers, IG experts and ethicists. NHS Digital has gone some way down this road for data it disseminates, but not for the PID it collects (q.v. the abolition of GPES IAG).

The temptation to use legislation to make population of &/or access to warehoused data easier is great, vide the Directions enabled by the HaSC Act 2012. **Al**l proposals to collect data by a warehouse (including NHS Digital) should be subject to approval by the CAG as well as the governance mechanism outlined in the previous paragraph □

DPA 1998 in the Brave New World

The DPA 1998 always had problems, such as overly generic descriptions of who may use personal data that is collected and what for, but data warehousing raises new & major issues.

Patient consent to share and the right to object to processing become meaningless where data is collected & warehoused for future uses & users unknown at the time collection is proposed & consent sought, **unless** the warehouse contacts every data subject each time a new use / user arises. This is usually impossible or too expensive to do, which is one of the major arguments for seeking consent (i.e. opting in) rather than assuming consent and permitting an opt out. Another approach is to collect more generalised consents in advance of collection (e.g. by major class of user and/or use).

So the definition of what is meant by 'fair processing', and some DPA principles need to be looked at afresh if they are to remain useful

The implementation of the new EU General Data Protection Regulation (GDPR) in the UK in 2018 will also make this necessary, as it is more insistent that data subjects get timely and more explicit fair processing notices, consent for 2ndry purposes must be more explicit, and that assumed consent is less acceptable than hitherto.

People have also asked that re-identification of de-identified data be made a crime, unless specifically & openly allowed for when it is de-identified, and this is being considered. □

A different approach

Currently we have a situation where most secondary uses involve **giving the end user a copy of the data**. With the increase in storage availability, networking speed and computing power, this isn't necessary, as the recent experiments with a data laboratory by HSCIC and various other data accumulators (SAIL in Wales, SHIP in Scotland, THIN) indicate. The end user sees his results, and can even use the most confidential patient data as he / she **never sees the individual patient data.** This makes it even more likely that the public would be content for new users / purposes of warehoused data to be vetted and approved (or not) on their behalf by a body such as CAG.

Greater use of de-identified data (including data which only uses pseudonyms generated at source to identify data subjects) may be sufficient to satisfy some end user demands. However much health data being collected is so rich that it would be identifiable in conjunction with other data the holder had, or could have access to. For example, the HSCIC holds so much identifiable data that almost any additional individual level data it collects would have to be considered identifiable.

I am as eager as any that the maximum use is made of NHS patient data for worthy 2ndry purposes, but not at the expense of patient data confidentiality. The evidence shows that GPs & the majority of the public & patients agree, as I believe do all clinicians caring for patients. As I hope this presentation shows, there are ways this can (and should) be done that should be explored ASAP.