



BCS Primary Health Care SG

CLICSIG 5/7/14 - Can Care.data be made compatible with patient confidentiality?

Palace Hotel, Buxton

Discussions held under the Chatham House Rule

CLICSIG Objectives

The of 19 members of the BCS PHCSG was held on the 5th July 2014 to consider the questions surrounding the NHS England Care.data programme and its implications for patient confidence and confidentiality. The intended outcomes were a report for BCS Health and for the House of Commons Health Select Committee.¹

Care.data² be will the largest and most comprehensive database of identifiable digital patient data in the UK, if not the world. Done well, it will in due course offer very significant new and enhanced opportunities for all sorts of secondary uses. Done badly it will severely undermine the trust between patients and their professional carers which is the bedrock of health and social care, and tarnish the reputation of big data in healthcare.

The Care.data programme was announced in 2012, The Technical Specification for the GP element was published in May 2013 following the agreement to proceed from the March meeting of the GPES IAG. The Direction³ from NHS England to HSCIC was formally published on 19th December 2013. There was a leaflet campaign to inform the public in January/February (the leaflet was generally considered to be confusing and the distribution method led to many households either not receiving it or not noticing it) and the media and public concern led to a 6 month delay soon afterwards.

This CLICSIG meeting was called to consider what needed to be done if public and professional confidence in the Care.data programme was to be restored and maintained. The high level conclusions covered:-

- Transparency - the information on the aims, data collection, governance and distribution need to be specified, published and preferably subjected to public consultation
- Details of the programme; accountability, governance (including independent external review), consent/opt out mechanisms, planned storage, QA and distribution of the outputs, specified and published.
- Patient Information Campaign designed to reach all of the population involved.

The meeting discussed the problems presented in three broad areas:

- The collection of Care.data - the programme, the risks and benefits of this approach including the ethical and legal aspects and implications, future plans where known and data limitations
- Handling of the data collected. Including its linkage with data from other sources
- Dissemination from the HSCIC

The *Care.data programme* extends beyond the collection of data from GP records under the powers conferred on NHS England by the HSCA, even though it is this aspect which has gained wide-spread media coverage. It will include the 15 or 16 data collections already held by the HSCIC, and further data collections will be added in the

¹ Full recommendations Annex 2

² A concise description of care.data is given in Annex 1.

³ <http://www.england.nhs.uk/wp-content/uploads/2014/01/cd-directions.pdf>

future⁴ including pathology data transmitted to GPs via DTP⁵ and later prescriptions transmitted via EPS (EPS has the facility to capture data on what has been dispensed but this has not yet been activated).

Patients are allowed to opt out of allowing the upload of their identifiable GP data for secondary uses (including care.data), release of their identifiable data from the HSCIC or both: NHS England arranged for an information leaflet to be distributed to households in England. However, the information contained in it was confusing and the leaflet - delivered via the Royal Mail leaflet distribution service - failed to reach many households⁶.

The first output is planned to be CES (Care Episode Statistics) produced by linking GP and HES data to provide a view of the care afforded to individual patients for commissioning purposes⁷ but the linked databases will be extended in the future using as yet undecided new and existing data collections.

There are *advantages in a central data repository model* including cleansing and quality assurance of the data being performed once rather than repeatedly for different queries.

On the other hand, the data in such a rich database is at *risk of re-identification*.

There is a *risk of 'legal abuse'* - decisions to change the legal framework protecting care.data's collection, use or distribution after the data has been collected⁸ and after fair processing information has been provided to NHS patients.

This *situation* - a persistent cumulative database with no defined purpose(s) before collection - *is not addressed in the DPA*, and although the HSCA 2012 specifically protects GPs from actions due to breach of confidentiality, the ethical obligation of confidentiality and the requirement for 'fair processing' (Data Controllers have to inform Data Subjects of changes in the use of their personal data) remain.⁹

The original *specification for the GP care.data extract* included data from April 2013 and limited the data use to commissioning purposes: currently NHS England has instructed the HSCIC to implement a rolling programme of collecting data monthly for the previous 4 months on a rolling basis i.e. only the previous 4 months will be included. It is not clear how this would be aligned with a cumulative permanent record.¹⁰

Before the programme is rolled out, NHS England has committed to assessing the programme with a small number of Pathfinder practices, probably 100-500.

The methods of selecting these practices, developing or specifying what is being trialled, evaluation of outcomes and incorporation of the findings into a National rollout are not yet known, although one participant had been told that several CCGs had expressed interest and the CCGs/Pathfinders would be financed.

NHS England has said that certain conditions need to be in place before the programme can proceed but it expects to proceed in October - which is before the HSCIC (Kingsley Manning) expect its Pseudonymisation working group to report (November 2014): it has never been clear whether the NHS England plans have taken the HSCIC plans and organisation into account: as *Joint Data Controllers* clarification of their respective roles and accountability is needed if public trust is to be gained/maintained.

Questions were raised about *data quality and comparability*.

When it comes to the intended first output - CES - concerns were expressed about the robustness of combining GP data of very varying completeness and quality¹¹, entered into the record during consultation and for the purpose of providing personal care, with HES data, derived from patient records by non-clinical coders for payment purposes. There has been little or no investment in *training* for GPs or other primary care clinicians in record keeping or optimal use of Read or CTV3 Codes and if the quality of the care.data data is to be dependable, it was felt that this was an area which needed to be addressed urgently at the undergraduate, specialist training and postgraduate levels.

⁴ NHS England Privacy Impact Assessment <http://www.england.nhs.uk/wp-content/uploads/2014/04/cd-pia.pdf> states intention to ask for historical data in future

⁵ This is being used by Leeds University in a data quality assessment project under s251

⁶ Relevant polls:

ICM Research, Feb 2014 "29% recall getting leaflet": <http://www.bbc.co.uk/news/health-26187980#?>

JRRT/ Ipsos Mori, May 2014 "51% never heard of care.data": Q4 <http://www.ipsos-mori.com/Assets/Docs/Polls/jrrt-privacy-topline-nhs-2014.pdf>

⁷ <http://www.england.nhs.uk/wp-content/uploads/2013/08/cd-ces-tech-spec.pdf> technical specification of GP data extraction.

⁸ The directive - <http://www.england.nhs.uk/wp-content/uploads/2014/01/cd-directions.pdf> p9 states will be renewed and updated when necessary.

⁹ It was noted that GPs have 'fair processing' obligations to inform their patients about care.data: hard especially when they themselves have little information about uses.

¹⁰ It has since (September 2014) been made clear that each month's collection will contain the previous four months data, the oldest three months replacing the latest three months of data already collected. This is to allow for data that reached the record well after it was generated, often by external bodies such as pathology laboratories. The date attached to a data item is when it was stated or generated, not when it was entered into the GP record, this is problematic and needs clarification..

¹¹ Unpublished data from PRIMIS+

Management of the data once collected. by HSCIC

Process

NHS England has directed HSCIC to collect specified identifiable data from GP practices and limited the distribution of the data derived from this, but the *Directions* state that it *will need to be renewed and/or updated* when new collections or uses are required, and have stated that the current GP care.dataset is only the initial requirement: more data will be requested in the future.

Once data is collected in identifiable form, it is linked to data in other datasets in the HSCIC, and then pseudonymised. However, the means to re-identify it will be retained by the HSCIC.

Following this, and depending on the nature and purpose of the linked dataset, data releases will be graded as:

- Green - aggregate, fully anonymised data fit for publication
- Amber - pseudonymised/de-identified data with some risk of re-identification - which would only be safe in a controlled environment¹² or
- Red where the data is either fully identifiable or the risk is so high that it is effectively fully identifiable: this type of data may not - at present - be released unless it is legal, e.g. s251, and then only under restricted conditions.

Storage

The Care.data Programme will accumulate - and retain - increasing amounts of data over time. In addition, linkage between different datasets will produce new datasets e.g. CES, which in turn will require storage and management.

Questions were raised about physical storage and jurisdiction, e.g. will the data be held in servers and by companies under UK or EU law, or in situations where it might be accessible to non-EU legal systems e.g. the US Patriot Act.

At present the data is held on physical servers owned and controlled by HSCIC: Cloud storage is not currently under consideration.

It was suggested that any contracts should be written to enforce this and the example of the MOD was cited: MOD medical record storage in mainland Britain was specified in the contracts - avoiding the possibility of future arguments with suppliers.

Data processing and access.

Once data is held by the HSCIC, it will undergo further processing (including de-identification) and the outputs will be linked with other datasets.

Access to both the raw data and linked datasets will be limited by organisational (limited and controlled access to the information) and legal means (contracts, including re-use agreements v.i.)

It was emphasised that all the data releases from the NHS IC - predecessor to the HSCIC - were legal: however, the HSCIC has brought outsourced collections in-house (last due to be completed in December 2014) and has reviewed all its existing contracts with a view to converting them to a standard contract now or on renewal.

Access

Previously the dataset specified in the application was passed - under legal contract - to the requesting organisation and queries run there. This led to the problems reported in the review by Sir Nicholas Partridge of the contracts awarded by the NHS IC and his recommendations for future contract management.¹³

The suggestion of retaining the data in the HSCIC and running queries against it in situ is under discussion and may be implemented in a manner similar to that used in QResearch and the NSO¹⁴. This would have the advantage - besides reassuring the public - of preventing unauthorised access and allowing the queries to be refined to provide useful results. (see ref 13)

The HSCIC has data analysts transferred from CfH who are currently not employed in data analysis, so the capacity already exists.

Safe Havens and ASHs (Accredited Safe Havens)

The HSCIC is a statutory 'safe haven' - but the precise definition is not laid down in statute.

¹² NHS England appear to have accepted that "Amber" won't fly for CES. The position with regard to HES and other existing releases is unclear, and. public information explicitly mentions RAG., e.g. <http://www.england.nhs.uk/ourwork/tsd/care-data/our-plans/> ,

¹³ See Partridge's 9 recommendations: http://www.hscic.gov.uk/media/14244/Sir-Nick-Partridges-summary-of-the-review/pdf/Sir_Nick_Partridge_%27s_summary_of_the_review.pdf

¹⁴ The secure haven model appears to be the plan for GPES care.data extracts during the Pathfinder phase

The Department of Health (DH) ran a consultation "Protecting Health and Care Information A consultation on proposals to introduce new Regulations"^{15 16}. This was a high level consultation which appeared to assume that multiple ASHs would be established but gave little detail on their structure, powers, accountability, legal status or ability to hold PID for unspecified purposes. The relationship to the only statutory safe haven - the HSCIC - was unclear in the consultation. Their establishment would be authorised by the Secretary of State: no mention was made of the possibility that they might be terminated or any provision for data transfer or deletion.

There are data analytical functions which would be best managed locally (and might be impossible centrally, e.g. management and QA of contracts awarded by CCGs). It was not clear what role an ASH could perform in this or how ASHs would fit into data protection legislation and the role of the ICO: it was felt by some that DH, NHS England and the HSCIC have ignored the ICO's concerns and input.

If ASHs are established, the security, governance, auditing and accountability need to be at least as good as that of the HSCIC.

Data recipients

The Care Act 2014 restricts the use of the data to "health purposes"¹⁷ without clearly defining such purposes: the GPES IAG was only asked to approve the GP care.data extract 'for commissioning purposes'. In the Technical Specifications, it is stated that this is release 1 and further information will be requested in the future without specific information that this will or might include wider distribution.

There was discussion about the nature of recipients and whether the regulations needed for different types of organisation differ. Academic research organisations are accustomed to working with ethical standards supervised by Ethical Committees and each research study is pre-planned: commercial organisations operate in a legal contract environment and it is not clear whether they can or would be able to specify their data requirements to the same extent - or be bound by the same ethical considerations: it is unclear how Information Intermediaries operate: different methods might apply to different types of institution, but this might be difficult to implement and be open to challenge.

Purposes of data release

Aggregated datasets can be - and are - publicly released and posted on the HSCIC website. Our discussion related to pseudonymised and potentially re-identifiable data. The richer the dataset the greater the risk of re-identification and the CES dataset will be very rich indeed.

At present the use of the GP data extract from general practice has only been approved for commissioning purposes by the GPES IAG¹⁸: it is expected that the IAG will be asked to approve extensions to the use and possibly content in the near future¹⁹.

Data is currently made available by HSCIC for a wide variety of organisations and purposes, some authorised by s251, and HSCIC is doing a lot of work reviewing recipients, purposes and agreements, in light of the change in mandate from the NHS IC under the HSCA 2012 and the restriction to use for healthcare purposes and the promotion of health under the Care Act: healthcare purposes, neither of which is defined in the Act.

We considered the situation in a research database, QResearch.

In this the data is held in secure servers in Nottingham University. Research queries are submitted for specified research projects by researchers who are vouched for by responsible academic legal entities (usually Universities), the data needed for the project is specified, the data is manipulated in a secure lab at Nottingham or in an approved equivalent hub in another university, and only the output - not the raw data - is released.²⁰

A condition of use is that the results of the study must be published in an academic journal.

It was felt that a similar security/governance/access model would be suitable for the Care.data Programme - and within the capacity of HSCIC, with the great advantage that unauthorised extensions of use or distribution of the entire dataset would become unlikely if not impossible.²¹

¹⁵ June 26th to August 8th 2014 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/323967/Consultation_document.pdf

¹⁶ <https://medconfidential.org/2014/the-department-of-health-protecting-personal-health-and-care-data/>

¹⁷ <http://www.legislation.gov.uk/ukpga/2014/23/section/122/enacted>

¹⁸ Discussion in http://www.hscic.gov.uk/media/14911/GPES-IAG-Minutes-for-10-July-2014/pdf/GPES_IAG_Minutes_10.07.14.pdf

¹⁹ IAG meeting 11.9.14. The proposal was for care.data to support extra purposes by additional undefined bodies <http://www.hscic.gov.uk/article/3525/Caredata>

²⁰ <http://www.qresearch.org/SitePages/Home.aspx>

²¹ HSCIC is very actively investigating such an option. The HSCIC would provide access to care.data via an internal Data Laboratory. It has stated that this will be the only way of accessing data from the care.data pathfinders.

Problems - some expected to be temporary

The abolition of PCTs (NHS bodies) and creation of CSUs (which are not defined in statute, are not legal NHS bodies and may be private companies in the near future) combined with the output from the Caldicott 2 review requiring commissioners to use de-identified data for commissioning has led to the approval of a short term s251 order allowing the use of identifiable patient data by named CSUs and CCGs pending development of means to remove the need for this.

Some research projects run by University departments e.g. Imperial College and Leeds have access to identifiable patient data under s251 approval. It is likely that the need for s251 would decrease if the data-lab model was introduced and applied to all of the data.

We concluded that it was possible to develop and provide secure access for research in a data lab environment, and there did not appear to be circumstances in which it would be necessary to release the entire data set to an outside environment, especially if data is only released for specific purposes and access is restricted to the data needed for those purposes.

Research establishments subscribing to and supporting the FARR institute have standards equal to those of the HSCIC and it was felt that all organisations receiving data requiring to be handled in a secure environment should have and be able to demonstrate the same safe environment and be subject to the same checks and audit as HSCIC.

Other issues

There were a number of other issues which might or will affect the management of data included in care.data

Status of data re. ICO and data protection

While there are differing views on the 'identifiability' of pseudonymised data²², according to the Data Protection Act 1998 s. 1(1) data must be regarded as personal data if it intrinsically identifies an individual (e.g. by their NHS Number) or a very small number of people (e.g. most full postcodes) or if a person could be identified in conjunction with data which the holder of the data also has or could readily have access to.

It was the view of this group that any linked, patient-level dataset as rich as CES would be re-identifiable under the provisions of the DPA : in the unlikely event that CES/care.data are *not* be regarded as personal data, then DPA might not apply.

Status of organisations requesting access to care.data

Commercial organisations are competitive and there is a possibility that they would take legal action if they perceived that there was not a 'level playing field' in giving - or refusing - access to the data held by HSCIC. It may be difficult to define the purposes for which the data may be released or accessed to avoid this unless included in secondary legislation, and so far there has been no political appetite for this.

It is not clear what, if any, effect adopting the proposed Transatlantic Trade & Investment Programme (TTIP) might have, but there are fears that it could severely damage the protection afforded by UK legislation. Concern was also expressed as to whether a "one strike and you're out" policy would be able to be applied, legally, in case of data and contract breaches: it was felt that this might require legislation.

IG governance within the system

There is known to be a review of the whole information governance structure of the HSCIC in progress: however there is no publicly available detailed information on its TOR or scope. The members of the IAG - set up to provide external governance on data extracts from general practice records using GPES - have been asked to extend their terms to November 2014²³ and it is understood that the role of the IAG - or its replacement - will change as a result of the governance review.

Both NHS England²⁴ and HSCIC²⁵ have published PIAs (Privacy Impact Assessments) of the care.data programme: it is not clear whether the structures outlined in these will survive the internal governance review at HSCIC.

The role of DAAG²⁶ is being strengthened and independent oversight of the whole of the IG arrangements may be provided by IIGOP and CAG but no details of the powers, TOR and extent of scrutiny of collection, management, safeguards and scrutiny of requests for data are available so far.

Concerns were expressed about the under-resourcing of the CAG - which already scrutinises many s251 requests - and whether it has the capacity to absorb the additional workload.

²² http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

²³ Now extended to the end of the financial year 2014/15

²⁴ <http://www.england.nhs.uk/wp-content/uploads/2014/04/cd-pia.pdf>

²⁵ http://www.hscic.gov.uk/media/12931/Privacy-ImpactAssessment/pdf/privacy_impact_assessment_2013.pdf

²⁶ DAAG is responsible for authorising the release of identifiable information: CAG approves collection and use of identifiable information under s251

Role of NHS England

Although there is some information available on the HSCIC, there does not appear to be any available for NHS England.

NHS England and HSCIC are deemed to be Joint Data Controllers by the ICO, which should require written agreements on their roles²⁷ but so far these have not been published.

The situation is complicated by the power of NHS England (under the HSCA) to direct HSCIC to collect any information it deems 'necessary or expedient': it is not clear whether this power extends to directing the use or distribution of such information.

The role of the NIB - National Information Board - which sits within the Department of Health is unclear.²⁸

Consent

The current model of patient consent is an 'opt out' one: patient data will be collected unless the individual patient registers an objection to their identifiable data leaving the practice for secondary purposes or from leaving the HSCA or both.

This ability to be able to opt out is not included in legislation (it is under the NHS Constitution and announced by the Secretary of State for Health) and could therefore be revoked.

Opting out after the data has been collected would not delete information previously collected.

Under the HSCA, practices have no power to refuse to allow the extraction of patient data even though some have threatened to do so: under the DPA, GPs have a duty ("fair processing") to inform all their patients about the changes being made to the use of their data. Once data has left the practice, NHS England and HSCIC become Joint Data Controllers: it is not clear what obligations they have towards the patients as data subjects. While they do not have a duty to act in the patient's best interests, the Care Act 2014 amends s253 of HSCA 2012 to say that the HSCIC must have regard to "the need to respect and promote the privacy of recipients of health services and of adult social care in England".

The BMA at the 2014 ARM adopted the policy that consent should be opt in and not opt out.

Public information.

In order for consent to be informed and the collection of care.data to be legal, information about care.data and the opt outs has to be available and known to be available to all patients.

Information for both public and professionals is generally agreed to be lacking or absent, leading to two "pauses" - in August 2013 and April 2014.²⁹ The first pause was intended to allow for a public information campaign.

In addressing the House of Commons Health Select Committee³⁰, Tim Kelsey, representing NHS England, made it clear that certain objectives had to be realised before data extraction could commence: these have not been specified but probably include completion and possibly evaluation (overseen by IIGOP) of the Pathfinders program.

²⁷ http://ico.org.uk/for_organisations/data_protection/the_guide/key_definitions#rights-obligations In relation to data controllers, the term jointly is used where two or more persons (usually organisations) act together to decide the purpose and manner of any data processing. The term in common applies where two or more persons share a pool of personal data that they process independently of each other.

²⁸ The Care Data Program Board has been established. there appears to be no published information on TOR, membership, meetings or minutes (30/09/14)

²⁹ In August 2013, GP practices received an information pack containing posters, a few leaflets and a letter saying that data extractions would start in eight weeks and they should inform their patients. The lack of any other information about the care.data programme or the GP data extract led to confusion among GPs and their staff about the program which in turn prevented patients being supplied with information.

Under media pressure, the programme was paused to allow the delivery of an information leaflet to all households in January 2014: however the leaflet itself has been judged not to supply the information needed and the means of delivery - a Royal Mail leaflet drop to those who had not opted out of junk mail - meant that many people either did not receive or did not notice the leaflet at all.

In view of this - and ongoing media debate - the programme was paused again until October 2014 "to allow communication" and is to be started with 100-500 Pathfinder Practices: it is not clear whether these have been identified as yet.

³⁰ <http://www.parliament.uk/business/committees/committees-a-z/commons-select/health-committee/news/14-02-25-cdd-ev/>

Patient trust

- If patients lack information and/or confidence in the whole care.data programme there are a number of possible adverse effects: It is likely to: dangerously inhibit them from discussing medical problems with their GP in the first place, with potentially harmful individual effects.
- They might conceal information or medical problems from their GP with potentially harmful effects for the community at large.
- opt-out in sufficient numbers may lead to significant if unknown bias in the data.

Research has consistently shown³¹ majority support for the concept of using individual patient data for secondary uses such as research, if patients are provided information on what their data is to be used for and individuals are given the opportunity to opt out. This majority falls to a minority if adequate information is not given, if patients have no choice or if data is used for purposes other than those people have been told about. The group felt that if patient confidence in the program is to be maintained or restored, this will depend on the provision of much more information about the program, how it will be implemented and governed and any future planned changes.

Planned outputs from the CLICSIG session

It was decided that the outputs from the meeting should include:-

- A report aimed at the House of Commons Health Select Committee following their three sessions on 25th February, 8th April and 1st July 2014³²
- A report for BCS Health to help inform BCS policy³³
- A report - this one - for publication on the website.

All of these to be preceded by further email discussion between participants to ensure the issues were clearly recorded.

³¹ e.g. by the MRC <http://www.mrc.ac.uk/public-engagement/open-meetings-and-consultations/public-consultation-on-the-use-of-personal-health-information-in-research-ongoing/> and Wellcome Trust <http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTX058859.htm>

³² The Health Select Committee's ongoing inquiry into the 'Handling of NHS patient data': <http://www.parliament.uk/business/committees/committees-a-z/commons-select/health-committee/inquiries/parliament-2010/cdd-2014/>

³³ This has been done, and is published in the September 2014 issue of the BCS Journal "IT Now", and is available to BCS members via the BCS secure page of the its web site. Non BCS members should contact stuartianherbert@gmail.com for additional information.

Annex 1: What is care.data?

- a) A comprehensive and permanent digital warehouse of individual-level identifiable patient data to be built by the HSCIC as directed by NHS England. It will comprise a record per English patient, consisting of linked data extracted from all organisations who have delivered state funded care to that person, starting with GPs and hospitals. Consent will be implied, i.e. 'opt out', rather than explicit ('opt in'). Ultimately it is likely to include other sources of personal data such as the Office of National Statistics (who hold the census data).
- b) The GP data will include patient diagnoses, observations and symptoms, prescriptions issued, test results received and referrals for treatment and investigation. Data about sensitive issues such as sexual health and HIV/AIDS will not be collected, nor will any data recorded more than 4 months before the collection starts. However NHS England has indicated that it may seek approval to collect an expanded data set in the future. It has not yet indicated a wish to collect free text data in patient records.
- c) Collection of GP data from 'pathfinder' practices was scheduled to start in October 2014, and after that the database will be updated at regular intervals. The HSCIC already collects and retains identifiable standard datasets about each episode of patient care from hospitals, and community and mental healthcare providers.
- d) The HSCIC will remove or obscure identifiers from care.data content it shares with end users unless authorised to release identifiable data by patient consent or the Health Research Authority's Confidentiality Advisory Committee (CAG). However the data in care.data is so rich that it will remain identifiable personal data as defined in the Data Protection Act even after this.
- e) There is a commitment that that individual-level linked data may only be accessed via a 'safe setting' in the pathfinder phase: there is no commitment beyond this phase.
- f) The ultimate intention is to enable a wide variety of NHS organisations, academics and companies to use it for a broad range of unknown purposes, including information intermediaries. Only very limited individual care.data could be published without offering a significant risk of patient re-identification.
- g) The General Practice Extraction Service Independent Advisory Group (GPES IAG) of the HSCIC approved the collection of GP data for care.data where used for commissioning, i.e. by those arranging contracts to supply NHS-funded care. It rejected a subsequent application for a much wider range of users and uses, though NHS England is planning to reapply shortly (and did so in September 2014 see footnote 19).
- h) NHS England opted to collect the GP input for care.data via a statutory 'gateway' created by sections 255 - 259 of the Health and Social Care Act 2012. If NHS England had decided to collect the data itself, approval would have been required from the CAG rather than GPES IAG. This might not have been given.
- i) The Secretary of State for Health has enabled patients to stop their identifiable GP data and/or identifiable information about them held by the HSCIC being shared for secondary purposes. These opt outs are 'grace & favour' arrangements, not legal rights.
- j) Even where identifiable data is collected without consent, the DPA 1998 states that patients should be informed about it. Without such 'fair processing' information, most patients would be unaware of the opt out. NHS England sees the provision of information about care.data as a job for the data controllers of the data they have directed the HSCIC to collect, as the DPA 1998 states, e.g. GPs for GP data.
- k) After meeting the Information Commissioner in autumn 2013, NHS England decided to send a leaflet about care.data to every English household rather than adult. Subsequent surveys showed that only 29% remembered the leaflet and so might be aware of the opt out, and that only 51% were aware of care.data. The leaflet did not mention care.data *per se* and gave misleading information about the opt outs and how to use them.

Annex 2: Draft recommendations re the care.data programme

It is essential to restore and maintain the trust of the population and clinical professions in this programme if the planned benefits are to be realised. We believe that implementation of the following recommendations is necessary to achieve this:

- a. Develop (and publish) detailed plan for programme, including:**
 - i. all proposed purposes and their expected benefits;
 - ii. how these will be realised and the risks involved;
 - iii. intended sources of data, including from GPs;
 - iv. consultation with patients and their clinicians on each plan to extend the data collected, e.g. additional codes, historical data and (much more problematically) free text;
 - v. how NHS England and the HSCIC will demonstrate that the data collected is necessary for the stated purposes. This includes but is not restricted to examination by the BMA / RCGP and clinical specialists, and review by an independent panel of patients and GPs;
 - vi. how fair processing information will be provided to all patients about any change to the data collected, the purposes it may be used for and the ultimate end users;
 - vii. management of applications from end-users;
 - viii. information governance (IG) and external oversight;
 - ix. that the collection respects any and all objections by patients to the sharing of their individual GP data in identifiable or anonymised form, and any patient objections to the sharing of any individual items in their records kept by care professionals;
 - x. ensuring that records which are confidential by statute (e.g. those of GUM clinics) are not collected or linked with care.data unless the patient provides explicit prior consent;
 - xi. clarity over dissemination of data other than properly-treated statistics (which should be published) or access in data-lab;
 - xii. independent vetting of applicants and applications for use³⁴;
 - xiii. no further distribution of entire datasets or of partial datasets containing individual-level patient data except where specific legal support (s251) has been given;
 - xiv. the default for access to individual-level linked data should be via a data laboratory ('safe setting') controlled by the HSCIC;
 - xv. tightly monitored and enforced controls over (or no) reuse;
 - xvi. audited data destruction by ultimate end users when use for the stated purpose(s) is complete.
- b. IG controls should include substantial representation of data controllers and their data subjects as well as IG experts.**
 - i. There should be independent audit of IG practices.
 - ii. 'Pathfinders' should be pilots, not part of a national roll-out. They should be used to evaluate and report on all aspects of care.data, not just communications. Pilot practices should either be chosen randomly, or systematically to include the true variety of attitudes to care.data. Their current number (100-500) is excessive for a pilot.
 - iii. Produce a Public Information Programme giving clear, comprehensive and accurate explanations and information to:
 - GPs and practice staff, to be conducted – and its effectiveness assessed – before ii starts
 - Patients and the public, including how they can control the sharing of data held by their GPs and other care professionals
 - iv. The Programme should establish criteria for success before it begins, and test them in the pilot implementations.
 - v. Consult /get approval from BMA & RCGP professional organisations, patient organisations, independent public interest bodies and patient representatives. The care.data Advisory Group's composition is not appropriate to provide this, and it is unlikely to be seen as sufficiently independent.
 - vi. Inform GPs and practice staff in pilot areas and evaluate their understanding.

³⁴ The HSCIC may only disseminate the data under paragraph 6(1) if:

(a) The recipient is a health service body, a provider of NHS-funded services or a local authority engaged in joint commissioning with an NHS organisation, or a person acting on behalf of any such body, and the information is to be used by that person or body for purposes relating to the exercise of public functions;

(b) The data are pseudonymised and the HSCIC is appropriately assured that the recipient would not be able to re-identify individuals from the information to be provided when linked to other information held by or likely to come into the possession of the recipient; and

(c) The recipient has signed a written agreement with the HSCIC which specifies the data to be made available, the purpose for which the recipient will use the data and the terms on which that data may be shared and re-used

- vii. Ensure all recommended changes at HSCIC are implemented and working, cf. Partridge report³⁵ and HSCIC CEO's 'work programme'³⁶.
- viii. Public phase of pilots + evaluation, including the publication of all materials, the results and their evaluation.
- ix. No extraction of data in pilot areas until all of the above have been satisfactorily completed.
- x. Only once pilots have been fully evaluated, evaluations have been published, the utility of data has been verified and a full business case has been approved by the Treasury should national roll-out be considered.
- xi. To ask, if the data quality in the pilots falls below anticipated levels,
 - what do NHS England/HSCIC propose to do about it, and;
 - what justification they have for continuing to extract low-quality data?
 - How this will affect the stated benefits of the programme?

c. Draft general recommendations

- i. Get roles, responsibilities and accountabilities of NHS England and HSCIC agreed, documented and published.
- ii. Establish and publish mechanisms for:-
 - Managing changes in data collection, management and distribution developed after the initial data has been collected and
 - How the public and professionals involved will be informed - which should be in such a manner that the information will reach both public, GPs and their staff and any others involved in providing, extracting, collecting, manipulating and making available the data will not be likely to overlook the changes.
- iii. Develop and enforce independent external review (including audit and IG oversight) of all care.data procedures - and provide resources to do this on an on-going basis.
- iv. The opt-out must be a statutory right, ideally guaranteed by primary legislation but at a minimum in secondary legislation by affirmative procedure rather than as a 'grace and favour' arrangement in Directions that may be amended without Parliamentary scrutiny. Or change the opt-out to an opt-in (i.e. explicit consent) as is now official BMA and RCGP policy.
- v. Invest in improving the quality of the data including PRIMIS+ type tools to assess it, resources such as training/'best practice' materials, financed time and requiring Read/CTV3/SNOMED use in GP training.
- vi. Ensure that those using GP & CES data understand its complexity and variability and non-standard nature. It is not at all like HES data.
- vii. Such a rich database will be a prime candidate for 'legal' misuse, and measures must be taken to prevent it and to minimise the damage if it does occur, e.g. by ensuring that the HSCIC does not hold identifiable data, e.g. by using pseudonymisation at source. By 'legal misuse' we mean future changes in statute which reduce the controls on the disclosure of identifiable data held by the HSCIC, including data already collected and linked for the care.data programme. (vide what HASCA 2012 did to GP patient data confidentiality).
- viii. Acknowledge, as the NHS Anonymisation standard (ISB 1523) points out, that only very simple extracts of individual care.data can be published without creating significant risks to patient data confidentiality.
- ix. Any organisation holding data that poses a risk of identifying the patient(s) – even if this may previously have been classed as 'non-significant' – must adhere to the same security standards as the HSCIC and be governed and audited in the same way and to the same standards.

³⁵ See Partridge's 9 recommendations: http://www.hscic.gov.uk/media/14244/Sir-Nick-Partridges-summary-of-the-review/pdf/Sir_Nick_Partridge%27s_summary_of_the_review.pdf

³⁶ See (different) 10 points in HSCIC's Partridge Review press release: <http://www.hscic.gov.uk/article/4780/HSCIC-learns-lessons-of-the-past-with-immediate-programme-for-change>