JAMIA

Privacy protection and research access mechanisms for National Health Service data: The Clinical Practice Research Datalink (CPRD)

Journal:	*Journal of the American Medical Informatics Association*
Manuscript ID:	Draft
Article Type:	Research and Applications
Keywords:	Electronic Health Records, Computer security, Information systems, Privacy, Confidentiality

Privacy protection and research access mechanisms for National Health Service data: The Clinical Practice Research Datalink (CPRD)

Tim Holt[1], Tarita Murray-Thomas[2], Tim Williams[2], John Parkinson[2]

[1]University of Oxford, Oxford, UK

[2]Clinical Practice Research Datalink (CPRD), Medicines and Healthcare products Regulatory Agency (MHRA), London, UK

Corresponding author:

Dr Tim A. Holt PhD MRCP FRCGP

NIHR Academic Clinical Lecturer

University of Oxford

Department of Primary Care Health Sciences

2nd floor

23-38 Hythe Bridge Street

Oxford OX1 2ET

Tel: +44 (0)1865 289281

Fax: +44 (0)1865 289287

Email: tim.holt@phc.ox.ac.uk

Key words: Electronic Health Records; Computer security; Information systems; Privacy; Confidentiality

Word count (excluding Abstract and References): 3509 words

ABSTRACT

Electronic health records were introduced into UK primary care during the late 1980s to support consultation based record keeping, prescribing, practice administration, and later audit and performance management. National Health Service (NHS) policies required standardisation of coding to support future interoperability and the integration of structures to enable equitable provision of care across different regions. These processes also led to the establishment of research databases containing large volumes of anonymised, routinely collected data extracted from participating general practices. Whilst the completion of NHS clinical software integration has proven elusive, these resources have benefitted from the data standardisation process designed to support it. They provide huge statistical power for addressing observational research questions including pharmaco-vigilance. Linkage to other clinical databases, hospital data, and the Office for National Statistics mortality data has been established through trusted third parties. Recently, the potential for supporting primary care based interventional trials has been developed. Whilst the information stored in these secondary databases contain no personal identifiers, governance arrangements take account of legislation introduced during the 1990s to offset public concern over the control of sensitive health information. 2012 sees the creation of the Clinical Practice Research Datalink, which will bring together anonymised data from a significant proportion of UK practices with extensive linkage to other national datasets for the benefit of patients, UK public health, life science industries and the international academic community. This paper describes how the challenges of privacy protection and data sharing will be addressed in this new programme.

BACKGROUND AND SIGNIFICANCE

Clinical software infrastructure in the UK has from the outset been influenced by the existence and requirements of the National Health Service (NHS), a system established in 1948 to provide equitable access to health care that is free at the point of delivery. Fifty years later, a proposal for NHS software integration was made by the *National Programme for IT* (NPfIT)[1]. This described a vision for NHS software development including the NHS Care Records Service, through which patient records could be accessed from outside individual practices and therefore beyond the team responsible for usual care. This was designed to support care at remote locations, such as accident and emergency facilities, treating acutely ill individuals away from their home base, and was generally welcomed, but raised issues over data control and privacy. Despite difficulties in completing this vision, standardisation of data coding and the integration of previously unconnected

domains (such as those of hospital laboratories and primary care records) succeeded in achieving the necessary interoperability to support the *Quality and Outcomes Framework* (QOF) established in 2004[2]. This 'payment by results' system required the use of QMAS (Quality Management and Analysis System)[3] software that extracts relevant data anonymously from practices to monitor performance remotely against QOF targets. These developments moved chronic disease management beyond individual patient care at the practice level, and closer to a nationally distributed public health endeavour.

From a research perspective, routinely collected primary care data were a potentially rich resource from an early stage, but required careful interpretation[4]. Data began to be extracted from multiple sites into the *General Practice Research Database* (GPRD) as early as 1987[5], and this usage increased during the following two decades, incorporating more recently innovations involving linkage to secondary data sources, enhanced methodologies and novel applications[6]. This led on to a range of data repositories and integrated data collection systems summarised by Gnani and Majeed[7]. In addition to GPRD and QMAS, they include MIQUEST (Morbidity Information Query and Export Syntax)[8], Prescribing Analysis and Cost (PACT) data[9], the RCGP Weekly Returns Service[10], the Primary Care Information Service (PRIMIS)[11], The Health Improvement Network[12] and QRESEARCH, a large database hosted at the University of Nottingham[13]. Electronic data recording was, at the start of the 1990s largely designed to support individual care. It then expanded to meet the needs of clinical audit, later becoming a tool for monitoring adequacy of care at the practice level and of comparing different practices by primary care organisations. These were able not only to extract anonymised data remotely (as GPRD already could) but also to feed the results back to practices on a regular basis. This process required a certain level of code standardisation that was unnecessary for the requirements of the decade before.

Central to this process were the concepts of disease registers and unique identifiers. Disease registers have become the focus of structured, systematic chronic disease management in UK primary care, and are an important basis for the interpretation of clinical behaviour and of health outcomes in research settings. NHS numbers have enabled linkage of information between records held in different clinical databases, as discussed below.

Privacy legislation

In 1980 the global Organisation for Economic Co-operation and Development (OECD) defined a number of key principles related to data security, access and accountability[14]. These have

influenced later policy development in regions and individual nations, including Europe and its member countries. In the UK, the Data Protection Act (DPA) was first introduced in 1984, and amended in 1998[15] in the context of the European Data Protection Directive of 1995 (which is itself currently in the process of revision)[16]. The DPA defines 'data controllers' and describes their responsibilities with respect to personal data, including more specifically 'sensitive personal data'. Such data (and the Act itself) relate to identifiable, living individuals. In the specific arena of health care, increasing public concern over the sharing of personal health information led to the Caldicott Report of 1997[17]. This established a clear 'need to know' principle through which the purposes of health data collection and usage must be defined and justified. It requires each NHS organisation (from large Strategic Health Authorities to individual general practices) to identify a 'Caldicott Guardian' who is responsible for all matters related to personal data control and privacy protection.

These developments largely relate to the handling of information on identifiable individuals. However, the 'need to know' principle has also influenced access to anonymised health data. The three major UK primary care databases (GPRD, QRESEARCH, and THIN) all collect pseudo-anonymised data that contain no strong identifiers, i.e. no details of the person's name, address, or other contact details. Record ID numbers are applied in the practices that enable individual records to be updated over time, but these numbers are not released to researchers[18]. Those registered with the practices contributing data are, in the GPRD system, informed by practice poster and leaflets that their anonymised data will be used and linked for research unless they specifically request their general practitioner to flag their record as not for use in such research. In practice the number of patients that opt out is very small.

Each database requires detailed justification of access to the data and each has a Scientific Advisory Committee. Such committees review all proposed study protocols; approving, suggesting amendments or rejecting as they decide. Ethical approval is also required for such work and in most cases exists under a blanket approval arrangement for observational studies. Individual studies requiring direct access to patient populations must in addition seek their own ethical approval from relevant approving bodies. Study protocols often define medical code sets that relate to the pre-specified research question and allow the identification of the anonymous records of relevant individuals in the repositories. In the case of QRESEARCH, only the coded events, observations, prescriptions and descriptors are released rather than the whole health record. In all cases, the investigators must confirm through a licence that no attempts will be made to identify individual people or practices. In GPRD, the lowest health administrative area of the UK (defined here as that

covered by a primary care trust) for which data is made available for research must comprise of at least three GPRD contributing practices within that area. GPRD is available to the international academic community through an on-line secured connection (GPRD GOLD)[19].

Linkage to other databases

The opportunity to link primary care databases to other sources of information greatly increases their potential to address a range of research questions. Despite moves to improve interoperability, there is a sharp division in the UK between the clinical software systems used in primary and secondary (hospital based) care settings. Linkage with Hospital Event Statistics (both inpatient and outpatient)[20] allows research to occur at the interface between these domains and provides a more inclusive capture of health outcomes. Linkage to the Office for National Statistics[21] enables death certificate data to confirm the cause of death of individuals, which is not always recorded in primary care after a person has died. Other arrangements involve national disease registries including linkage of GPRD to the National Cancer Intelligence Network[22]. However, these linkages require strong identifiers including NHS number, sex, date of birth and post code which, as noted above, are not included in the primary care databases as this would violate the anonymity principle. To overcome this, trusted NHS third parties are used that provide a linkage service without either database 'seeing' the other's data, as discussed further below.

The established model of voluntary linkage of practices to the large research databases is evolving. The latest move will establish a comprehensive data collection system that will involve a large proportion of UK practices linked to numerous other health data sources as the Clinical Practice Research Datalink.

ESTABLISHMENT OF THE CLINICAL PRACTICE RESEARCH DATALINK (CPRD)

CPRD is jointly funded by the Department of Health's National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA). It combines the original GPRD operation with the extensive work undertaken within the Research Capability Programme[23] over the last 4 years. CPRD was introduced in April 2012 as part of the UK government's Plan for Growth[24] and will use a federated approach to integrate many NHS datasets and other data which may be useful for health research. Through the iterative development of data linkages with relevant sources CPRD will enable access to observational data to facilitate epidemiological research, drug safety/effectiveness and risk-benefit research, help to support surveillance activities and more importantly facilitate interventional research within the database setting.

Greater access to data and more extensive linkage arrangements under the CPRD initiative could potentially increase the risks of re-identification. However, CPRD proposes to safeguard patient privacy at all levels of its operations starting with the use of appropriate privacy enhancing technologies for privacy consolidation at the design stage, the use of a trusted third party to undertake data linkage and the implementation of privacy impact assessments, performance evaluations, legal contracts and audit. It will use anonymisation methods that protect privacy without losing the functionality of data sources for research. Figure 1 depicts a summary of the processes for ensuring overall data stewardship in CPRD.

Non-interactive and interactive frameworks for protecting privacy in CPRD

CPRD will build upon the privacy mechanisms supporting GPRD, integrated with input from the Department of Health Research Capability Program[23]. In the case of the existing GPRD privacy is protected mainly under a non-interactive framework in which original data is first 'sanitized' and a modified version is then released to users[25]. In contrast to an interactive framework, the non-interactive solution allows data sanitization to be conducted offline as interactions with users are not required[26]. The risk of accidental disclosure of sensitive data is avoidable under this framework. The potential limitation of using a non-interactive model is that as CPRD expands over time it may become more difficult to provide utility that has not yet been specified at the time that the sanitization is undertaken[27]. The anonymisation of GPRD data is implemented at source as part of the data extraction process from practices. Strong identifiers (e.g. name, address, post code, telephone number) are removed and other fields are generalised where necessary (e.g. date of birth becomes year of birth)[28].

Interactive mechanisms will be used, under certain circumstances, for handling highly sensitive information such as infectious disease data. Under this framework, data queries will be submitted through a mechanism that can either deny queries, or alternatively modify or suppress the query output in order to ensure privacy[26]. This query auditing approach combined with output perturbation methods have been shown to be of comparable and even of better quality than some non-interactive solutions[29]. However, significant levels of data perturbation will have potential effects on analysis possibly introducing bias and misclassification to observational research.

Data encryption

CPRD will markedly expand the patient populations available for research by integrating the collection of longitudinal patient data from different practice management software systems. Such

data will be accessible under a non-interactive framework and privacy will be assured through the use of Privacy Enhancing Technology (PET)[29, 30]. Privacy legislation such as the European Union (EU) Data Protection Directive 95/46/EC, Article 17 Security of Processing provides the legal basis for the use of PET in securing data at such levels[31]. PET will be used to achieve data encryption without the need to collect information such as names, addresses and NHS numbers. Coupled with appropriate levels of governance relating to access and use of data, PET will help to minimize the risk of re-identification of individuals in the database. PET would operate during the data collection process to encrypt the identifiers of patients, doctors and other practice staff who enter data into the practice management system. In this way, all data regardless of its origin of collection by the CPRD group will be pseudonymised. As an additional safeguard, patient and practice identifiers will be encrypted for a second time prior to release to researchers via the CPRD data warehouse. CPRD policies will be consistent with guidance provided by the Information Commission Office (ICO) on the use of PETs for maintaining privacy[32].

In those circumstances where patient identifiable information must be made available for research CPRD will only make this available if individuals have given informed consent for their data to be used or where researchers have been granted exemption by the National Information Governance Board for Health and Social Care (NIGB)[33] to use identifiable data under Section 251 of the NHS Act 2006 (formerly Section 60 of the Health and Social Care Act 2001)[34].

Trusted Third Party (TTP)

CPRD will have access to patient demographics data for the whole of the National Health Service (NHS) in England using data maintained by the national electronic database of the Personal Demographic Service (PDS)[35]. While the PDS does not capture clinical or sensitive data items it provides access to patients NHS number which will enable the deterministic linkage of persons across data sources. CPRD will not have direct access to PDS data but will have access to this via a trusted third party, the Information Centre for Health and Social Care[36]. At the level of the trusted third party, PDS data will be used to generate and store a unique encrypted identifier for each person (CPRD ID). This will provide the infrastructure needed to support record linkages among any combination of data sources. Use of the CPRD ID will enable audit and replication of analysis in key studies of high public health importance. This is a key feature of emerging guidelines such as Good Pharmacoepidemiological Practice[37] and recent ENCePP Guidelines[38].

CPRD will use the established protocol for record linkages via TTP as developed by the GPRD and

further extended by the work of the Research Capability Programme. It will also extend this approach to include privacy preserving linkage mechanisms. This extension is important as the number of data linkages under CPRD is expected to surpass significantly that currently undertaken by GPRD. Under this proposed arrangement, the TTP will continue to be independent of both the NIHR and MHRA and will serve as the mediator of the linkage process. In principle, data holders and the TTP will agree on the identification data to be used for the linkage and all data holders will be required to supply unique serial identifiers and encrypted identification data to the TTP. Using deterministic and probabilistic techniques, the TTP will link individual records and create a unique linkage ID. Once the linkage is complete, the TTP will send the unique serial identifier and linkage IDs back to each respective data holder and will destroy all encrypted identification data used to generate the linkage. Data holders will then be able to add the linkage IDs to their dataset using the unique serial ID and send the required non identifiable dataset with linkage IDs to researchers. Once a common linkage ID exists across systems and datasets, it will be possible for researchers to create linked data sets that are de-identified and which they can then use to examine important public health and drug safety issues.

Privacy Impact Assessments (PIA)

CPRD will conduct privacy impact assessment at pre-specified intervals to oversee privacy, confidentiality and security. PIAs are therefore likely to reduce privacy risks to individuals, build public trust and confidence and at the same time identify where potential vulnerabilities may exist. Although privacy impact assessments are not a statutory requirement in the UK, many government departments are required to undertake them. CPRD will conduct PIA according to the processes and guidance outlined by the Information Commissioner's Office[39] and as required under Department of Health and other government regulations.

Information governance

CPRD will adopt a multi-layered approach to information governance similar to that previously proposed under the Research Capability Program[23]. Under this model, CPRD will formulate policies and provide technical solutions to protect patient privacy and will work collaboratively with partner organisations such as the NHS Information Centre, research and other communities and agencies to safeguard patient confidentiality.

NIGB approval will be a necessary prerequisite for record linkage under CPRD. This would be required on a linkage-by-linkage basis and would be pursued by CPRD on behalf of researchers

according to its record linkage policy. At the level of data access by researchers, scientific approval for undertaking research involving unlinked or linked patient level data will be adjudicated by the MHRA Independent Scientific Advisory Committee (ISAC). Where data owners may have additional governance relating to access and use of their data, CPRD will develop collaborative partnerships with such organizations to ensure that the potential for patient re-identification is minimized and appropriate use of the individual data source is maintained.

Accessibility arrangements

Data held in the GPRD is currently available via two mediums: across a virtual computing environment (VCE) with additional inbuilt security features and as ad-hoc datasets/analysis files. CPRD will build on the VCE technology of GPRD to produce secure robust e-based systems to access all aspects of CPRD services, from data set provision to clinical trial feasibility assessments. In terms of the underlying infrastructure CPRD is likely to consider implementation of Wide area or Local area distributed database solutions, or potentially newer technology such as cloud based solutions[40, 41]. Adopted solutions will need to provide security, service continuity, scalability and appropriate levels of response time (depending upon actual tasks). Whilst cloud computing offers good solutions regarding service continuity, scalability and access security it involves spreading data over a wide network with physical duplication or mirroring which may well provide a barrier to its use with healthcare data from a data governance point of view. Local solutions, however, enable a greater degree of control over the physical data, but have less IT resource at their disposal in terms of ensuring high quality uninterrupted research service provision.

For online access, CPRD will use the general governance and security procedures of the GPRD to assign and monitor security. Access will only be granted to users who hold the appropriate secure electronic passes and meet the strict criteria for holding the same.

Potential to support clinical trials

CPRD will support capability to enable pragmatic randomised clinical trials (p-RCT) to be undertaken in the primary care setting. The technical and operational mechanism for doing so has already been developed by the GPRD[42]. This involves a system developed to facilitate patient recruitment at the practice, informed by lists of patients identified as potentially eligible. Patients may be invited to attend appointments or clinics for recruitment into studies, or alternatively recruitment may take place opportunistically as part of a face to face consultation. Once recruited and consented to the study patients are randomised to an intervention. As with the GPRD, this will be mediated by a study

specific e-based system within CPRD and external to the primary care setting. No identifiable data will leave the practice. Primary care data will be accumulated and downloaded as per normal data collection processes on a daily basis. These data will then be processed into a separate security ring-fenced data repository, where patient follow up data are collated with TPP linked secondary data repositories such as Hospital Episode Statistics and ONS data. The p-RCT system is Good Clinical Practice (GCP) compliant[43] and includes processing systems to facilitate adverse event (AE) reporting, blinded database creation and fraud detection. CPRD will extend these capabilities to also enable phase 3 clinical trials to be conducted within the primary care setting. Using patients' electronic health records as the backbone for collecting clinical and non-clinical data, CPRD will integrate and unify processes to produce an electronic case report form. The proposed system will enable real time access to recruitment information, resource utilisation, AE reporting data, outcome identification and long term follow-up of patients. CPRD will work with various primary care software vendors and interested parties to achieve both technical and semantic interoperability[44] to ensure that processes and data capture are harmonised.

SUMMARY

The Clinical Practice Research Datalink (CPRD) is a major development in the integration of NHS health data building on the established processes of the General Practice Research Database (GPRD) and the Research Capability Programme. It takes advantage of two decades of work designed to promote interoperability of component systems within the UK National Health Service. It will significantly exceed the functionality of existing resources in terms of data volume, linkage and accessibility to support research and health care delivery. This brings with it challenges to the protection of privacy, challenges to be addressed through a range of privacy protecting arrangements, including trusted third parties and privacy enhancement technologies. The experience of GPRD indicates that a rolling programme of system redevelopment will be necessary over time to keep pace with the expanding volume of data, innovations in systems operating within the health care environment and development of new clinical software solutions.

Ethical approval

No ethical approval required

Funding

No external funding

Competing interests

Dr Tim Holt has no competing interests. Tarita Murray-Thomas, Dr Tim Williams and Dr John Parkinson are employees of the Medicines and Healthcare products Regulations Agency that is responsible for developing the Clinical Practice Research Datalink. They have no competing interests.

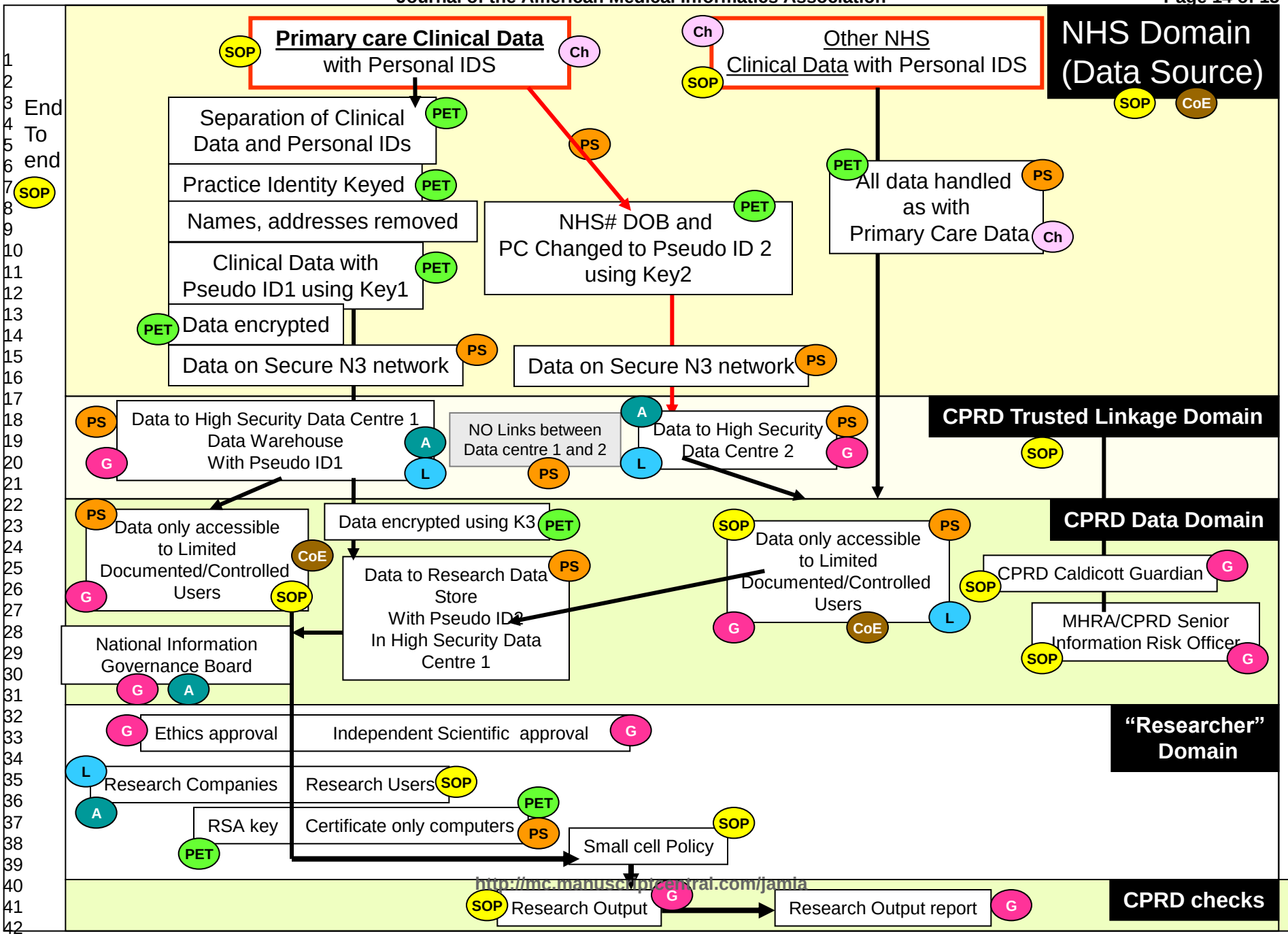
References

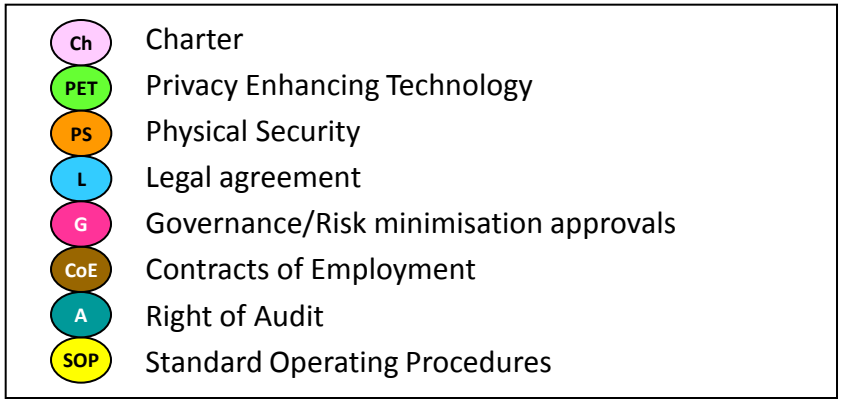
1. Department of Health. *Information for health: an information strategy for the modern NHS 1998-2005.* London: DoH, 1998.
2. www.qof.ic.nhs.uk/ (Accession date 29.2.12)
3. http://www.connectingforhealth.nhs.uk/systemsandservices/gpsupport/qmas.
4. Pringle M, Hobbs R. Large computer databases in general practice. BMJ 1991;**302**(6779):741-2.
5. http://www.gprd.com/home/default.asp (Accession date 29.2.12)
6. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Therapeutic Advances in Drug Safety 2042098611435911, first published on February 2, 2012 as doi:10.1177/2042098611435911
7. Gnani S, Majeed A. *A user's guide to data collected in primary care in England.* Eastern Region Public Health Observatory, 2006.
8. http://www.connectingforhealth.nhs.uk/systemsandservices/data/miquest (Accession date 29.2.12)
9. Lovejoy AE, Savage I. Prescribing analysis and cost tabulation (PACT) data: an introduction. Br J Community Nurs 2001;**6(2)**:62-7.
10. RCGP Weekly Returns Service. http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/RealtimeSyndromicSurveillance/SyndromicSystemsAndBulletinArchive/primcRCGPWeeklyReturnsService/.
11. University of Nottingham. http://www.primis.nhs.uk/ (Accession date 29.2.12)
12. http://www.thin-uk.com/
13. University of Nottingham. http://www.qresearch.org/SitePages/Home.aspx (Accession date 29.2.12)
14. http://www.oecd.org/ (Accession date 29.2.12)
15. http://www.legislation.gov.uk/ukpga/1998/29/contents (Accession date 29.2.12)
16. http://ec.europa.eu/justice/data-protection/index_en.htm (Accession date 29.2.12)

17. The Caldicott Committee. Report on the Review of Patient-Identifiable Information. London: Department of Health, 1997.
18. http://www.qresearch.org/SitePages/Confidentiality.aspx (Accession date 29.2.12)
19. http://www.gprd.com/services/online.asp (Accession date 29.2.12)
20. http://www.hesonline.nhs.uk/ (Accession date 29.2.12)
21. http://www.statistics.gov.uk/hub/index.html (Accession date 29.2.12)
22. http://www.ncin.org.uk/home.aspx (Accession date 29.2.12)
23. Research Capability Programme Information Governance Framework (PD16). www.nihr.ac.uk/...Programme.../PD16%20IG%20framework.pdf (Accession date 22/02/2012)
24. Research & Development Directorate. *The Government plan for a secure data service: Strengthening the international competitiveness of UK life sciences research.* London: Department of Health, 31st October 2011.
25. Chawla S, Dwork C, McSherry F, Talwar K. On the utility of privacy preserving histograms. In: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, 2005.
26. Domingo-Ferrer J. A Three-Dimensional Conceptual Framework for Database Privacy. Computer Science 2007;**4721**:193-202.
27. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. In: Proceedings of the 3rd Theory of Cryptography Conference, pages 265–284, 2006.
28. Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression. International Journal on Uncertainty, Fuzziness and Knowledge based Systems 2002;**10(5)**:571-588.
29. Tavani HT, Moor JH. Privacy protection, control of information and privacy-enhancing technologies. Readings in cyberethics. Maynard, MA: Jones and Bartlett, 2004: 436-449.
30. European Commission Community Research and Development Information Service (CORDIS). Privacy Protection and Electronic Identity Management http://cordis.europa.eu/fp7/ict/security/eid-management_en.html#priv (Accession date 22/02/2012)
31. European Union (EU) Data Protection Directive 95/46/EC. http://www.issa.org/Library/Journals/2011/February/Sorensen-European%20Union%20Data%20Privacy%20Directive.pdf (Accession date 22/02/2012)
32. Information Commission Office – Privacy Enhancing Technology. www.ico.gov.uk/.../data.../privacy_enhancing_technologies.pdf (Accession date 22/02/2012)
33. National Information Governance Board for Health and Social Care : http://www.nigb.nhs.uk/s251

34. http://www.legislation.gov.uk/ukpga/2006/41/section/251
35. http://www.connectingforhealth.nhs.uk/systemsandservices/demographics/pds (Accession date 28.2.12)
36. http://www.ic.nhs.uk/ (Accession date 28.2.12)
37. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf 2008;**17(2)**:200-8.
38. ENCePP Guide on Methodological Standards in Pharmacoepidemiology EMA/95098/2010 (Amended: 23 August 2011) available at: http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE.pdf (Accession date 29.2.12)
39. Information Commission Office- Privacy Impact Assessment. http://www.ico.gov.uk/for_organisations/data_protection/topic_guides/privacy_impact_assessment.aspx (Accession date 22/02/2012)
40. Nelson MR. Briefing Paper on Cloud Computing and Public Policy, Prepared for the OECD ICCP Technology Foresight Forum, October 14, 2009.
41. SIIA White paper: Guide to cloud computing for decision makers http://www.siia.net/blog/index.php/2011/07/siia-releases-guide-to-cloud-computing-for-policy-makers/ (Accession date 22/02/2012)
42. Van Staa T-P, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, et al. Pragmatic randomised trials using routine electronic health records. BMJ 2012;**344**;e55.
43. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf (Accession date 28.2.12)
44. Chan LM, Zeng ML. Metadata Interoperability and Standardization – A Study of Methodology Part I - Achieving Interoperability at the Schema Level. D-Lib Magazine 2006;12(6). ISSN 1082-9873.

CPRD (MHRA) Summary Data Stewardship Map




Ch
Charter
PET
Privacy Enhancing Technology
PS
Physical Security
L
Legal agreement
G
Governance/Risk minimisation approvals
CoE
Contracts of Employment
A
Right of Audit
SOP
Standard Operating Procedures