

Resource report

An outline framework for the efficient onward-sharing of linked Longitudinal Study and NHS Digital records

Andy Boyd^{1, 2}, Garry Coleman³, Estelle Spence³, Alison Park⁴, Rebecca Hardy²

¹ ALSPAC, Population Health Sciences, Bristol Medical School, University of Bristol

² CLOSER, UCL Institute of Education

³ NHS Digital

⁴ University College London

December 2019



Correspondence:

Andy Boyd
University of Bristol
a.w.boyd@bristol.ac.uk

First published in 2019 by
CLOSER
UCL Institute of Education
20 Bedford Way
London WC1H 0AL
www.closer.ac.uk

© CLOSER



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

About CLOSER

CLOSER, the home of longitudinal research, is an interdisciplinary partnership that brings together world-leading longitudinal studies with participants born throughout the 20th and 21st centuries, the British Library and the UK Data Service. Our work aims to maximise the use, value and impact of the UK's longitudinal studies in order to help improve our understanding of key social and biomedical challenges. CLOSER was funded by the Economic and Social Research Council (ESRC) and the Medical Research Council (MRC) from 2012-17, and by the ESRC from 2017 to present. Visit www.closer.ac.uk.

This document is available in alternative formats.

Please contact CLOSER:

tel: +44 (0)20 7612 6938

email: closer@ucl.ac.uk

Citation of the Report:

Boyd A, Coleman G, Spence E, Park A, Hardy H. (2019). An outline framework for the efficient onward-sharing of linked Longitudinal Study and NHS Digital records. London, UK: CLOSER, University College London.

Contents

Executive Summary	3
Current status and key areas for consideration.....	3
Recommendations	4
Next Steps.....	5
1. Introduction	6
2. Prerequisite conditions	7
3. Data Sharing within NHS Digital and within LPS: setting the scope of ‘onward-sharing’	10
4. Different approaches to how data are shared and accessed during research	13
5. Derived Data.....	15
6. National Opt-Out.....	17
7. Cost Recovery.....	18
8. Model 1: Anonymous data sharing.....	18
9. Model 2: Anonymised data sharing (also referred to as ‘effectively anonymised’, ‘anonymised in context’)	19
10. Model 3: Sub-licence model	22
11. Model 4: Multiple agreement model	24
12. Model 5: Trusted Third-Party model	26
13. Cross-LPS studies.....	27
14. A decision-making framework for onward-sharing.....	28
15. Alignment with other providers of health records	30
16. Conclusions	30
Annex 1: Summary of timeline of tasks for NHS Digital and LPS Community.....	32
Annex 2: NHS Digital Pre-Requisite Conditions (for Model 2)	33
Annex 3: NHS Digital Pre-Requisite Conditions (for Models 3-5)	35

Executive Summary

Current status and key areas for consideration

- i. Current data sharing agreements between Longitudinal Population Studies (LPS) and NHS Digital (typically) do not permit for the onward-sharing of linked data beyond the LPS's host institution or provide mechanisms to add new research hypotheses without a full data sharing amendment.
- ii. This joint piece of work between the CLOSER longitudinal study consortium and NHS Digital has been successful in identifying five distinct models to enable onward-sharing and the investigation of new research purposes in efficient ways. Aligned with these models are possibilities for onward-sharing inherent when creating derived data to the point where ownership and Data Controller status transfers to the study (and host institution). These models have been developed with contributions from many UK LPS, key UK secure research platforms/infrastructure providers and feedback from the NHS Digital Research Advisory Group (RAG)¹.
- iii. The most likely routes to be used by LPS for onward-sharing are firstly, to establish a formalised sub-contract arrangement with NHS Digital; or secondly to render the data anonymised² by pseudonymising the data and applying a relatively resource intensive 'wrapper' of safeguards (the 'anonymised' model).
- iv. The relevant Information Commissioner's Office (ICO) code of practice is currently being updated, and therefore there is a risk that the anonymised model may not meet any new guidance. This can be mitigated by considering the ICO's and other data controllers' current positions. To use other models might introduce resource issues which could act as a barrier to onward-sharing in some studies or by some groups of users.

¹ <https://digital.nhs.uk/services/research-advisory-group>

² Key terms relating to states of identifiability (i.e. pseudonymisation, anonymised, anonymous) are defined in section 4.3.

- v. While a number of recommendations have been made (summarised below), this report recommends that those relating to the development of the anonymised and ‘sub-contract’ models should be prioritised for further development, together with joint work to minimise the need for such approaches.

Recommendations

- vi. Recommendations are split into two groups: those which require actions within the control of the author’s organisations including those based around further collaborative discussions with the LPS community; and those which relate to the wider health system.

The recommendations are outlined below.

Recommendations within control of the author’s organisations or for further collaborative discussion with the LPS community:

- **Recommendation 1:** For **NHS Digital** to revise their guidance to data applicants to make clear the potential routes for onward-sharing; and for CLOSER to promote these routes to the LPS and wider data science community.
- **Recommendation 2:** For the **LPS community** to consider whether they are likely to adopt a common onward-sharing model (from the options identified within this report) and to consider whether there are efficiencies to be gained from addressing some development and ongoing tasks at a community level rather than study level.
- **Recommendation 3:** For **NHS Digital** to issue practical guidance in relation to disseminating data beyond the European Economic Area in line with ICO guidance.
- **Recommendation 4:** For **NHS Digital** to revise their guidance on when data processing meets the threshold of producing ‘derived data’. For this guidance to be illustrated with exemplars based on diverse data types and processing contexts; and for CLOSER, other LPS, and studies with other research designs (e.g. RCTs) to help nominate these examples.

- **Recommendation 5:** For **NHS Digital** to consider the sensitivity of ‘National Opt-Out’ status flags and whether these can be shared (under controlled conditions) in order for research users to apply them.
- **Recommendation 6:** For **NHS Digital** to consider whether the cost recovery model should vary across onward-sharing models in line with its costing review.
- **Recommendation 7:** as the ICO update the ‘Anonymisation: managing data protection risk code of practice’, for the **LPS community** and **NHS Digital** to ensure that guidance on key points related to onward-sharing and the required controls are in alignment.

Recommendations which relate to the wider health system:

- **Recommendation 8:** For the **LPS community** to consider and consult on the potential for the ‘anonymised’ model to form a gateway for the sharing and processing of health data, and potentially other data, from across the UK Home Nations and to promote this option if it is seen as being valid and acceptable.
- **Recommendation 9:** For the findings within this report to be disseminated to all UK providers of health care records for public good research.

Next Steps

- vii. NHS Digital and the LPS community to build on the initial group to take forward the recommendations above. The work will include publishing precedents, wider and more effective communications, and encouraging a co-ordinated approach by the LPS community (a timeline for these is included in Annex 1).
- viii. LPS studies wishing to implement onward-sharing agreements are to amend their Data Sharing Agreements to reflect this new data use. NHS Digital staff will facilitate these amendments in line with the models set out in this report and emerging good practice identified through LPS setting study specific precedents.

1. Introduction

- 1.1 UK longitudinal population studies (LPS) frequently operate as ‘data banks’ with a remit to share processed data with the wider ‘bona fide’ or ‘safe’³ research community. At a high-level, the development of these resources can be characterised as having a broad scientific basis, meaning LPS collect diverse data to inform diverse investigations. It is not possible to predict what these investigations may be as new requests to use study data are received on a regular basis, and LPS operate over very long time periods and have a remit to be responsive to wider scientific developments and innovation.
- 1.2 LPS collect data directly from participants and also via linkage to participants’ routine records which are held/owned by other organisations. To maximise the value of these linkages, given their remit and operating models, LPS require mechanisms to onward-share linked records with ‘bona fide’/‘safe’ researchers in an efficient, flexible and dynamic manner. This mechanism must be acceptable to participants; equally, data owners need to ensure their records are used and curated in a legal manner that is compatible with public expectations.
- 1.3 Such arrangements will need to allow;
 - i. broad data extracts that are minimised to study samples, but not specific research hypotheses (e.g. the full life-course Hospital Episode Statistics record for a study participant);
 - ii. flexible and efficient mechanisms to agree new research purposes (as future databank use is unknown at the point of submitting any linkage proposal);
 - iii. to enable managed access/sharing with bona-fide users from other institutions (in line with NHS Digital requirements).
- 1.4 CLOSER and NHS Digital have conducted this joint piece of work to establish a generalisable framework for LPS to onward-share linked NHS Digital records.

³ ‘Bona Fide’ and ‘Safe’ are similar models proposed by the research funders and the Office for National Statistics for determining whether a research applicant is legitimate or not.

- 1.5 CLOSER and NHS Digital have set out a roadmap by which we seek to define, through stakeholder consultation, this framework. This roadmap has been reviewed and approved by NHS Digital Research Advisory Group (NHS Digital RAG).
- 1.6 A meeting of LPS, research infrastructure providers and LPS consortia⁴ was held to discuss ‘Onward-Sharing’ (17th December 2018, UCL Centre for Longitudinal Studies, Gordon Square, London). A draft of this report was presented to the NHS Digital RAG for comment and guidance (25th July 2019, UK Research and Innovation, Victoria Embankment, London). This report has been updated to reflect the insights from these meetings.
- 1.7 This report describes the framework for onward-sharing, the specific models for onward-sharing, and the permutations within these.

2. Prerequisite conditions

- 2.1 Institutions wishing to hold and use NHS Digital data must have a valid NHS Digital Data Sharing Framework Contract (DSFC) and Data Sharing Agreement (DSA) in place. The terms within these contracts/agreements will need to be extended to cover the onward-sharing of linked data. They are not described in this document (see the NHS Digital Data Access Request Service website⁵), but adherence to the terms are essential requirements in all the onward-sharing models. Broadly, the controls and consideration given to sharing via an onward-sharing model must, at minimum, mirror those in the DSFC and DSA established between NHS Digital and the study/institute.
- 2.2 Research data can be held indefinitely⁶. However, the permissions surrounding the holding of research data under licence may be time-bound or subject to regular

⁴ Delegates attended from 13 longitudinal studies – 100k Genomes, Avon Longitudinal Study of Parents And Children, Born in Bradford, 1970 British Cohort Study, Millennium Cohort Study, Million Women Study, 1958 National Child Development Study, 1946 MRC National Survey of Health and Development, Next Steps, TEDS, Twins UK, UK Biobank, Understanding Society; the UK Data Service and UKSeRP; research infrastructure providers; CLOSER, the UK Dementia Platform study consortium; NHS Digital.

⁵ <https://digital.nhs.uk/services/data-access-request-service-dars>

⁶ Boyd A, Woollard M, Macleod J, Park A. The destruction of the ‘Windrush’ disembarkation cards: a lost opportunity and the (re)emergence of Data Protection regulation as a threat to longitudinal research. Wellcome Open Research. 2018;3.

review. The NHS Digital DSA is currently time-bound, meaning there is no guarantee that an LPS can hold linked study-NHS Digital data indefinitely.

- 2.3 The suggested models (except Model 1) rely on additional pre-requisite conditions, however these pre-requisites differ by the identified model (see Annex 2 for model 2; and, Annex 3 for models 3-5) and on establishing a suitable technical-governance infrastructure. The technical-governance infrastructure will need to accommodate the NHS Digital contractual requirements; be sufficiently flexible and adaptable to meet the nuances of different LPS requirements; be affordable and sustainable; enable efficient research access; and be acceptable to key stakeholders (e.g. public/participants, LPS funders)⁷.
- 2.4 There are five distinct high-level models in the framework. The first of these relates to sharing anonymous data, and typically will only relate to statistical outputs rather than individual-level data. The second suggests a lighter touch governance arrangement for onward-sharing where the data to be shared can be rendered ‘anonymised’ using controls applied to both the content and context of the data. The remaining three models rely on introducing new controls to counter the risks relating to sharing anonymised individual-level data and ensuring this flow is acceptable and transparent.
- 2.5 Following the stakeholder consultation, the models will be refined into a core set of acceptable processes that will enable onward-sharing of linked NHS Digital records. NHS Digital will issue clear guidance to applicants on how to incorporate these processes into their DSAs and CLOSER will promote the potential for these routes to the LPS and population data science communities. The guidance will need to inform study and institution considerations as to which is the most appropriate option(s) for the context of the specific LPS.

⁷ While the majority of LPS will already have appropriate mechanisms in place, some will need to invest resources to meet these requirements and some recipient institutions will also need to invest in order to meet these requirements. Where the recipient institution is not a regular user of NHS Digital data this may introduce a barrier to data sharing or the use of some of the recommended models.

Recommendation 1: For **NHS Digital** to revise their guidance to data applicants to make clear the potential routes for onward-sharing; and for CLOSER to promote these routes to the LPS and wider data science community.

2.6 It is likely that LPS will adopt only one or two of the framework onward-sharing models, given that their needs are broadly similar and that most LPS tend to have a broadly similar operating context. There is therefore the potential for LPS to jointly develop and implement study-led mechanisms to meet these pre-requisite requirements (e.g. LPS could operate a joint mechanism to review applications for linked data, and thus provide mechanisms for dynamically adding new research investigations. Alternatively, LPS could develop common fair processing materials, such as websites detailing use of information across LPS, or standardised text to explain LPS specific issues). These mechanisms could be implemented in conjunction with organisations, such as CLOSER, or as part of the new Wellcome-MRC-ESRC initiative considering a Population Research Resource.

Recommendation 2: For the **LPS community** to consider whether they are likely to adopt a common onward-sharing model (from the options identified within this report), and to consider whether there are efficiencies to be gained from addressing some development and ongoing tasks at a community level rather than study level.

2.7 Each mechanism will have possible permutations depending on whether the data are stored within a facility owned and operated by the LPS host institution or where the data are stored by a third-party software service provider, and also depending on whether data are directly shared or analysed within a secure research setting.

3. Data Sharing within NHS Digital and within LPS: setting the scope of ‘onward-sharing’

- 3.1 It is important to consider the context in which data are managed, processed and shared in order to identify robust frameworks for onward-sharing. The following considerations are significant and impact on sharing models.
- 3.2 Both NHS Digital and LPS share data with different *ownership* considerations:
- 1) their own data only in its original form,
 - 2) their own data in a processed form (e.g. a derived variable),
 - 3) their own data linked with data owned by a different organisation.
- 3.3 Both NHS Digital and LPS share data in various states of *identifiability*:
- 1) identifiable and containing direct identifiers⁸ (considered Personal Data in the Data Protection Act 2018);
 - 2) pseudonymised and where direct identifiers have been replaced with a pseudonym identifier (again, considered Personal Data in the Data Protection Act 2018⁹);
 - 3) pseudonymised and where the data are contained within a ‘wrapper’ of data and technical and social controls sufficient to ethically and legally treat the data as ‘anonymised’¹⁰ whilst under the protection of those controls, despite retaining some theoretical potential to be related back to an individual (not considered Personal Data while protected by the wrappers);

⁸ Including name, full address, date of birth, NHS ID or other well-used system ID. Other direct identifiers could include pictures of individuals and – since the introduction of the EU General Data Protection Regulations – mobile device and internet connection details and genomic sequence information.

⁹ Mourby M, Mackey E, Elliot M, Gowans H, Wallace SE, Bell J, Smith H, Aidinlis S, Kaye J. Are ‘pseudonymised’ data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*. 2018 Apr 1;34(2):222-33.

¹⁰ Also known as ‘Anonymised in context’. Both terms refer to data that has been transformed to the point that the risk that it will be related back to an identifiable individual is so low that the data can be treated as being anonymous (i.e. not Personal Data). The reference standard for achieving this is set out in the Information Commissioner’s Office Anonymisation Code of Practice.

- 4) and, anonymous data where the subjects cannot be re-identified (not Personal Data)¹¹.

This report uses the definition of Personal Data as provided within the Data Protection Act 2018.

3.4 Onward-flowing data will take different *forms*:

- 1) raw data which is shared in an unchanged form;
- 2) a newly created derived value that is the product of multiple data sources (e.g. NHS Digital data and LPS data);
- 3) the triangulation of multiple data sources (e.g. multiple sources of health data, linked LPS data, linked social administrative records), with provenance information that provides an overarching value with the highest probability of representing the true real world value;
- 4) the products of algorithms based on one or more sources of data;
- 5) and, statistical outputs and aggregated data presented at a population rather than individual-level.

3.5 The recipient of the data will take different *roles*:

- 1) they can be an active or passive user of non-Personal Data;
- 2) they can take the role of Data Processor and use the data for defined and controlled purposes under the direction of the Data Controller;
- 3) or, they can take the role of Data Controller and determine (either alone or jointly with other Data Controllers) how the data will be processed and used.

3.6 Different data sets will have different degrees of sensitivity and the subjects within the data may have properties which increase the risk of their information being

¹¹ Typically anonymous data will be aggregated information contained in tables and statistical outputs. While individual-level data can be anonymous (e.g. a file containing a single variable recording smoking status as a binary value for every citizen in the country), most files used within longitudinal research contain sufficient information to have at least some potential for disclosure

disclosed or the motivations for their information to be attacked¹². While all health information is deemed sensitive (a special category of data), it is acknowledged that some health information is more sensitive than others.

- 3.7 Different permutations of *ownership, identifiability, form, role* and *sensitivity* will generate different considerations as to the risks related to the data sharing and which safeguards and oversight considerations are necessary when data flows from one organisation to another or is released into the public domain. Following the introduction of the EU General Data Protection Regulations (GDPR), LPS must consider these risks formally through conducting Data Protection Impact Assessments (DPIA). The DPIA process will consider the context of the proposed data sharing, identify risks and identify safeguards to control risk. These safeguards extend from involving no additional requirements (e.g. for an approved researcher to publish anonymous statistical findings into a peer-reviewed publication), to requiring additional requirements (e.g. for a databank sharing pseudonymised individual-level data linked to LPS data with a research group based at a different institution). This report seeks to define these additional requirements and the circumstances in which they are used.
- 3.8 LPS commonly request access to NHS Digital ‘Flagging and Tracing’ data which provides updated personal identifiers, address details, General Practice registration and fact of death of their participants. These data are used to re-contact participants, to contact their General Practice in order to seek assent for primary care record linkage and to update study databases. These data may be used by contractors (e.g. survey fieldworkers). Sharing these data with contractors must be fully described in the LPS Data Sharing Agreement with NHS Digital but does not constitute ‘onward-sharing’, given that the contractors will take the role of Data Processor working under contract to the LPS (who will be Data Controller). It is therefore out of scope for this report.

¹² See the UK Anonymisation Network’s anonymisation decision making framework for more information. Elliot M, Mackey E, O’Hara K, Tudor C. The anonymisation decision-making framework. UKAN; 2016.

4. Different approaches to how data are shared and accessed during research

- 4.1 Researchers can use data within a secure facility where they are not able to remove the data or can be provided with copies of data to use in their own facility. These distinct options have been described using the ‘reading library and lending library’ analogy¹³. Controlling the movement and potential proliferation of data by using a ‘reading library’ model reduces the risk relating to data sharing.
- 4.2 Research infrastructures such as the Office for National Statistics Micro Data Laboratory are long established examples of the ‘reading library’ model and have been used for research on health data (e.g. cancer registration records). However, these infrastructures have traditionally been restrictive and acted as ‘silos’ which – given they do not permit the input of external data - prevent the linkage of different sources of information and therefore limit research potential.
- 4.3 The ‘reading library’ model can be extended to allow linkage to new data sources (i.e. to move away from the ‘silo’ of a secure server only hosting information from one owner). The linkage and addition of new records would be conducted by those running the server (e.g. the LPS or a trusted third party), not the researcher. This potential for sharing health records has been recognised in academic thinking (e.g. the Data Safe Haven model¹⁴) and in the development of infrastructure, such as UK Secure eResearch Platform (UKSeRP)¹⁵ or the Clinical Practice Research Datalink (CPRD)¹⁶ which provides functionality to enable this.
- 4.4 The ICO have given some considerations that personal data accessed using ‘reading library’ models still constitutes a dissemination of Personal Data. As such the ‘reading library’ model alone does not constitute a solution to onward-sharing,

¹³ Davies SC. Annual report of the chief medical officer 2016, generation genome.

¹⁴ Burton PR, Murtagh MJ, Boyd A, Williams JB, Dove ES, Wallace SE, Tasse AM, Little J, Chisholm RL, Gaye A, Hveem K. Data Safe Havens in health research and healthcare. *Bioinformatics*. 2015 Jun 25;31(20):3241-8.

¹⁵ Cornish RP, John A, Boyd A, Tilling K, Macleod J. Defining adolescent common mental disorders using electronic primary care data: a comparison with outcomes measured using the CIS-R. *BMJ open*. 2016 Dec 1;6(12):e013167.

¹⁶ Cornish RP, Henderson J, Boyd AW, Granell R, Van Staa T, Macleod J. Validating childhood asthma in an epidemiological study using linked electronic patient records. *BMJ open*. 2014 Apr 1;4(4):e005345.

although it may provide a useful control which contributes to the solution (i.e. a framework to support the effective anonymisation of data).

- 4.5 NHS Digital recognises the value of ‘reading library’ controls and have approved the use of these in LPS (e.g. the ALSPAC Data Sharing Agreement which includes the use of UKSeRP¹⁷). The use and details of the ‘reading library’ needs to be explicit in the DSA and their needs to be an appropriate legal basis to support this.
- 4.6 It was noted that the location of the user is important in this situation. For example, a researcher accessing the ‘reading library’ remotely from the USA constitutes a dissemination of the data to the USA, even though the data remain on a server in the UK. It is not permitted to disseminate data to all countries, given that local legislation may not offer sufficient safeguards or that other local legislation may compromise the confidentiality and security of the data (e.g. the US Patriot Act). It is not currently clear as to which countries - beyond those in the European Economic Area – that studies can disseminate data to; and it is not clear which safeguards (e.g. secure ‘reading libraries’) mitigate which risks.
- 4.7 NHS Digital recognises the collaborative nature of research and the potential for a researcher(s) from one institution to work with the LPS host institution under the basis of an honorary contract. Sharing linked NHS Digital and LPS data with a researcher under these conditions is not considered onward-sharing if the outputs of the research¹⁸ are owned by the host institution and that the honorary contract contains robust confidentiality clauses and meaningful penalties. This model is not within the scope of this report.

Recommendation 3: For **NHS Digital** to issue practical guidance in relation to disseminating data beyond the European Economic Area in line with ICO guidance.

¹⁷ Boyd A. NHS Digital: Precedents Set. London: CLOSER, 2018.

¹⁸ To include: derived data, intellectual property, copyrights and attributions on published outputs.

5. Derived Data

- 5.1 The NHS Digital DSFC recognises¹⁹ that users of NHS Digital supplied data may create derived data, and that ownership and intellectual property of these – and as such Data Controller status – lies with the study and their host institution, and not NHS Digital.
- 5.2 Schedule 1 of the DSFC sets out the definition of when processed data can be considered ‘derived data’ (see Panel 1). NHS Digital require (Section 3.10 of the DSFC) that LPS retain a register of the data they have derived and that NHS Digital reserve the right to assess whether the processed data meets the threshold set out in the definition. Furthermore, NHS Digital reserve (Section 3.9 of the DSFC) the right to apply ‘Special Conditions’ regarding the copying, processing, distributing, manipulating, creating, storing and any other use of the Derived Data.
- 5.2 Consideration will have to be made of the intellectual property inherent in data derivation. Once a derivation is added to the NHS Digital ‘derivation register’, and this derivation can be achieved through NHS Digital data alone, then NHS Digital will from that point on create the derived value for subsequent users and ownership would not pass to any study/data user.

¹⁹ “All Derived Data and any and all Intellectual Property Rights in Derived Data shall be owned exclusively by the Data Recipient. In consideration for the grant of the license set out at Clause 3.1, the Data Recipient agrees to be bound by any **terms set out in the Special Conditions** regarding the copying, processing, Distributing, Manipulating, creating, storing and any other use of the Derived Data”. Taken from Section 3.9 of the Data Sharing Framework Contract.

Panel 1: Definition of the terms ‘Data’, ‘Derived Data’ and ‘Manipulate’, taken from Schedule 1 of the NHS Digital Data Sharing Framework Contract.

Data [are] the health or social care data specified in and that is provided by NHS Digital to the Data Recipient under a Data Sharing Agreement; and such term shall also be deemed to include Manipulated Data unless otherwise specified;

Derived Data [are] any Data (wholly or in part) that is Manipulated to such a degree that it:

- a. cannot be identified as originating or deriving from the Data and cannot be reverse-engineered such that it can be so identified; and
- b. is not capable of use as a substitute for the Data; and
- c. has not at any time been verified by NHS Digital as not fulfilling the criteria (a) and (b) above;

Manipulate means:

- combine (wholly or in part) with other data or information; or
- aggregate (wholly or in part) with other data or information; or
- adapt (wholly or in part);

- 5.4 Derived data are not necessarily anonymous. Rather, the processing and considerations that generates derived data and de-identified or anonymizes data (anonymized in context) are different tasks.
- 5.5 It is currently not clear to many studies as to where the threshold at which data can be defined and treated as being derived data lies. Clearer guidance is needed using applied exemplars across the different classes of data supplied by NHS Digital (e.g. current status tracing information, care pathways, disease status, fact of death).

Recommendation 4: For **NHS Digital** to revise their guidance on when data processing meets the threshold of producing ‘derived data’. For this guidance to be illustrated with exemplars based on diverse data types and processing contexts; and for CLOSER, other LPS, and studies with other research designs (e.g. RCTs) to help nominate these examples.

- 5.6 Given that the LPS (or rather, their host institution) will be Data Controller for the derived data, then the LPS may determine to whom the data can be shared, the purpose for which it can be used and the safeguards/conditions attached. The approach to using derived data may vary depending on the nature of the data in question and must link back to any Special Conditions applied by NHS Digital. This is not considered to form a broad model for onward-sharing, given that it will only apply to a subset of data.

6. National Opt-Out

- 6.1 The national data opt-out was introduced on 25 May 2018, enabling patients to opt out from the use of their data for research or planning purposes, in line with the recommendations of the National Data Guardian in her Review of Data Security, Consent and Opt-Outs.
- 6.2 National data opt-out status is centralized and held by NHS Digital. LPS would need an efficient mechanism to access current opt-out status in order to implement this in an onward-sharing process in the future. This process is challenging due to Opt-Out status being ‘sensitive’ and not for sharing in its own right.
- 6.3 The current guidance and mechanism for how LPS should apply national opt-out is not clear. There is a need for clearer guidance on when to apply opt-out (i.e. at the start of every new project) and how frequently opt-out status is refreshed.

Recommendation 5: For **NHS Digital** to consider the sensitivity of ‘National Opt-Out’ status flags and whether these can be shared (under controlled conditions) in order for research users to apply them.

7. Cost Recovery

- 7.1 NHS Digital operate on a cost recovery basis. They currently charge £10,000 per annum to enable onward-sharing in an LPS²⁰. This has the potential to add a burden to LPS which – in addition to LPS staff time required to manage this – would need to be passed on to either the research funders or to data users. It also has the potential to act as a barrier to enable onward-sharing in all LPS (e.g. it may not be viable to recoup costs in a small LPS with low numbers of users).
- 7.2 Different onward-sharing models will have different associated maintenance costs, i.e. a light touch onward-sharing model should require less input from NHS Digital and thus generate lower costs. It should be considered whether a flat cost recovery fee is appropriate, or whether the fee should vary across the models.
- 7.3 NHS Digital cost recovery charges do not apply to the onward-sharing of derived manipulated data.

Recommendation 6: For **NHS Digital** to consider whether the cost recovery model should vary across onward-sharing models in line with its costing review.

8. Model 1: Anonymous data sharing

- 8.1 There is an existing principle that anonymous NHS Digital data can be onwardly-shared without the need for additional oversight, for example anonymous research findings are ‘onwardly-shared’ into the public domain in peer reviewed journals.
- 8.2 This approach has a sound legal basis as anonymous data are not Personal Data (i.e. they do not relate to an identifiable individual), and therefore the terms of the Data Protection Act (2018) and Common Law Duty of Confidentiality do not apply.
- 8.3 While it is possible for individual-level data to be anonymous, generally rich individual-level data used in longitudinal investigations are likely to retain the potential for being re-identified²¹.

²⁰ <https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-charges-2018-19>

²¹ Elliot M, Mackey E, O'Hara K, Tudor C. The anonymisation decision-making framework. UKAN; 2016.

- 8.4 For this reason, Data Controllers attempting to balance disclosure risk with data utility will not seek to render the individual data *themselves* anonymous (i.e. to change the content of the data). Rather, Data Controllers will tend to try and achieve the same benefits through rendering the data effectively anonymised by applying sufficient socio-technical controls to ensure that, taking account of all means reasonably likely, that individuals accessing the data would not be able to identify the individuals (i.e. to apply controls to the context in which the data are used).

9. Model 2: Anonymised data sharing (also referred to as ‘effectively anonymised’, ‘anonymised in context’)

- 9.1 The principle that anonymous data can be shared can be extended to sharing individual-level data if the risk of identification can be controlled to the point where the user does not have the “means reasonably likely” to re-identify the data subjects. In these circumstances linked NHS Digital and LPS data would not be considered Personal Data and could be treated, while the controls were applied, as being anonymous. The ICO anonymisation code of practice²² establishes a framework to achieve this, which is subsequently elaborated on in the UK Anonymisation Network guidance²³.
- 9.2 A range of data processing, technical and social controls are needed for data to be considered ‘anonymised’ (see ICO code of practice). Example controls include effective pseudonymisation, secure research servers (i.e. the ‘reading library’ approach), data sharing contracts, user training, and penalties for misuse. The requirements extend beyond using a ‘reading library’ approach with no other controls. Many of these principles are found within existing data sharing models (see Figure 1) developed by LPS operating Data Safe Haven approaches. These can include the use of local secure research servers (managed by the host institution)

²² Information Commissioner’s Office. Anonymisation: managing data protection risk code of practice. UK. 2014-04-10. <https://ico.org.uk/media/1061/anonymisation-code.pdf>. 2012.

²³ Ibid. 21.

or national platforms including UKSeRP (University of Swansea) and UK Data Service (University of Essex).

- 9.3 It is important to note, for those seeking to interpret these models, that data can be considered ‘anonymised’ to one user (e.g. the researcher) and identifiable to another (e.g. the LPS data managers) at the same time. This principle has been clearly articulated in legal judgements²⁴ and is reflected in the ICO guidance²⁵.
- 9.4 For this model to be acceptable to NHS Digital, studies would need to clearly describe the onward-sharing processes within their DSA and to demonstrate an appropriate legal basis for this and they would need to robustly demonstrate how they were meeting the requirements of the ICO code of practice. NHS Digital would not need to determine each new use case of the data (given they are no longer Personal Data). This means that the LPS DSA would articulate a broad case for using the data and the means (setting out the precise rules) by which new use cases were considered, approved and managed; but studies would not need to seek specific approval from NHS Digital for new users or new projects.

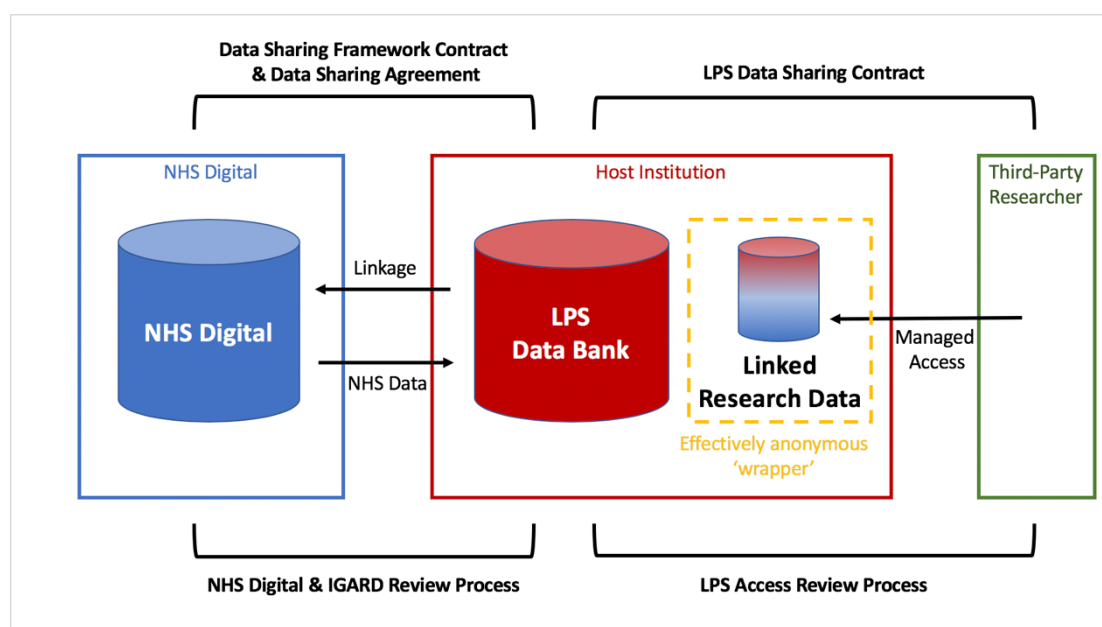


Figure 1: Onward-sharing of ‘anonymised’ linked data

²⁴ Department of Health v Information Commissioner [2011] EWHC 1430 (Admin)

²⁵ Ibid 22.

- 9.5 Sub-licences are still required between the LPS host institution and any external institution (as such a licence is a specified control within the ICO code of practice). One sub-licence is needed per institution, and new users and use cases can be agreed by the LPS host institution (subject to the appropriate checks taking place, as described above). The NHS Digital sub-licence cost-recovery models will apply to the LPS institution only.
- 9.6 The feasibility of establishing a framework for onward-sharing ‘anonymised’ information is currently restricted as the regulatory code of practice defining this area has not yet been updated to the Data Protection Act 2018. The guidance is currently being updated, and therefore there is a risk that the anonymisation process may not meet any new guidance. This can be mitigated by considering the ICO’s and other data controllers’ current positions. Here, the regulator has indicated that the current code of practice is ‘a good starting point’ for considering anonymisation in context under the new Act²⁶. Any applications developed based on this model can only be provisional until the code of practice has been updated. To explore this model the following is based on the current ICO code of practice.
- 9.7 There is potential that this approach could provide a gateway to enable the pooling of NHS and potentially other data gathered from across the UK. There should be further consideration of this potential and the safeguards that would be needed to make this acceptable once the ICO guidance is updated (this lies outside the scope of this project).
- 9.8 It was noted in the consultation workshops that anonymous data has a different standing in Freedom of Information (FOI) rules, where anonymous data is subject to FOI requests, but Personal Data is not. However, it is important to note that anonymised data can only be considered such while within the ‘wrappers’ of its protective controls. Once outside of these controls, the data will revert to being Personal Data and therefore outside the scope of FOI.

²⁶ <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/>

Recommendation 7: as the ICO update the ‘Anonymisation: managing data protection risk code of practice’, for the **LPS community** and **NHS Digital** to ensure that guidance on key points related to onward-sharing and the required controls are in alignment.

Recommendation 8: For the **LPS community** to consider and consult on the potential for the ‘anonymised’ model to form a gateway for the sharing and processing of health data, and potentially other data, from across the UK Home Nations and to promote this option if it is seen as being valid and acceptable.

10. Model 3: Sub-licence model

10.1 This is an existing model, but used by only a few organisations (UK Biobank, Clinical Practice Research Datalink (CPRD)), where NHS Digital shares data with organisation *A*, who are in turn licensed to share data with other organisations *N*, subject to agreed controls. This is permitted where *A* adds additional value to these data in a manner that cannot be achieved by NHS Digital (e.g. linking LPS data with routine health records). It is required that the data sharing controls in place between NHS Digital and *A*, are replicated between *A* and *N* (see Figure 2).

Examples of this include:

- 1) NHS Digital require the ability to directly audit organisation *N*’s compliance with these controls;
- 2) Organisation *N* will need to be using the data both for the benefit of health and social care, and within the scope of data use agreed between NHS Digital and organisation *A*.

The advantage of this model relates to the flexibility offered to *A* (i.e. it is similar to existing study data sharing models). It should be noted however that NHS Digital only has an agreement in place with *A*, and therefore *A* is accountable for *N*’s actions. Model 3 permits organisation *A* to provide linked data to organisation *N* through physical transfer (e.g. UK Biobank) or through requiring users to conduct analysis within a physical ‘safe haven’ (e.g. CPRD).

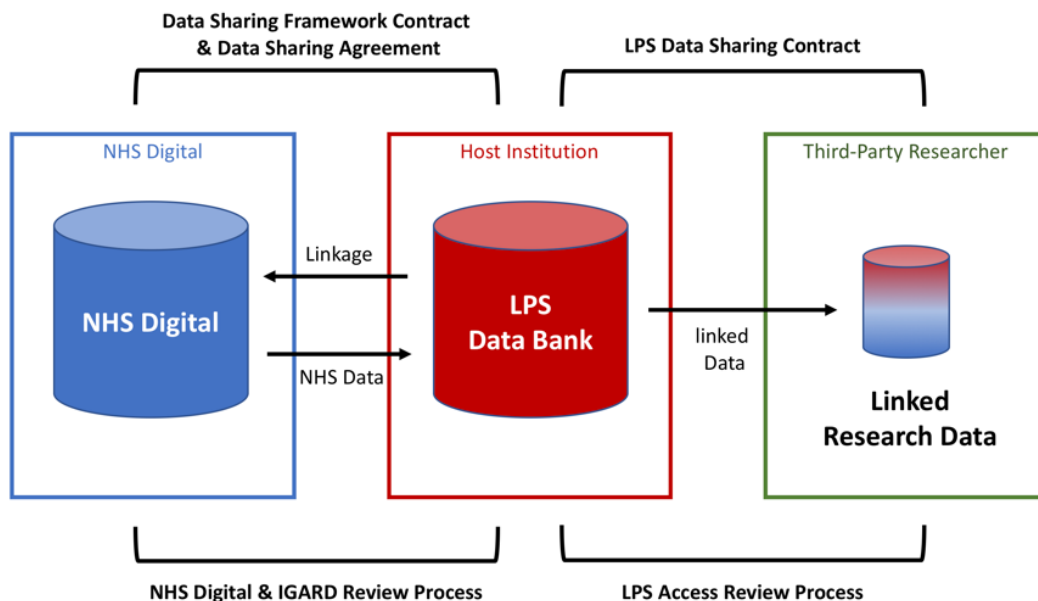


Figure 2: Sub-licence model

- 10.2 Model 3 can be extended to include secure research environments (see Figure 3) provided by third party infrastructure providers (e.g. UKSeRP or UK Data Service). This model reduces organisation A's risk through requiring organisation N to access data in a managed and audited manner which does not involve the physical transfer or proliferation of data. It is a requirement of NHS Digital that an onward-sharing arrangement can only involve three Data Controllers (i.e. NHS Digital – Organisation A – Organisation N) which implies that the third-party infrastructure provider operating in Model 3 must do so in the role of Data Processor. The inclusion of a secure research environment does not fundamentally change the principles of Model 3, however, making the substantial investment to maintain a secure research environment may make Model 2 (the anonymised model) more feasible.

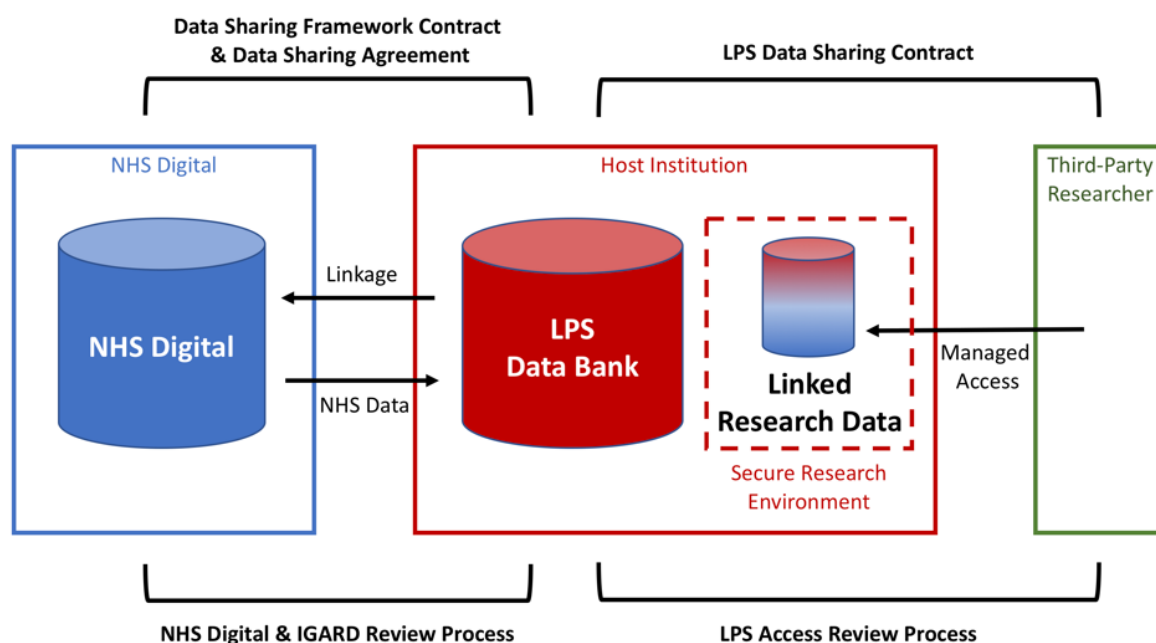


Figure 3: Sub-licence model using secure research environment

11. Model 4: Multiple agreement model

- 11.1 In this model, NHS Digital enters into a data sharing agreement with organisations *A* and *N* separately. Additionally, *A* enters into a separate agreement with *N*, in relation to the sharing and use of *A*'s data. Subject to usual approvals, NHS Digital data can then be shared with *N*, and *A* can provide study data with *N* (see Figure 4). Bridging files will be shared to enable the data to be combined²⁷. The benefit of this model is that NHS Digital retains direct agreements (and accountability) with all organisations using their data. The disadvantages relate to the burden of multiple agreements and related processing requirements. This additional burden will impact mainly on NHS Digital and the research users, and may act as a deterrent to data use.

²⁷ Currently, studies linking to NHS Digital records use a study participant reference ID as a reference number ('study ID'). In this model, both the study and NHS Digital would agree to transform this study ID at a project specific level (e.g. using a hash encryption algorithm with a project specific 'seed' number). This number could then be used on both data files and act as a linking 'bridge'.

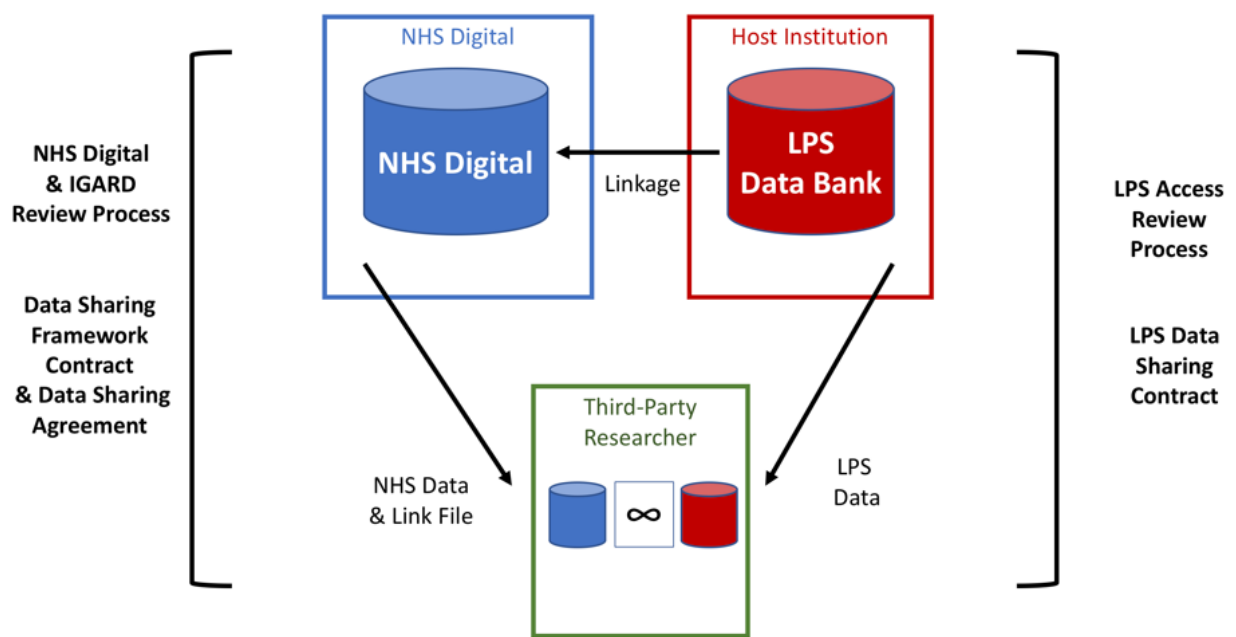


Figure 4: Multiple agreement model

- 11.2 Model 4 could be adapted so that LPS A initially receives the linked NHS Digital data which they subsequently process to adapt the data to the needs of the longitudinal research community (see Figure 5). LPS A will then provide both the LPS and the linked, post-processed, NHS Digital data to *N* once NHS Digital confirms that *N* had been approved and had entered into a DSFC and DSA.

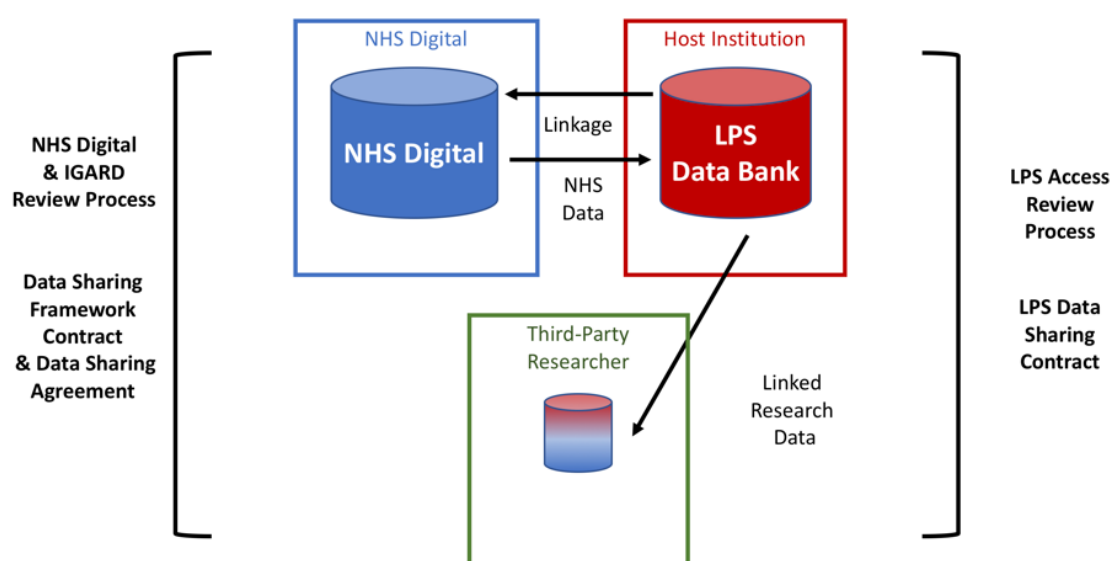


Figure 5: Multiple agreement model with LPS conducted processing of linked NHS Digital data

- 11.3 A secure research server could be used to host the linked data described in either 12.1 or 12.2.
- 11.4 The emerging NHS Digital Data Services Platform (DSP) could be used to accommodate and facilitate this data-sharing model. In this scenario, the DSP could host LPS data and provide it pre-linked with NHS Digital Data to research users.

12. Model 5: Trusted Third-Party model

- 12.1 In this model, LPS data and NHS Digital data are linked and managed by a trusted third-party who manages onward-sharing on behalf of an LPS (see Figure 6). This model is distinct from the ‘third-party’ arrangements described earlier (in Model 3 or Model 4) as the third-party will operate as Data Controller²⁸ where the data sharing contract(s) are established between NHS Digital and the third-party provider (i.e. NHS Digital – third-party – Organisation N). There will be an additional contract between the LPS and the third-party infrastructure provider which sets the terms under which the LPS data can be processed and managed.

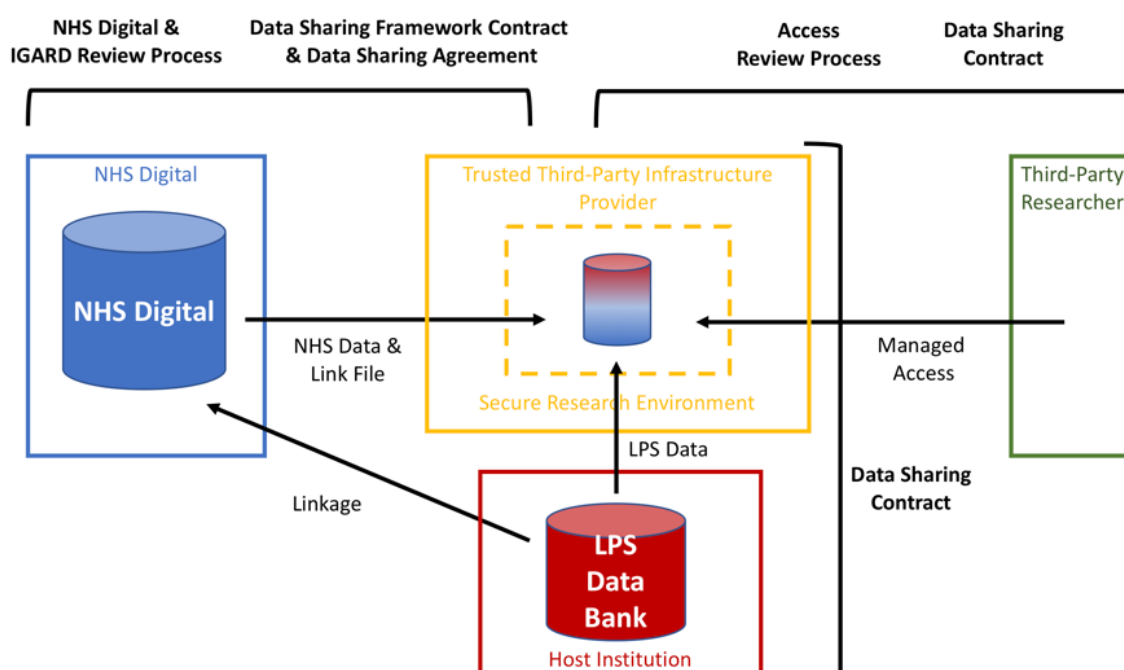


Figure 6: Trusted third-party model

²⁸ i.e. will make day-to-day operational decisions as to how the data are processed and managed: typically following the terms and conditions of a contract from the LPS (organisation) who ultimately own the data.

- 12.2 A potential variation to this model could see NHS Digital acting as both Data Controller of the centralized health record, and also acting in the role of trusted third-party (although not meeting the true definition of this term) and providing linked LPS and NHS Digital data to research users. This model is likely to take advantage of the NHS Digital DSP as the infrastructure in which the data are linked and shared. In addition to the DSFC and DSA, there would need to be a contract between the LPS and NHS Digital which sets the terms under which the LPS data can be processed and managed.

13. Cross-LPS studies

- 13.1 Cohort and other LPS are increasingly operating in consortia or in cross-LPS research designs where information is used from multiple studies. Within these designs the analysis takes two distinct forms: firstly ‘meta-analysis’ studies, where information from each study is analysed separately in discrete assessments and where statistical outcomes from different studies are compared and combined into a single estimate; and, secondly, ‘pooled analysis’ studies where the data are combined into a single location and jointly analysed.
- 13.2 Meta-analysis studies can be managed in a similar manner to the arrangements described above, as each study-level assessment is conducted in isolation, i.e. this model does not demand Personal Information. Care will need to be taken when sharing statistical outcomes as some outcomes are disclosive (e.g. it is possible to recover underlying information when sharing regression co-variance matrices). These will need treating as either Personal Data or will need further controls in order to treat them as effectively anonymised information.
- 13.3 Pooled studies will require data sharing, as the model requires minimum necessary data from all studies to be centralized into a single location for harmonization and analysis. In this situation, the studies will need to identify the most suitable data sharing arrangement from the models described above. Care will have to be taken when determining which parties are Data Controllers and which parties are Data Processors.

- 13.4 While there are many LPS within the UK, there is the potential that clusters of studies – which are ultimately controlled by a smaller group of institutions – may comprise a disproportionate number of cross-cohort studies. If this proves to be the case this may influence the choice of sharing model (e.g. the reduced institutional liability within Model 4 may be prioritized given efficiencies of reusing the investment in establishing the model over multiple investigations).
- 13.5 Privacy preserving systems such as DataSHIELD²⁹ and VIPER³⁰ have been developed to enable anonymized cross-cohort pooled-analysis. There is potential for privacy-preserving technologies such as these to be deployed as additional controls and to provide a mechanism to support international cross-cohort studies.

14. A decision-making framework for onward-sharing

- 14.1 The most appropriate model for onward-sharing will be specific to the context of the LPS and the proposed data sharing. It is therefore important to consider what would comprise an onward-sharing decision-making framework.
- 14.2 LPS will need to consider the project specific context through conducting DPIAs and also the broader context through aligning with stakeholder principles (e.g. the ‘Fair’ principles³¹, the ‘Safe’ principles³² and the principles set out in the Global Alliance for Genomics and Health Framework for Responsible Sharing of Genomic and Health-Related Data³³) and through engagement with study participants and the wider public.

²⁹ Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, LaFlamme P, Tobin MD, Macleod J, Little J, Fortier I. DataSHIELD: resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data. *International journal of epidemiology*. 2010 Jul 14;39(5):1372-82.

³⁰ Carter KW, Francis RW, Carter KW, Francis RW, Bresnahan M, Gissler M, Grønberg TK, Gross R, Gunnes N, Hammond G, Hornig M. ViPAR: a software platform for the Virtual Pooling and Analysis of Research Data. *International journal of epidemiology*. 2015 Oct 8;45(2):408-16.

³¹ Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*. 2016;3.

³² Desai T, Ritchie F, Welpton R. Five Safes: Designing data access for research.

³³ <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/#fp>

- 14.3 LPS will then need to consider the resources they have at their disposal, e.g. the availability of a secure research server needed to operate a ‘reading library’ model, and whether the deployment of all possible infrastructure is proportionate to the risk. When considering proportionality, it will be necessary to consider the need to provide information in a timely manner with a sustainable cost model which is aligned with the considerations of the research funders and does not introduce a barrier to shared data use.
- 14.4 Strong consideration will need to be made to how to control disclosure risk in the requested data. An appropriate balance will need to be made between controlling risk and preserving data utility. Given that the institution who controls the LPS will ultimately take much of the liability it is likely that they will take decisions on this balance in consultation with their LPS and regulatory/legal advisors. It is highly unlikely that rich individual-level data could ever be processed to the point where they are truly anonymous while retaining research utility. Therefore the adopted controls and sharing models will need to be suitable for the exchange of de-identified data (i.e. Model 1 is unlikely to ever be used for the sharing of individual-level data). The UKAN Framework³⁴ provides a key resource for assessing disclosure risk.
- 14.5 LPS will need to gauge the ‘risk tolerance’ of both their institution (who are liable for penalties resulting from data misuse or breaches) and their participants (whose data are being shared). Risk tolerance, along with the data context and study financial resources, is likely to inform the decision as to whether to use a ‘lending library’ or ‘reading library’ approach.
- 14.6 For LPS who are likely to only have one or two key onward-sharing flows, which are potentially sustained over many iterations of flows and long periods of time (e.g. an LPS run by one institution in partnership with another), Model 4 may prove to be the best solution, given this has a relatively high cost to establish but shares risk and accountability across the institutions.

³⁴ Ibid. 21.

- 14.7 For LPS who are likely to have many data sharing flows with many institutions, then Model 2 and Model 3 may be most suitable. It is possible to adopt multiple models e.g. it may be desirable in some cases to have Model 3 in place with a ‘heavy user’ of the data (e.g. to share data to a centre or unit located in a different institution), while using the more light touch approach found in Model 2 with institutions making one-off or occasional use of the data.
- 14.8 LPS will need to provide fair processing information to participants (and the wider public) to help ensure the data share is transparent and fair, and to drive wider support for health data science.

15. Alignment with other providers of health records

- 15.1 LPS are seeking to link to routine health and social care records held by a range of health providers. These range from national providers (e.g. Public Health England), regional providers (e.g. the community mental health care providers who are part of the UK-CRIS network), specific datasets held at a Trust level, or focused resources (e.g. Clinical Practice Research Datalink).
- 15.2 An efficient framework for onward-sharing implemented by NHS Digital would ideally be replicated across sharing arrangements with other health record data owners (e.g. Public Health England).

Recommendation 9: For the findings within this report to be disseminated to all UK providers of health care records for public good research.

16. Conclusions

- 16.1 This report sets out five distinct models for the onward-sharing of linked NHS records within longitudinal research studies operating as databanks. It provides an outline framework for applying these models.
- 16.2 It is anticipated that the most likely routes to be used by LPS for onward-sharing are Model 2 to the data ‘effectively anonymous’, and Model 3 to share under a sub-licence agreement.

- 16.3 The report makes a series of recommendations. Those required to enable Model 2 and Model 3 should be prioritized for discussion at the NHS Digital RAG.
- 16.4 LPS should work together as a community to develop standardised tools for efficient onward-sharing that is both efficient for LPS to deploy, and familiar and standardised for external users. CLOSER, or potentially a future Population Research Resource, could provide support or a forum for this work.
- 16.5 Both LPS and NHS Digital will need to clearly describe these activities to participants, and the public, to help ensure that onward-sharing is transparent and fair.

Annex 1: Summary of timeline of tasks for NHS Digital and LPS Community

Task	Timeline
For NHS Digital to continue processing applications and application amendments to include onward-sharing provisions. For precedents to be publicised through NHS Digital DARS and CLOSER.	Ongoing
Disseminate the report insights through:	
• NHS Digital DARS website	Qtr4/2019
• CLOSER social media channels	Qtr4/2019
• Academic paper	Qtr1/2020
• Academic conference	Qtr4/2019
For CLOSER to discuss with the joint funders Population Research Resource panel as to how to encourage coordinated approaches to developing standardised solutions across LPS community.	Qtr1/2020

Annex 2: NHS Digital Pre-Requisite Conditions (for Model 2)

- A2.1 **Pre-requisite 1:** The LPS's host organisation must have a DSFC in place and the LPS must have a DSA in place which specifies the mechanism for onward-sharing.
- A2.2 **Pre-requisite 2:** The LPS's legal basis must clearly describe the intention to onwardly-share data with third-party research users (i.e. fair processing materials must make the intention to onwardly-share data clear; Section 251 applications must make the intention to onwardly-share clear)³⁵.
- A2.3 **Pre-requisite 3:** The LPS must make careful consideration of 'anonymisation in context' requirements. A clear legal basis is needed at each stage of the process, and moving from one legal basis for identifiable data to another legal basis for anonymised in context data requires robust justification against the criteria in the pending update to the ICO Anonymisation Code of Practice.
- A2.4 **Pre-requisite 4:** The researchers must demonstrate their own, their projects' and their institutions' 'Bona fide' or 'Safe' credentials (i.e. they must demonstrate they are legitimate researchers).
- A2.5 **Pre-requisite 5:** The researchers must demonstrate how their project objectives will benefit the health and social care system and how their project outputs will be disseminated to help realise these benefits.
- A2.6 **Pre-requisite 6:** Fair processing information describing the specific data use must be provided to LPS participants and the general public.
- A2.7 **Pre-requisite 7:** Where an LPS proposes onward-sharing via a third-party infrastructure provider then there must be a formal agreement between the LPS and the infrastructure provider in which the infrastructure must operate in a role of

³⁵ The LPS must have a valid Article 6 and Article 9 basis to meet the Data Protection Act 2018 requirements. In relation to meeting Common Law duties of Confidentiality: where LPS rely on Section 251 as their legal basis then their Section 251 application/approvals must make clear the intention to do this and that the LPS is operating as a research database; and that sufficient fair processing is conducted to ensure this use would be a reasonable expectation (no surprises) of a typical participant. Where consent is used as a legal basis, then the onward-sharing of linked information must be explicitly described in the consent materials. It may be possible to update existing consent which does not describe this explicitly through providing updated fair processing and providing a means to object. Determination as to whether an LPS has a valid legal basis is subject to change as the Law changes and as new Codes of Practice are issued.

contracted Data Processor rather than in a Data Controller role; or where the infrastructure provider acts as the hub for the exchange of data, they will act as Data Controller rather than the LPS.

- A2.8 **Pre-requisite 8:** The LPS and, where used the provider of the secure research server, must demonstrate sufficient management of Information Security to the good practice standards required by NHS Digital (demonstrated through ISO27001 certification, NHS Data Security and Protection Toolkit³⁶, or adherence to a sufficiently robust System Level Security Policy).

³⁶ Was the NHS Digital Information Security Toolkit. See: <https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit>

Annex 3: NHS Digital Pre-Requisite Conditions (for Models 3-5)

- A3.1 **Pre-requisite 1:** The LPS's host organisation must have a DSFC in place and the LPS must have a DSA in place which specifies the mechanism for onward-sharing.
- Where Model 4 is used, the third-party data user's institution must have a DSFC and DSA in place.
 - Where Model 5 is used, the trusted third-party infrastructure provider must have a DSFC and DSA in place.
- A3.2 **Pre-requisite 2:** The LPS's legal basis must clearly describe the intention to onwardly-share data with third-party research users (i.e. fair processing materials must make the intention to onwardly-share data clear; Section 251 applications must make the intention to onwardly-share clear)³⁷.
- A3.3 **Pre-requisite 3:** The researchers must demonstrate their own, their projects' and their institutions' 'Bona fide' or 'Safe' credentials (i.e. they must demonstrate they are legitimate researchers).
- A3.4 **Pre-requisite 4:** The researchers must demonstrate how their project objectives will benefit the health and social care system and how their project outputs will be disseminated to help realise these benefits.
- A3.5 **Pre-requisite 5:** Fair processing information describing the specific data use must be provided to LPS participants and the general public.
- A3.6 **Pre-requisite 6:** Where an LPS proposes onward-sharing via a third-party infrastructure provider then there must be a formal agreement between the LPS and the infrastructure provider in which the infrastructure provider must operate in a role of contracted Data Processors rather than Data Controller role.
- A3.7 **Pre-requisite 7:** Both the LPS and the third-party researcher(s) must demonstrate sufficient management of Information Security to the good practice standards required by NHS Digital (demonstrated through ISO27001 certification; NHS Data Security and Protection Toolkit³⁸; adherence to a sufficiently robust System Level Security Policy).

³⁷ Ibid 35.

³⁸ Ibid 36.