Medicines & Healthcare products Regulatory Agency

CPRD

Non-Standard Linkage Service Application Form Guidance

These guidance notes are to help users to complete and understand what is required to complete the Non-Standard Linkage Service (NSLS) application form. The NSLS Application Form is used by the CPRD Linkages Working Group to evaluate proposed non-standard linkages, the process of evaluation is described on the CPRD website (cprd.com/non-standard-linkage). The NSLS Application Form is in addition to an ISAC application, with ISAC approval a requirement for any study using CPRD data.

- Each section of the application form should be completed in full. Any section not applicable should be completed as 'Not applicable' and justification provided. Applications with incomplete sections without justification will be returned as invalid;
- If you are a new applicant who has not requested CPRD data before, please download the **'CPRD New client request for access to CPRD Data form'** or email enquiries@cprd.com to request a copy
- This form should be sent in with the NSLS application form to the CPRD linkages mailbox (cprdlinkages@mhra.gov.uk) with the email subject header '**For the attention of Non-Standard Linkage Service'**.
- The application form should be completed in Arial 10pt font only.

Section A: Study details
Question 1: Study Title Please provide a brief title that summarizes your proposed research, avoiding the use of abbreviations, vague or excessively long titles. *Limited to 120 characters*
Question 2: The Chief Investigator The Chief Investigator should be a senior scientist that will take responsibility for ensuring that the research is undertaken with full adherence to CPRD guidelines. Please include the title, full name, job title, affiliation/organisation and email address for correspondence of the Chief investigator on the form.
Question 3: The Corresponding Applicant The name and contact details of a Corresponding Applicant who will be the direct point of contact for the application must also be provided for each application. It is acceptable for the Chief Investigator to be the corresponding applicant. It is not acceptable for the Corresponding Applicant to be a person not involved in the

NHS National Institute for Health Research

study (e.g. personal assistants).

Question 4: Summary of the study (Max. 250 words)

Please provide a succinct overview of your proposed research in non-technical language. The summary should cover the aims of the study as well as an overview of the study design and statistical methods. It should include a clear justification, avoiding jargon, of the expected public health benefits from the study, which must be capable of being understood by a member of the public without a scientific or medical background. Abbreviations should be avoided.

Question 5: Proposed completion date of the study

Provide the proposed completion date of the study. Include the delivery date for the linked dataset from CPRD that would be required in order to meet the given study completion date.

- For area-level linkages based on practice postcode a minimum period of 9 months from initial application to the CPRD Linkages Working Group is advised.
- For individual patient level linkages, a minimum period of 18 months from initial application is advised. The timeframe will be variable due to external dependencies that vary on a case-to-case basis.

Question 6: Does this study involve requesting any additional information from GPs, or contact with patients?

Please contact enquiries@cprd.com to speak to the Interventional Research team **BEFORE** submitting the **Non-Standard Linkage Service application form** to discuss the feasibility of including additional GP or patient information.

Explain the proposed data collection and the legal basis for this collection.

Question 7: Funding source(s) of study

Please specify the funding source(s) for the study. Where applicable, please also specify the institution/organisation name and country of the funding source(s). It is important that the source of funding for a study is made explicit on the application form.

In relation to studies funded from non-health related sectors, the following criteria must be met before they could be approved:

- The research must have a health-related purpose
- The investigators should all be employed in health-related institutions
- Persons employed in other sectors should not be investigators and should not have access to the raw data
- The funding source must be transparent

The findings should be published and their intellectual property rights reside with the applicants.

CPRD follows HM Treasury rules on fees and charges as a publicly funded body and therefore operates on a cost recovery basis. CPRD will recoup the costs of the work associated with facilitating the linkage. In addition, CPRD data has associated access charges that are dependent on the size of the dataset and the linked data provided. CPRD can provide estimated costs for a non-standard linkage by contacting the CPRD linkages mailbox cprdlinkages@mhra.gov.uk. A quote will be provided once an application has been formally agreed by CPRD.

The applicant must also meet any costs from NHS Digital to process the linkage, these can be estimated from: https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-charges-2018-19
Any third-party data controller may also have their own charges which will be the responsibility of the applicant.

Section B: Data Requested

Question 1: Primary Care Data

Vision® and EMIS® are different clinical software systems used by general practices in the United Kingdom primary care setting. GPRD/CPRD has historically collected data from VISION primary care practices (CPRD GOLD). More recently, CPRD has been able to release primary care data collected via the EMIS software system (CPRD Aurum).

Question 2: Please select any CPRD linked data or data products being requested

Primary care data collected by CPRD has been linked to a number of other datasets, including Hospital Episode Statistics, Office of National Statistics death registration data, Cancer Registration etc. These linked datasets are only available for English practices that participate in the linkage scheme with the exception of the GP practice postcode area level data. For more information about CPRD linked data please follow the link www.cprd.com/linked-data

Question 3: Proposed dataset(s) to be linked to CPRD data

Please provide information about the dataset(s) you wish to link to CPRD data. This should include the name of the dataset, the type of data (area level, patient or practice level etc.), the time period of data availability and the study period.

At present CPRD **ONLY** undertake linkages that can be taken forward under CPRD's section 251 approval. This means that linkage to any datasets that are not within CPRDs section 251 Master dataset would require CPRD to submit an HRA CAG amendment. The decision whether to submit a CAG amendment will be assessed by the CPRD Linkages Working Group.

Please note that CPRD's approvals do not permit linkages to non-research or non-health data which does not

support health service improvements. **Requests for linkages for such studies therefore will not be undertaken by CPRD.**

Under CPRD's HRA-CAG support CPRD is not able to link datasets that require free text data and CPRD does not hold or provide free text information. **Therefore, a proposed linkage to a dataset containing free text information would not be possible unless the free text was removed from the data in advance of any linkage.**

Question 4: Identifiers to be used for the proposed linkage

Please tick all identifiers the third-party data controller will send to NHS-Digital to enable the proposed linkage. Data linkage is enabled by NHS Digital, the statutory body in England legally permitted to receive patient identifiable data. CPRD and NHS Digital have a data sharing agreement in place governing the linkage process. CPRD has approval to link data at the individual patient level using some or all of the following identifiers - NHS number, date of birth, gender, postcode. CPRD cannot facilitate individual patient-level linkages based on any other identifiers.

If non-standard identifiers are planned to be used to establish a linkage, for example the NHS spine, you must be aware that this may not be feasible, may incur additional charges for services provided by NHS Digital and would need discussion with CPRD **BEFORE** submitting this application form. Please contact the CPRD linkages mailbox (cprdlinkages@mhra.gov.uk) to discuss this in advance of submission of this application.

For area level linkages please specify the linkage variable(s).

Question 5: Please confirm what legal basis the applicant receiving the final dataset from CPRD will be processing the data under.

Please confirm the legal basis you will be processing the final dataset released by CPRD. For more information about legal basis of processing please refer to the GDPR regulation (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679) and the ICO website (https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/) for more information.

If you have chosen Article 9(2)(h) please submit, as part of the application form, a legitimate interest summary. More information about this can be found on (https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/)

Under GDPR, CPRD's legal bases for processing data is usually Article 6(1)(e) and Article 9(2)(j). For more information about the legal basis of processing please refer to the GDPR regulation (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679) and the ICO website (https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/).

CPRD obtains annual research ethics approval from the UK's Health Research Authority (HRA) Research Ethics Committee (REC) to receive and supply patient data for public health research. CPRD must also gain annual section 251 support through the Confidentiality Advisory Group (CAG) at HRA to supply anonymised linked data for public health research. **For patient and practice level linkage, appropriate regulatory approval must also be held by each data custodian for their dataset to be linked to CPRD primary care data.**

Question 6: Coverage of the proposed dataset(s)

Any proposed non-standard linkage must have sufficient coverage overlap with CPRD data.

CPRD may be able to provide an estimate of the number of patients/practices/geographical areas/temporal coverage that are in both CPRD and the proposed dataset(s). Applicants can request this service by emailing the CPRD linkages mailbox (cprdlinkages@mhra.gov.uk) and applicants may be requested to provide additional information on the proposed dataset(s). If this service has been provided applicants should include the details of this in their application.

If an assessment of coverage overlap has not been conducted by CPRD, the applicant must provide an estimate of the expected number of patients available in the proposed dataset(s) and a justification of the feasibility of the proposed linkage.

Please be aware that under CPRD's section 251 support, patient level linkages to secondary care datasets can only be conducted with data from contributing English practices. Area-level linkages based on the GP practice postcode linkages can cover all countries within the UK.

Section C: Data flows and Data Security Assurance

Question 1: Proposed data flows and data flow diagram

CPRD will need to understand and evaluate which organisations will be sending or receiving data during the linkage process and the study. CPRD has certain requirements for the flow of data for any linkages, and an example is provided as a data flow diagram on the CPRD Non-standard linkage web page. In summary, NHS Digital receive identifiers from GP-system providers and expect to receive identifiers from the third-party data controller. These identifiers are used by NHS Digital to create a bridging file which is provided to CPRD to link variables from the new third party dataset to create the final anonymised linked dataset for study. Some variation on the data flow is permitted (please discuss with CPRD) but **CPRD will not undertake any linkage where the applicant or another party proposes to conduct the linkage. CPRD will always provide the final anonymised linked dataset for the study.**

Applicants should provide a data flow diagram with their application. This should include:

- All relevant data processors and data controllers
- Where the data flow includes identifiers or other variables.
- All storage locations of the data

- Who will be analysing the final linked dataset received.

To aid applicants, CPRD has provided an example data flow diagram on the CPRD Non-standard linkage web page. Applicants can also contact cprdlinkages@mhra.gov.uk to request a template created in Visio.

Question 2: Third-party data controller type (Sole, Joint, In-common) of the datasets to be linked to CPRD.

A data controller is defined as a person who (either alone or jointly or in common with other persons) determines the purposes for and the manner in which any personal data are to be processed.

The data controller must be a legal entity and will need to be able to sign a CPRD Data Sharing Agreement Include full contact details and address and registered name of the data controller(s). Often commercial companies may be part of a wider group or have multiple relationships - the data controller needs to be very clear in such circumstances.

Please be aware that the information required is about the organisation who own the data and not the applicant requesting a linkage. In addition, these terms are to understand ownership of third-party data that is to be linked to CPRD primary care data and not terms with the defined meaning under GDPR.

Question 3: Data Controller(s)

As explained in question 2, there may be more than one data controller. Please populate as appropriate to account for each data controller who has overall responsibility in the dissemination of variables and identifiers.

For this question please include the title, full name, job title, affiliation/organisation & e-mail address and telephone number of the Data controller. The person whose details are provided must be authorised signatory for such permissions within the data controller organisation.

Question 4: Formal written approval from the appropriate governance body within the third-party Data Controller.

CPRD will need to confirm that the third-party data controller(s) have provided approval for:

- the purpose of the linkage
- the identifiers mentioned in section B being sent to NHS Digital
- the clinical variables with pseudo ID's being shared with CPRD.

While CPRD does not proscribe a specific format for the written approval, it MUST:

- Include confirmation that the third-party data controller has relevant REC and CAG approvals (if appropriate) to share their data for research purposes with the relevant REC and CAG approval reference numbers.
- State the data controller type (refer to guidance on Question 2 in this section)

- Explicitly state that the third-party data controller(s) approve of:
    - the purpose of the linkage
    - the proposed study
    - the flow of identifiers mentioned in section B to NHS Digital
    - the flow of variables to CPRD.
- Include confirmation that the third-party data controller gives approval for CPRD to create and provide the final anonymised linked dataset to the Chief Investigator for the study
- Be on headed institutional letterhead with a named contact, job title, e-mail address and telephone number.

If the data has multiple controllers, please submit letters from the additional third-party data controllers.

Question 5: Would the third-party data controller consider making the linked dataset available to other researchers?

It is the vision of the CPRD to optimise research outputs by maximally linking person level data from different healthcare domains. Linkage between CPRD and datasets can be established on a non-standard or standard basis. A non-standard linkage is a one-off linkage that is undertaken to enable the conduct of a specific study. Such linkages are usually researcher-led, though public health bodies may request them to investigate a one-off public health concern.

A standard data linkage is an ongoing linkage that is led by the relevant data custodians and data release is approved for a broader purpose (for e.g. public health research, audit and service evaluation) rather than a specific study. CPRD has offered a range of standard linkages data services to researchers since 2013. Regular updates of standard data linkages are valued by the research community and as such, two thirds of research requests for CPRD data are for primary care data linked to other datasets. CPRD has existing agreements with other data custodians such as the Office of National Statistics (ONS), Public Health England (PHE) and NHS Digital (NHS D) to offer standard data linkages for public health research (www.cprd.com/linked-data).

Non-standard and standard linkages follow many of the same processes and as non-standard linkages can often take considerable time and work it may be preferable and more beneficial to the research community to make the linkage routinely available to other researchers.

If making this linkage available as a standard linkage is an option the third-party data controller would consider, please provide contact details for the relevant person within the third-party data controller organisation for the CPRD team to contact.

Question 6: Data Processor(s). Please list (with reference your data flows) each organisation that will be involved in processing, accessing or storing the linked dataset.

A data processor is responsible for processing personal data on behalf of a data controller. We require information on any processor/organisation that will have administrative access to the data released by CPRD to the applicant.

If there is more than one processor, please provide for each organisation, the name, address, lead contact name and email as well as the address where the data will be processed if separate to the address given. Please be aware that processor information is only required for organisations that will have access to the final dataset released by CPRD. Any processors involved before the data controller sends data to CPRD or NHS-Digital are not required.

If for any reason the processing of data will be at a separate location to the main address of the processors submitted, please include this information. Further guidance and information can be found on the ICO – https://ico.org.uk/

Question 7: Submit evidence of security arrangements for each data controller and data processor

CPRD requires information on the security arrangements in place for each data controller and/or data processors. The three acceptable pieces of evidence of these security arrangements are:

- an ISO27001 certification, for which the certificate must be provided
- a System Level Security Policy (SLSP), for which the policy document should be provided
- an Information Governance Toolkit, for which details must be provided on the form

For the Information Governance toolkit please provide the Organisation Data Service (ODS) code for each organisation that has been mentioned as a Data Controller or a Data Processor in this study.

CPRD will not review a provided SLSP however NHS Digital will review any SLSPs in detail for a DARS application (required for a patient level linkage). Therefore if submitting a System Level Security Policy it is recommended that you follow the guidance on the NHS Digital webpage (https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-process/data-access-request-service-dars-guidance-notes-on-security)

Please also provide the data protection registration number. This can be found on the Information Commissioner's Office (ICO) webpage (https://ico.org.uk/), please provide this registration detail (registration number, expiry date) for each data processor and data controller that has been mentioned in this application.

Question 8: Data Storage location - If separate to the processor location

If the data received will be stored in a separate location that is different to the processing location, please state where that will be and what organisation will hold the data. Please replicate as needed if there are multiple storage locations.