PIONEER
Health Data Research Hub

Data Request Form User Guide

User Guide

1. Introduction
2. Abbreviations
3. Data Access Process
4. Section A – The Project
5. Section B – The Data, Setting and Analyses
6. Section C – The Applicant
7. Data Trust Committee/Patient & Public Engagement (PPIE)
8. Appendices

1. Introduction

PIONEER is an ethically approved, secure research database and analytical environment containing data from healthcare organisations that provide unplanned or emergency care. The database is one of seven Health Data Research UK (HDR-UK) hubs.

Our aims

PIONEER aims to improve healthcare pathways and treatments by understanding the symptoms and diseases people have when they become unwell, either due to a flare or progression of a chronic disease or a new, acute illness. The database not only includes details of the acute presentation to a highly granular level, but also longitudinal data of preceding and subsequent health care contacts, diagnostic pathways, and detailed care plans.

Our objectives

PIONEER enables innovative researchers (including from academic, NHS based, policy, commercial and third sector organisations) to understand clinical problems in more detail (including guideline compliance and comparative medicines use) and then to develop, test and deliver advances in clinical care with real-time information about how unwell people access and use health services.

PIONEER is committed to working with researchers, patients and the public to ensure that the research we facilitate has benefit and meaning for patients. By working transparently, actively engaging members of the public in our research and enabling public oversight in our decision-making process, we are building public trust in the benefits of responsible data sharing.

This document is designed to provide guidance in the application process for data from the PIONEER database. It should not be taken as an instruction or providing information or answers to the questions asked in the Data Request form. If you have any queries, please do not hesitate to contact the PIONEER team.

2. Abbreviations

Title	Meaning
NHS	National Health Service
DTC	Data Trust Committee
PPIE	Patient & Public Involvement & Engagement
TRE	Trusted Research Environment
UHB	University Hospitals Birmingham NHS Foundation Trust

3. Data Access Process

Submitting a data request to PIONEER is straightforward. This User Guide will take you through the various stages, giving examples and explanations of each section to allow you to populate the necessary areas of the form. An online version of this guide and a downloadable Data Request form in Word are available on the PIONEER website – www.pioneerdatahub.co.uk – fill in all the sections, upload any additional information and click submit.

The form must be completed in its entirety. If at any point you would like to discuss your data study/project prior to raising a data request, then please contact us and our team will arrange a convenient time to talk through this with you.

Prior to submitting a Data Request Form, we strongly suggest talking through your project with the PIONEER team. We can advise you if the needed data is available within the PIONEER database, and the costs of data curation and access. We can discuss any additional consultancy services you may require, such as support with analytics; clinical, patient or healthcare provider expertise; regulatory support or clinical trial/algorithm design. We can also discuss options for staging data for access, with private and secure PIONEER Trusted Research Environments available on demand. These services, and more, are available to help you gain the most from the data requested.

Once you are sure of the data and consultancy services you require, the process is –

- Submit your application (the Data Request Form) to PIONEER.
- Our internal review process will ensure your form is fully complete, meets the requirements of each section of the document and that the request meets our due diligence and governance requirements.
- Your request is presented to our Data Trust Committee - a group of patients and members of the public who review every data request to ensure there is the potential for public benefit, were the request supported.
- If the request passes our due diligence, governance and Data Trust Committee review, a Data Sharing Licence Agreement is developed and signed by all relevant parties.
- Data is extracted and curated.
- Data is placed in a secure Trusted Research Environment (TRE), identified and agreed in advance.
- If this is a PIONEER TRE, access is permitted for the agreed time.
- Additional consultancy services will start, as agreed within the project, to support you.

The following sections of this document relate to the PIONEER Data Request form and will take the 3 sections built into the form. The information contained in this document is for guidance only and if

when completing the Data Request form there are questions or queries please refer to a member of the PIONEER team by contacting us on pioneer@uhb.nhs.uk

Section A – The Project

This section of the form is split into eleven separate elements. The purpose of this section is to allow you to tell us more about your study/project, its aims and objectives, your involvement with patients and the public, the type of dataset you require, etc. This area is split as follows –

- **A1 Project Title**. We would like you to state in no more than 200 characters, the title of your study/project. This could be, for example - *An exploration of the number and reason for pregnant and newly postnatal women's attendance at Emergency Department at University Hospital (119 characters)*
- **A2 Aim(s) of your project**. We would like you to explain in no more than 200 words what are the aims of your project. Here you should consider outlining what your project will assess or what questions will be answered by the project.
- **A3 Background and scientific rationale of the proposed project**. In no more than 300 words we would like you to describe the background to your study/project area. This could include how you have arrived at the data request, why the study area needs further work or why there is a particular issue. It should include the scientific rationale to this area of study along with any scientific background or previous work that may have been carried out in this area. It might help to describe the current situation if one exists and why the area of study may help to minimise risk, improve patient outcomes, or assist in recording information more appropriately. This section will be reviewed by the academic teams within PIONEER and should be justified with references where possible.
- **A4 Lay summary**. This section should be no more than 500 words. It is of the utmost importance that this is written for a lay audience, and the form will be returned if it is unclear or contains jargon. This section is designed to allow the Data Trust Committee to understand your project. It should include a brief introduction to the area or problem you are trying to address, the aims of your project, how you will meet those aims, how the results will be shared and used (being as specific as possible) and how this study might benefit patients. It is important to recognise that there are risks in allowing access to sensitive health data, and you should justify how these risks have been considered and mitigated, where possible.
- **A5 Patients and the public involvement.** This section is to understand how you have engaged with patients and the public in developing this project. This includes actively listening and interacting with patients and the public so that they are aware of and can comment on your project. We also want to understand how patients and members of the public are or will be involved with the project. Involvement gives patients and members of the public the opportunity to help you shape the research, ensuring the research is delivered with patients and members of the public. We would like you to be specific about involvement and engagement activities, how they have influenced your proposal, and what activities will be planned for the duration of the study. The word count in this section is 300 words.

We recognise that not all researchers have access to patient groups or links to established public forums. You may find our **Patient and public involvement and engagement best practice playbook** helpful.
This guide describes why patient/public involvement and engagement is important, the benefits it can bring, and how to start your engagement and involvement activities. The PIONEER team can also help you with your patient and public involvement and engagement work, putting you in contact with various teams and public groups.

At PIONEER, we are committed to working with researchers, patients, and the public to ensure that the research we facilitate has benefit and meaning for patients. By working transparently, actively engaging and involving members of the public in the research we support, and enabling public oversight in our decision-making process, we are building public trust in the benefits of responsible data sharing.

- **A6 How will access to PIONEER data help your project?** This is your opportunity to describe how access to PIONEER data will help with your study/project. This is to provide assurances to our Data Trust Committee that access to our data is necessary and important for you to meet your aims. This section should be 300 words or less.

- **A7 The type of dataset required**. In no more than 100 words please describe the type of dataset that you require. Please describe the general scope of the data or provide an overview of what data fields you need. For example, "data on pregnant women presenting to hospital with an unplanned healthcare need. This includes background demographics, medical conditions presenting complaint, all tests and investigations, all usual and acute medications and outcomes over the subsequent 6 months". If there are any sensitive fields or demographics or disease areas which make reidentification more likely (such as at the extremes of age or with rarer diseases), please state this here. PIONEER can provide bespoke synthetic data, and data can be structured involve images or be derived from free text using natural language processing. We are happy to support you in making your data selection. We will need an exact data specification for contracting, so please attach that to your DRF once available.

- **A8 The expected value of the project to patients and the NHS.** PIONEER supports data access that is likely to lead to patient or public benefits. In 300 words or less, please explicitly state how supporting this project will benefit patients or members of the public. Please give specific examples of potential benefits and give timelines where possible.

- **A9 Up to 6 keywords which best summarise your proposed project –** it is here that we would like you to choose 6 key words that best describe your study area. For example, if your study relates to pregnant or postnatal women then some of the keywords' you might wish to consider could be – Pregnancy, Maternal health, Obstetric, Care pathways, etc.

- **A10 The estimated duration of the project**. Please give an indication of the predicted timescale of your proposed study/project. This will inform the sunset clause on the Data Licensing Agreement.

- **A11 How will results be shared/disseminated.** PIONEER is committed to open and transparent science, and in publicising the benefits of data sharing across different sectors. We expect the results of this analysis to be available in peer review publications, reports, and lay summaries. We would prefer for them to be made available in open access journals. We require PIONEER to be acknowledged in any outputs. In no more than 300 words, please state how you will meet these requirements/expectations and how you plan to share/disseminate the results of your study/project to a wider audience. Think about the different beneficiaries of your research and how you will target them. If the results may be commercially sensitive, please state this here and provide your pathway to sharing results in time.

Section B – The Data, Setting and Analyses

This section is split into 6 sections to allow you to provide us information relating specifically to the data, the environment you will be working in, requirements to comply with data security and safety standards, etc. The areas in this section are –

B1 Level of data access requirement. This section relates to the level of data access that you need. This is to meet our requirements for data minimisation and the 5 safes, ensuring people are only exposed to the level of data they need. Most analyses undertaken by researchers require access to anonymised individual patient-level data. If PIONEER are undertaking the analysis for you, or you are undertaking a pilot or relatively simple study, you may only require access to aggregate data. Aggregate data is high-level data which is acquired by combining individual-level data.

B2 and B3 Data Environment. This section refers to the environment that the data will be placed into. Our preferred option is for you to access the data solely within a bespoke and private PIONEER Trusted Research Environment (TRE). This ensures that all security and regulatory requirements are met, and the data is controlled and managed in a safe and secure way.

Question B2 asks if you would like to use a PIONEER TRE and what tooling you would like to use for data analysis. Please contact a member of the PIONEER team to discuss your TRE requirements including pricing and tooling.

Alternatively, you may wish to transfer the data to a data safe haven of your choice. Question B3 asks for more information about this. The reasons for data transfer are required, with some justification as to why this is your preferred option. The name of the legal entity receiving and responsible for data is required for contracting. Data can only be transferred if the safe haven you wish to use meets all the necessary regulatory requirements that we may specify. The data security standards we refer to are –

- ISO27000 Series compliant - ISO - ISO/IEC 27001 — Information security management

- NHS Data Security & Protection toolkit - Data Security and Protection Toolkit - NHS Digital
- The National Data Guardian (NDG) Standards for Healthcare data - National Data Guardian - GOV.UK (www.gov.uk)
- The Data Security Standard Overall Guide - Data Security and Protection Toolkit (dsptoolkit.nhs.uk)
- NDG Review - Review of Data Security, Consent and Opt-Outs (publishing.service.gov.uk)
- Data Security & Awareness Training - Data Security Awareness - eLearning for healthcare (e-lfh.org.uk)

We are happy to discuss PIONEER's TREs and your own data safe haven and can support you with technical conversations with your data safe haven providers, should you be unsure about the technical standards. We are likely to verify the security of your chosen haven, and this will form part of the Data Licensing Agreement.

B4 Statistical Analysis. Please describe the form of statistical analysis you will undertake and how it is intended that you will present this to your intended audience(s). There is reference made to smallest cell value that is likely to be generated. This includes aggregate data presented in a table. Dependant on the field we will ensure that no data field >10 is disclosed and when analysing the data, we expect that you will also not present any data with a cell count of <10. If you feel this is necessary, please contact the PIONEER team to discuss.

B5 Machine Leaning. Please confirm if machine learning will be undertaken on the dataset. If the answer is "No" then please move to B6 Ethical Approvals. If the answer is Yes, please describe the techniques you intend to use and the likely output of the analysis, including whether this is for algorithm generation and training; internal validation; external validation or "Other", with an option to describe the data use.

B6 Ethical Approvals. PIONEER has its own overarching ethical and CAG approvals which can be used to support your project. Alternatively, you may have ethical approvals of your own in place and wish for the research to be performed using your own ethically approved protocol. If you wish to use the PIONEER ethical approvals, please state so here. If you intend to use your own ethical approvals, we will require evidence to support this.

B7. Cost Model. PIONEER operates a cost recovery model for activities which support data access. Please either confirm that costs have been discussed. If not, please contact the PIONEER team to discuss costs.

Section C – The Applicant

This last section relates to the applicant, and the organisation(s) they work for or are working with if this is a joint application. This section will be used for the Data Licensing Agreement and for contracting, so it is important all sections are complete.

C1 Lead Applicant – we require name, email address, current position, workplace organisation and the specific role this person is undertaking in the project/study area. The details contained here will allow us to maintain contact with and update this person as the data request progresses through our

systems. This person needs to be substantively employed by the lead organisation. If this application is part of a studentship, please give the primary supervisors name here, and the student's name as a co-applicant.

C2 Evidence of Lead Applicant's expertise and experiences relevant to delivering the project. As part of the 5 safes, we only share data with "safe people", meaning people who have the relevant expertise to complete the work. Evidence of this can be in the form of their CV, a list of up to 5 of the most relevant publications or descriptions of previous work.

C3 Sponsoring/Lead Organisation. This is the name of the organisation who is leading on the project or is the sponsor of the project. This must be a legal entity who can sign legal documents between parties. We require the Sector of the lead organisation, as well as its size in terms of numbers of employees. This will also be used for our due diligence check.

C4 Co-applicants – please provide details of any co-applicants who may be part of the study. We require their names, their current position and details of the origination or institution they work for. If the project is forming part of a studentship, please place the student's name here and state that they are student. PIONEER is committed to supporting the training of researchers with an interest in health data and welcomes applications which form part of doctoral and post-doctoral training degrees.

C5 Other significant project team members –please provide us with details of any other team members who may be involved with the project. We need to know their name, current position, organisation name and their specific role in the project. This is especially important if you wish these people to have access to the data, which is provided on a named basis.

C6 Contracts lead in Organisation – this is the person who will agree the terms of the Data Licence Agreement and will be the signatory of any legal documentation. We need their name, email address and contact telephone number.