

**BHF Data Science Centre Cardiovascular and
Diabetes Cohorts Trusted Research
Environment (TRE) Platform: Call for Funding
(CFF)**

23 November 2022

Overview

Funder:	HDR UK (through the BHF Data Science Centre)
Funding type:	Grant
Amount:	Up to £580,000 (inclusive of VAT)
Duration:	Up to 3 years (with potential for funding beyond this duration)
Opening date:	23 rd November 2022
Closing date:	30 th December 2022 at 5pm

Summary

This is an opportunity to apply for funding to co-develop a Trusted Research Environment (TRE) with the British Heart Foundation Data Science Centre (BHF DSC) at Health Data Research UK (led by HDR UK). The aims of the TRE platform are to fulfil the secure and efficient data storage, linkage, sharing and analysis requirements of UK-wide disease-specific cohorts, with an initial focus on cardiovascular and diabetes cohorts. This TRE platform will provide the technical infrastructure, capability and associated data governance processes to host and enable approved access to data from consented cardiovascular and diabetes research cohorts (which may include bespoke clinical data collection, imaging, omics and wearables data). There will be a key focus on enhancing these cohorts through the incorporation of linked data from routinely collected, mainly structured health-relevant data from national (primarily NHS) data custodians. The overarching aims are to coordinate and enable these linkages and to provide secure, remote access to data from each cohort together with linked health-relevant data for (i) cohort owners to manage their cohort dataset and the associated linked data; (ii) accredited, approved researchers to carry out approved research projects. The call is to select a TRE platform provider with the appropriate experience, expertise, accreditation, organisational commitment and collaborative ethos to work closely in partnership with the BHF DSC in developing, providing and running the platform, initially to support UK-wide cardiovascular and /or diabetes cohort studies but with the ambition of scaling to provide a disease agnostic platform that can support the needs of a wide range of research cohorts that recruit people with or at high risk of specific diseases across a diverse range of conditions.

Context

Cardiovascular disease (CVD) remains one of the most common causes of mortality and morbidity worldwide. The United Kingdom (UK) has a rich and diverse collection of studies that recruit people with or at high risk of developing various diseases, including cardiovascular diseases or diabetes (CVDD), collecting

detailed information about them, and following their health over many years (cohort studies) to understand causes and consequences of disease. This has led to insights which have been translated into benefit for patients and the health of the public.

To date, however, such cohorts have not been able to take full advantage of the enhancements offered by linkages to routinely collected health-relevant data, particularly data collected at national level. Whilst the focus of recruitment and initial data collection usually occurs at the point of acute or specialist secondary care, the integration of data from nationally collected routine health records has huge potential to substantially enhance research based on these cohorts by providing rich health-relevant information: (i) from before the event or diagnosis triggering recruitment, so enhancing research on the antecedents, causes, predictors, subtypes and diagnosis of disease; and (ii) for long term follow-up of health after recruitment, so enabling research on outcomes following disease, including response to treatments. Such cohorts often collect detailed information about the condition at a given point and may include additional measures such as genomics and blood-based biomarkers, structured and unstructured imaging data, data from mobile phone apps and wearable devices.

The key aims of developing the TRE platform are to:

- provide a cost- and time-efficient solution to enable linkage of multiple cohorts to routinely collected health-relevant data;
- host data collected by the cohorts together with the data from these linkages, and;
- provide a secure, privacy-preserving, trustworthy and scalable means for cohort owners and approved researchers to access and analyse the enhanced, multimodal cohort data.

This will overcome several key challenges that cohorts currently face, including:

- navigating the complex processes for access to and linkage of routinely collected health-relevant data, principally but not exclusively from national NHS organisations, involving multiple data custodians and varying requirements across the four nations of the UK;
- meeting the escalating costs and providing the necessary expertise required to establish, develop, run and maintain the secure data infrastructure(s) and associated capabilities required for linking, cleaning, curating and sharing cohort data safely and equitably with a broad community of approved researchers.

Currently, these challenges tend to be dealt with by each cohort separately, with significant demands on time and resources (both for cohorts and for data custodians), and variable, often limited success. The proposed TRE platform and its associated services will provide an expert-led, centrally coordinated approach. This will overcome these challenges and so accelerate translational findings from cohort-based research, in several ways, including the following:

- reducing overall costs and duplication of effort by multiple cohorts, through providing a coordinated, facilitated process for requesting access to linked health-relevant data;
- improving efficiency and achieving economies of scale through establishing a single, multi-tenant, shared TRE to support the secure data upload, storage, linkage, remote access, analysis and sharing needs of multiple cohorts;

- reducing the burden on and costs of data custodians, through a unified, streamlined approach both for requesting access to and for the provision and linkage of health-relevant data across multiple cohorts within this TRE;
- further reducing the costs of data linkage through specifying a set of common requirements for linked datasets - and data items within those datasets - across cohorts (recognising that individual cohorts tend to have very similar/identical linked data needs);
- avoiding the data security risks associated with the dissemination of health-relevant data from data custodians to multiple separate local NHS and/or university systems for linkage to multiple cohorts, through enabling well-documented, streamlined, secure data transfer into a single, multi-tenant TRE, using robust protocols that data custodians can readily understand, trust and – where required – audit;
- maximising consistency, transparency and efficiency of data access, linkage, governance and data management processes (which currently vary substantially, often without a clear reason).

Scope

The aims of this call are to enable the BHF Data Science Centre to:

- i. **support UK-wide cardiovascular and diabetes cohort research studies** to create new knowledge, accelerate and expand research outputs, provide new insights for policymakers and benefit patients and their carers;
- ii. **co-develop, improve and enable broad use of platform and data infrastructure, capabilities and processes** for secure data upload, storage, efficient linkage to national NHS records and other health-relevant data, and remote access, providing a single ‘off the shelf’ resource for cohorts and their research communities;
- iii. **inform the platform analytical environment** to support multi-modal data analysis integrating a wide range of data types and advanced analytics, including machine learning approaches;
- iv. **co-develop processes and establish mechanisms** to provide well curated data which will allow reproducible analysis;
- v. **collaborate with relevant parties to develop and implement streamlined data governance processes** to support approved access to linked cohort data by accredited researchers;
- vi. **provide reassurance to funders and other stakeholders** of an efficient and cost-effective mechanism to collect, store, link and provide access to data, while satisfying all necessary legal and ethical governance and security requirements.

This call is to fund a partner organisation who will **work in collaboration with the BHF Data Science Centre** to co-develop a secure and scalable platform infrastructure and related processes within an accredited TRE with the ability to:

- provide a separate, partitioned workspace for each adopted cohort study (see **Appendix A** for summary details on the initial set of ‘vanguard’ cohorts);

- provide the capability to store an agreed subset of direct identifiers for each cohort separately from the de-identified participant data for research, to enable the provision of identifiers for linkage to relevant national data custodians and to facilitate legitimate participant recontact by cohort owners;
- provide secure, authenticated, remote access mechanisms to allow named cohort staff to access their cohort data and to use a variety of common tools for data management and/or analysis;
- support entry of data directly into the TRE as individuals are recruited into cohorts with new or ongoing recruitment (including being able to handle research subject data requests, such as deletion);
- transfer existing data (including demographic and clinical data, and data derived from omics assays and images) from already recruited cohorts into the TRE and provide suitable data models;
- collaborate with national (principally NHS) data custodians to coordinate, streamline and enable the secure transfer and accurate linkage of NHS data from national data custodians across all UK nations into the TRE;
- collaborate with an expanding number of BHF DSC adopted UK-wide research cohorts (for a listing of key feature of initial 'vanguard' cohorts and other relevant background information, see **Appendix A**), initially focusing on cardiovascular and diabetes disease (CVDD) cohorts, but with the ambition of scaling to provide a disease agnostic platform;
- maintain secure access control to prevent unauthorised use or download of data;
- provide secure, authenticated, remote access mechanisms to enable remote secure access to subsets of linked de-identified cohort data for accredited researchers to conduct approved research projects;
- provide the technical capability and co-develop processes to enable each cohort to retain control of how their participants' data is used and which research projects are provided with access;
- co-develop and integrate reproducible data management, curation and analysis pipelines developed by researchers and the BHF Data Science Centre health data science team, including algorithms to define disease phenotypes from linked health data;
- co-develop methods and standard operating procedures to support the ingestion and analysis of multi-source and multi-modal data (e.g. data from images, multi-omics, wearable devices and apps) to enable efficient and faster outputs of cohort-based research;
- co-develop robust governance and security processes ensuring protection of participant confidentiality, accurate linkage and secure access to health data for research;
- implement appropriate technical and operational measures to ensure an acceptable level of security appropriate to the risk as required by Data Protection Legislation in relation to the personal data in the datasets;
- work within the [Five Safes Framework](#);
- demonstrate enhanced data security standards with ISO certification and other relevant certification (e.g., NHS Data Security and Protection Toolkit (DSPT), Cyber Essentials Plus certification);
- provide knowledge and capacity to comply with the Secure Data Environment accreditation framework (being developed by the NHS Transformation Directorate) following its rollout.

What we are looking for

We are looking for applications from parties that have:

- a track record in co-development and delivery of a secure virtual environment with remote access for hosting new and legacy cohort data and enabling data linkage and analysis;
- capability to manage and provide access to appropriate hardware and cost-effective provision of compute for at scale, multimodal analysis;
- appropriate software packages and tooling such as SQL, SPSS, R Studio, Python Pyspark, STATA and MS Office with an ability to provide regular version updates and to enable the provision of more specialised software as required for particular data types;
- a track record of obtaining the trust of, coordinating, receiving, and linking multiple health data sets from data custodians of health and health-relevant data (in particular – but not exclusively – national NHS data custodians);
- a track record of collaborative partnership working – in particular, with academic and NHS organisations – in the linkage of cohorts to routinely collected health data;
- expertise in establishing mechanisms for the collection and secure separate storage of personal identifiers to enable linkage of cohorts to health relevant datasets;
- demonstrated experience of supporting research at scale with different data types (e.g., imaging, omics) and demonstrable ability to develop new solutions to support ingestion and analysis of new data types;
- experience of co-developing data, security, governance, and operational processes for the secure linkage and access to health data for research, in particular in response to a changing governance and regulatory environment;
- ability to advise on the aligned use of data standards across cohorts and identifying potentials for interoperability;
- experience of curation and data management support and facilitation;
- ability to develop infrastructure components that are fully portable to an alternative secure data environment;
- appropriate data security standards and other relevant certification, with a comprehensive awareness of - and ability to adapt and comply rapidly with – a changing regulatory environment, including new/emerging accreditation requirements;
- ability to meet the MRC and HDR UK Terms and Conditions of funding detailed in **Appendix B**;
- ability to rapidly review and agree collaboration T&Cs, invoice within strict, non-negotiable and tight timelines for upfront payment of the award and be able to start working with BHF Data Science Centre rapidly after the award is made;
- a highly credible team;
- value for money – noting our public sector funders, we operate for the public good;
- alignment to HDR UK and BHF DSC values (p8 below).

Eligibility

- You must be based in the UK.
- Accredited TRE platform provider.
- You must have experience and an excellent track record in delivering platform infrastructure to link cohort and health data from across the 4 nations of the UK, and ideally worldwide.
- You must be in a position to capitalise funds in the timescales available (see below and **Appendix B**).

Funding available

Up to £580,000 inclusive of VAT is available for this call.

All costs and – where appropriate – different options for how to make most appropriate use of this investment and align it to the call - should be fully described in the Justification of Resources section of the Call Response Form and detailed in the budget Excel spreadsheet.

Terms and conditions and details of eligible and ineligible costs are in **Appendix B**.

Application process

Clarification questions from potential applicants must be received by 17.00 GMT Friday 16th December 2022 or earlier. These must be submitted by email to bhfdsc@hdruk.ac.uk with the clear subject heading BHF DSC Cohorts TRE funding call. Any questions received after this date may not be responded to.

Applications must be submitted using the clear subject heading (BHF DSC Cohorts TRE funding call), by 17.00 on Friday 30th December 2022 to bhfdsc@hdruk.ac.uk and must be provided in a single compressed .zip file containing:

- Completed Call Response Form;
- Completed budget template (Excel template provided);
- Accompanying materials collated into one pdf document, including:
 - Programme plan Gantt chart indicating clear milestones for deliverables or project phases
 - Risk register
 - CVs for all named applicants (max 1 page per applicant detailing key information in relation to the call).

Text should be single-spaced, with page margins of at least 2cm, using Arial 11 font. Applications received after the deadline passes will not be considered.

Timing

All timings below, including proceeding with this process, are subject to change at HDRUK sole discretion.

- Funding call launched 23rd November 2022;
- Clarification questions from potential applicants submitted by 17.00 GMT 16th December 2022;
- Applications submitted by 17.00 GMT 30th December 2022;
- Notification of intention for interview 12th January 2023;
- Interview TBC (w/c 16th or 23rd January 2023);
- Successful application announced 30th January 2023;
- Grant commences 13th February 2023, subject to contract and due diligence.

About Health Data Research UK

Health Data Research UK (HDR UK) is the UK's national institute for health data science. Our mission is to unite the UK's health data to enable discoveries that improve people's lives. Our vision is that every health and care interaction and research endeavour will be enhanced by access to large scale data and advanced analytics.

We have established national research programmes that use data at scale, and we are building an infrastructure to enable the responsible access and analysis of this data. Our work is structured around three key themes:

- Uniting health data – including our work with data custodians through the [UK Health Data Research Alliance](#) and on making health data discoverable and accessible through the [Health Data Research Innovation Gateway](#);
- Improving health data – incorporating contributions from HDR UK's [Human Phenome](#) and [Applied Analytics](#) priority areas, the [Health Data Research Hubs](#) and the [BHF Data Science Centre](#), including our work on improving data quality and standards, on data curation, and on developing tools and methods for sharing, linking and analysing data;
- Using health data –including our work on research discoveries and skills development across four national priority areas: [Understanding the Causes of Disease](#), [Clinical Trials](#), [Public Health](#) and [Better Care](#).

How we work

Our values are our guiding principles:

1. **Transparency:** we will share information, insights, and innovations so that we learn faster together;
2. **Optimism:** we believe that we can make things better, that we can do things differently and that we can overcome challenges to create a new and thriving health data ecosystem that benefits patients and the public, the NHS, scientific discovery, and industry;
3. **Respect:** we deliver better results when we work in a truly interdisciplinary way. We listen, share, and respect a diversity of thought and opinion, perspective and experience. We are inclusive - leveraging and fairly attributing the expertise and capabilities of others;

4. **Courage:** we are leading the way and will be prepared to try new things, take risks, embrace ambiguity, and challenge the status quo. We will contribute opinions to shape the future of health data research;
5. **Humility:** we have a lot to learn from others; and aim to be free from pride and arrogance.

About the BHF Data Science Centre

The [British Heart Foundation \(BHF\) Data Science Centre](#), is building on a £10m initial investment from the BHF to deliver the data and data science needed to address some of the most pressing challenges in heart and circulatory health research.

The centre works in partnership with patients, the public, NHS, researchers and clinicians to promote the safe, ethical and scientifically robust use of data for research into the causes, prevention and treatment of all diseases of the heart and circulation (including, for example, heart attacks, heart failure, heart rhythm disorders, stroke, peripheral vascular disease and vascular dementia).

Extensive and ongoing engagement with key stakeholders has shaped the development of the centre's six thematic areas:

- Better access to and use of **structured health data** UK population-wide for cardiovascular research;
- Better access to and use of **unstructured health data** (including imaging data) at scale for cardiovascular research;
- Enabling large-scale use of **personal monitoring data** in a wide range of cardiovascular research;
- Developing and refining **computable cardiovascular phenotypes** for different applications;
- Supporting discoveries of cardiovascular disease causes, prediction, early detection, prognostic tools and treatments using **disease-based cohorts**;
- Developing methods and infrastructure for large, efficient, **data-enabled cardiovascular trials**.

The BHF Data Science Centre does not hold data itself. Instead, it works with relevant data custodians, including through the UK Health Data Research Alliance and Health Data Research Innovation Gateway, to provide knowledge and expertise to help researchers from the NHS, academia and industry find, access, understand, connect and analyse the UK's unique cardiovascular 'big data' that is distributed across national registries, NHS electronic medical records and other relevant datasets.

Contact

For queries relating to this funding call, please contact: bhfdsc@hdruk.ac.uk

Appendix A

Summary Information on Vanguard Cohorts

Study	Stage	Recruitment	Study data	Data linkage requirements	Long-term requirements
HFpEF	New study due to start recruitment in July 2022	~10,000 participants (heart failure with preserved ejection fraction), UK-wide	Collect study data to database in TRE, includes imaging data	Hospital records (e.g. HES), primary care, death registry, CVD specialist datasets (e.g. NICOR audits)	Research cohort that can be used for other researchers to carry out approved research in this area
SIGNUM	Cohort recruited	~10,000 participants (stroke), UK-wide	Existing study database, includes blood sample data	Hospital records (e.g. HES), primary care, death registry, CVD specialist datasets (e.g. NICOR audits, SSNAP)	Research cohort that can be used for other researchers to carry out approved research in this area
COVID-HEART	Cohort recruited	~380 participants (COVID-19 and cardiac damage [raised troponin]), England, Wales and Scotland	Existing study database, includes imaging data	Hospital records (e.g. HES), CVD specialist datasets (e.g. NICOR audits)	Follow up at 1,3 and 5 years. Link to other cardiovascular COVID cohorts (~1000 participants)
ADDRESS-2	Recruitment ongoing (registry)	~8,500 participants (type 1 diabetes), UK-wide	Existing study database, includes blood sample data	Hospital records (e.g. HES), primary care, death registry, national diabetes audit	Research cohort that can be used for other researchers to carry out approved research in this area (Governance in place)
R4VAD	Recruitment finishing Summer 2022 and data linkage will be required in the coming months	~2,500 participants (cerebrovascular small vessel disease), England, Wales and Scotland	Existing study database, includes imaging data	Hospital records (e.g. HES), medications, primary care, death registry, specialist psychiatric data, place of care	Research cohort that can be used for other researchers to carry out approved research in this area (Governance in place)

Beyond these initial vanguard cohorts, we estimate that there are tens (perhaps 20-40) of further cardiovascular and diabetes cohorts with a UK-wide geographical footprint, similar complexity and sample size, and potential to benefit from the proposed platform. We estimate that there are likely to be hundreds of further similar cohorts that have recruited, are recruiting and plan to recruit patients with a wide range of other conditions for a future disease-agnostic expansion.

Appendix B

Funding conditions and requirements

The funding is subject to MRC and HDR UK conditions of funding as well as to the awardee organisation's ability to sign the required agreements with HDR UK (e.g., grant award, collaboration agreement) by 10th February 2023. The funding is for up to 3 years, but the awardee organisation must be in a position to invoice for the entire award by the start of March 2023. The expectation is of rapid co-development and set-up of a functional TRE with essential processes and services in place within the first 2 months of the award start date, followed by co-development of further functionality and processes and ongoing service provision. Subject to excellent delivery and research cohort demand, there is the potential for further funding beyond this award.

Funding available from HDR UK is up to £580,000 (inclusive of VAT). For this award, HDR UK will fund 100% of direct costs to the grant, and not on a proportion of full economic costing (fEC) basis. All directly attributable costs are eligible, including direct support staff. Organisational overheads and any activities not directly contributing to the TRE platform are not eligible for funding.

Eligible Costs

- Costs for equipment such as hardware, software for the development and implementation of platform infrastructure;
- Database space and storage for new and legacy cohort data as well as storage for linked data including data types such as imaging, omics data and routinely collected data;
- Database space and memory (e.g., GPU) for innovative AI and machine learning functionality;
- Service support costs for the environment;
- Staff salary costs for development and implementation of infrastructure, protocols and processes. Note that costs incurred as a result of staff spending time on capital projects can be capitalised, provided that time can be linked to bringing a specific, separately identifiable asset into working condition, or substantially enhancing the working life of an existing asset.

Ineligible costs

- Travel and subsistence;
- Costs related to research or development of new technologies or software packages;
- Maintenance costs;
- Depreciation costs;
- PhD studentships;
- Indirect and estate charges.

Terms and Conditions

This award will be subject to the following terms and conditions: [UKRI-021122-fECGrantTermsAndConditions.pdf](#) (N.B. 100% direct costs are funded and this award does not use the fEC model); [MRC-25052022-MRC-Additional-terms-and-conditions-April-2022.pdf \(ukri.org\)](#); and additional HDR UK requirements, including but not limited to the open access policy <https://www.hdr.uk.ac.uk/about-us/policies/open-access-statement/>, and the attribution policy <https://www.hdr.uk.ac.uk/about-us/policies/hdr-uk-attribution-policy/>