Integrating Our Community

Year 1

Kay Snowley, Lara Edwards, Ben Crosby, Helen Tatlow

Contents

Introduction

Welcome to the first HDR UK 'Integrating Our Community' pack! This pack is intended to give you a snapshot of all the fantastic activities underway across our community, signpost opportunities for wider involvement and highlight key contacts for collaborations. It summarises the highly useful data provided in the Year 1 workplans across Regions, Driver Programmes and Infrastructure and Services Programmes. Furthermore, it highlights important Partnership programmes such as the BHF Data Science Centre, Hubs and DARE UK, promoting opportunities for integration, collaboration and cohesion. Strategies and workplans that are currently still in development will be added to future updates of this pack as they evolve, along with any new major developments that offer future opportunities for integration and collaboration.

We would appreciate feedback on the pack, including what you would like to see included in future versions (we expect to send out the next version in six months' time).

Programme Team

Kay Snowley - Programme Director, Infrastructure and Services

kay.snowley@hdruk.ac.uk

Lara Edwards - Programme Director, Driver Programmes

lara.edwards@hdruk.ac.uk

Ben Crosby - Programme Officer

ben.crosby@hdruk.ac.uk

Helen Tatlow - Programme Officer

helen.tatlow@hdruk.ac.uk

Driver Programmes

Medicines in Acute and Chronic Care

Key Activities in Year 1

- Develop a draft international framework for the identification of medicines-related harm.
- Prioritise which medicines research should be included in initial projects.
- Deliver a minimum viable product (MVP) for each use case, with tools to be refined in Year 2.
- Agree data specifications for prioritised projects and data access processes/technologies.
- Develop a metadata catalogue for data, sharing this with the Innovation Gateway.
- Operationalise the identification of medicines-related harm, building automated tools for testing.
- Build the first national medicines public stakeholder group, while co-developing patient and public involvement and engagement (PPIE) strategy.
- Recruit PhDs and academic staff.

Workstreams	• WS1: Medicines innovation. • WS2: Data and enabling technologies. • WS3: PPIE. • WS4: Capacity/capability building. • WS5: Communications and stakeholder engagement. • WS6: Operational delivery.
Wider Community Linkages	• Regions: HDR North, Midlands, Wales, Scotland, South West. • PIONEER Data Hub. • Useable Data Pillar. • Trusted Research Environments (TREs); Northern Ireland Honest Broker Service (NI HBS), Secure Anonymised Information Linkage (SAIL) Databank, Scottish National Data Safe Haven and regional TREs.
Dataset Priorities	• Primary care: SAIL Databank, Clinical Practice Research Datalink (CPRD). • Secondary care including laboratory data: PIONEER, Scottish National and Regional TREs. • MHRA Yellow Card and FDA FAERS. • Throughout Year 1 we will explore wider access to regional and national level data via NHSE SDEs, OpenSAFELY, NI TRE and UK Biobank.
Technology and Infrastructure Priorities	• We will use SAIL and PIONEER environments from the outset, and with the Technology Services Ecosystem (TSE) Pillar we will explore novel technology services and solutions.

Key Opportunities	• **Data Strategy**: Developing a comprehensive, UK-wide medicines data strategy. • **Technology and Standards**: With the TSE Pillar, testing novel technology services. • **Streamlining Information Governance**: Working closely with UK HDR Alliance members to streamline information governance processes, developing best practices and frameworks. • **Capacity building**: Recruiting post-doctoral research assistants (PDRAs) and PhDs. • **Policy and practice**: Convening a stakeholder group. • **Levering funding and partnerships**: Aiming to leverage 10-fold initial investment through new partnerships and funding.
Key Challenges	• Loss of public trust in health data access. • Staff/partner organisation recruitment. • Challenges in accessing representative, longitudinal medicines data. • Challenges with data linkage. • A lack of involvement/engagement from diverse public and other stakeholder groups. • A lack of involvement from healthcare providers/practitioners.

How to get involved / PhD and training opportunities

• Non-clinical and clinical PhDs will be recruited across Year 1 and 2.

• PDRAs will be recruited throughout Year 1.

• Training programme (currently in development) that will co-design development plans for short and longer-term career goals and how to achieve and resource them.

• Contact: Stephanie Robinson Larkin, *S.Robinson-Larkin@liverpool.ac.uk*

Inflammation and Immunity

Key Activities in Year 1

• Develop whole-system capacity to map the epidemiology, healthcare utilisation and outcomes for common allergic and respiratory conditions for each of the UK nations in near-real-time.

• Use completed curated datasets for asthma, chronic obstructive pulmonary disease (COPD) and interstitial lung disease (ILD) across all four nations, in order to provide UK-wide burden of disease estimates.

• Extend recent work on the impact of respiratory illnesses on NHS winter pressures and a new curated dataset for respiratory syncytial virus (RSV).

• Begin to advance mapping of variations in care processes and health outcomes.

• Begin to identify and develop novel near-real-time linkages with other suitable datasets.

• Recruit of five more PhD students across the Programme.

• Create a strategy to develop, validate and test next-generation risk-prediction algorithms.

• Develop a PPIE strategy for the Programme.

• Extend engagement work to focus specifically on policy influence.

Workstreams	• WS1: Develop whole-system capacity to map epidemiology, healthcare utilisation and outcomes for common allergic and respiratory conditions. • WS2: Develop novel near-real-time data linkages. • WS3: Map variations in care processes and health outcomes and identify opportunities for reducing inequalities. • WS4: Develop, validate, and test - in large-scale, UK-wide clinical trials - next generation risk-prediction algorithms. • WS5: PPIE. • WS6: Policy influence and impact.
Wider Community Linkages	• Driver Programmes: Molecules to Health Records, Social and Environmental Determinants of Health, Medicines in Acute and Chronic Care and Big Data for Complex Disease. • BREATHE Hub. • TSE Pillar. • HDR Northern Ireland, HDR Wales, HDR North, HDR London, HDR Scotland.
Dataset Priorities	• Initial data sets and access: curated respiratory datasets, CPRD, DataLoch, NI HBS, SAIL Databank. • NHS England Secure Data Environment (NHS SDE) – in development. • Link UK Severe Asthma Registry to primary care data. • **Priority data need**: Regular, UK-wide GP data.

Technology and Infrastructure Priorities	• Whole UK (federated data) opportunities with linkage to key datasets in socioeconomic / genomic / imaging data.
Key Opportunities	• **Data strategy**: key goal replication of four nation curated inflammation and immunity-related health data. • **Technology and standards**: explore approaches to federation with Pillar 1, DARE UK. • **Streamlining Information Governance**: keen to work with Driver Programmes and Pillars to explore solutions to regulatory hurdles. • **Policy and practice**: Focus on mechanisms that can pivot to respond to urgent policy needs. • **Capacity building**: Second round of PhDs in Year 2, explore cross-Driver Programme opportunities. • **Leveraging funding and partnerships**: Continue to develop new partnerships and opportunities.
Key Challenges	• Access to data across all four nations, including permissions for onward sharing of curated datasets. • Regular (near-real-time) primary care data updates in all four nations. • **Harmonising datasets**: Consistency of methodology, coding, and data definitions alongside close collaboration and communication between analysts across the UK, to enable effective UK-wide analyses. • Ensuring that PPIE is fully embedded; plans are in development that will include significant learning from other studies. • Lack of team capacity to drive this work forward, either through skills, sickness or leadership focus on HDR UK reporting and meetings.

How to get involved / PhD and training opportunities

• A second round of PhD opportunities is likely in Year 2; interested in collaboration opportunities with other Driver Programmes.

• We are keen to discuss collaborations with other Driver Programmes on projects for early career researcher (ECR) posts.

• Contact: Wendy Inglis Humphrey, *winglis@ed.ac.uk.*

Molecules to Health Records

Key Activities in Year 1

• Explore with the UK Longitudinal Linkage Collaboration (UK LLC) and TSE Pillar 2 the requirements for multi-omic data TRE infrastructure development.

• Partner/explore partnering with UK-based multi-omic cohorts focused on non-European ancestries.

• Evaluate mass spectrometry glycosylation assays deployed in several UK population cohorts.

• Enrich the Genomics England TRE data, focusing on specific data modalities – developing working groups and expanding the HDR UK-Genomics England post-doctoral fellowship scheme.

• Co-develop a PPIE strategy and identify key public stakeholder groups for each activity.

• Explore partnerships with international cohorts to progress collaboration on informatics tools and resources.

• Develop more diverse and global cohorts - begin to evaluate methods for linking cohort participants to health records, and assess tools to ascertain causes of death reliably in existing low-to-middle-income (LMIC) countries' cohorts.

• Commence recruitment of Fellows and academic staff.

Workstreams	• WS1: Population system genomics and Electronic Health Records (EHRs). • WS2: Genomic medicine and EHRs. • WS3: Molecular informatics tools and resources. • WS4: Diverse and global cohorts.
Wider Community Linkages	• British Heart Foundation (BHF) Data Science Centre. • Big Data for Complex Disease Driver. • Health Data Research Hubs. • TSE.
Dataset Priorities	• Build on existing linkages to multi-omic cohorts to EHRs (NHS England, UK Health Security Agency, UK HSA). • Sub-licensing NHS England datasets to external researchers. • Genomics England TRE data enrichment: Patient Reported Outcome Measures (PROMS), administrative data, clinical imaging. • International cohorts: Focus on obtaining collected data, generating molecular data and outcomes follow-up.

Technology and Infrastructure Priorities	• Currently no TREs are available nationally with the necessary technical capabilities to enable efficient analysis of complex multi-omic EHR data. To explore with TSE Pillar and UK LLC.
Key Opportunities	• **Data strategy**: Enrich longitudinal health-related data within the Genomics England TRE, and multi-omics population cohorts. • **Data access**: Make multi-omic data accessible via the Gateway, integrate with the Cohort Discovery tool. • **Technology and standards**: With the TSE Pillar and UK LLC, we will create relevant methods and tools, such as federated analytics of multi-omics and EHR data across TREs. • **Trust and transparency**: Align to accelerate trustworthy data access, and PPIE strategy. • **Partnerships**: Build on major collaborations, such as UK Biobank, NIHR BioResource and build on major international collaborations such as the Global Alliance for Genomics and Health. • **Leveraging funding and partnerships:** Aim to leverage initial investment with external funding; and engaging multiple industry partners.
Key Challenges	• Not achieving priority EHR data linkages due to application delays or changes in governance. • Lack of a Secure Computing Environment required for data analysis of complex multi-omic data alongside EHR data. • Accessing representative, longitudinal medicines data. • Data linkage issues. • A lack of involvement/engagement from diverse public and other stakeholder groups. • A lack of involvement from healthcare providers/practitioners.

How to get involved / PhD and training opportunities

• We plan to extend a joint initiative with Genomics England's Clinical Interpretation Partnership (GeCIP) to establish a cadre of clinical and non-clinical Fellows in Year 1 to support ongoing GeCIP rare disease research.

• We plan to recruit a Clinical Informatician in the second quarter of Year 1.

Contact: Richard Houghton, *rh12@sanger.ac.uk.*

Social and Environmental Determinants of Health

Key Activities in Year 1

• Develop governance principles for sharing Unique Property Reference Numbers (UPRN) across TREs, avoiding re-identification in attribute data.

• Carry out a scoping review of literature on methods for protecting privacy.

• Advance best practice technical principles for importing large Geographic Information System (GIS) data across assets/TREs.

• PPIE activity on geoprivacy issues, UPRN and health data linkage.

• Achieve permissions to link to UPRN data in Imperial College London's Small Area Health Statistics Unit (SAHSU) (Workstream 2).

• Achieve permissions for Education and Child Health Insights from Linked Data (ECHILD) cohort and the UCL Kids' Environment and Health Cohort (KEHC) to link to UPRN.

• Achieve support and permissions for enhancing and linking historic cohorts – UK LLC cohorts with UPRN, Scottish cohort analysis of health outcomes, mortality and UV/sunlight exposure.

• Begin PhD and staff recruitment.

Workstreams	• WS1: Infrastructure, data, methods and governance. • WS2: Heath in Aging Populations. • WS3: Health and Development from birth to adulthood. • WS4: Linkage to historic cohorts.
Wider Community Linkages	• Molecules to Health Records, Inflammation and Immunity, and Medicines in Acute and Chronic Care Driver Programmes. • UK Longitudinal Linkage Collaboration (UK LLC). • CLS national cohort studies. • HDR Scotland, HDR Wales, SAIL Databank. • Pillars: Trust and Transparency, TSE. • HDR Alliance.
Dataset Priorities	• SAHSU Research Database. • ECHILD. • KEHC Research Database. • UK LLC. • Scottish historic Population Platform. • SAIL Databank.

	• Explore TRE capability for linking, generating and analysing large-scale environmental exposure data, and federated approaches – such as collaborating with the Natural Environment Research Council (NERC) Digital Solutions Programme, linking in with the TSE Pillar.
Technology and Infrastructure Priorities	
Key Opportunities	• **Data Strategy**: We are keen to explore cross-links in Workstream 4 with other Driver Programmes on the association between environment and health in older adults (respiratory, cardiovascular disease outcomes). • **Data access**: Widening access to SAHSU dataset. • **Technology and Standards**: We are keen to work with the Infrastructure and Services Pillar and HDR Alliance to inform harmonisation of methods for linking to place across TREs. • **Trust and Transparency**: We will work with the HDR Alliance on governance aspects and public support for geospatial data linkage. • **Partnerships**: Data custodians of health and health-related data, recruited cohorts and surveys, organisations that generate environmental data, TREs, regulators and policy bodies. (eg Office for National Statistics (ONS), Ordnance Survey, UKHSA, Geospatial Commission, NHS England, SAIL Databank, the Met Office, NERC). • **Levering funding and partnerships**: Priority is to leverage sustainable funding for ECHILD.
Key Challenges	• Public support: Risk that privacy groups and the wider public will not support the work. • Governance: Risk that the Health Research Authority will not approve the health data assets in the scope of this programme to hold UPRN. • Data providers: Risk that data providers (e.g. NHS England and ONS) do not provide UPRN to the health data assets within the required timescales. • TREs: Risk that changes to TREs could delay deliverables. • Progress against the relevant novel deliverables. • Risk of delay with staff and student recruitment.

How to get involved / PhD and training opportunities

• Two three-year research posts will be recruited at Swansea in Autumn 2023, providing the opportunity to pool resources across Driver Programmes to make five-year posts.
• Contact: Matthew Lilliman, *m.lilliman@ucl.ac.uk.*

Driver Programmes

Big Data for Complex Disease

Key Activities in Year 1

• Awarded six Big Data for Complex Disease (BDCD) PhD projects and studentships across the HDR UK community (five have been recruited, one will be readvertised).
• We are preparing to launch a BDCD fellowship competition led by Queens University Belfast with the aim of launching in September 2023.
• Enhance disease-based cohorts: A data linkage platform is under development.
• Expand our set of reproducible data curation and analysis pipelines, including phenotype algorithms.
• Discuss solutions to share/merge environments, data access and analysis, sharing of code and PPIE.
• Recruit key posts (eg, programme manager).

Workstreams	• WS1: Better predict diseases such as cancers and cardiovascular diseases, thereby improving screening, detection, early diagnosis and prevention strategies. • WS2: Better understanding of the inter-relationship between these complex diseases. • WS3: Better understanding the impact of inequalities in personal and system-level characteristics and national/ regional geography in order to influence and mitigate the negative impacts on incidence and outcomes of these conditions. • WS4: Capacity- and capability-building. • WS5: PPIE, Stakeholder and Policy Engagement. • WS5: Operational delivery, Data and Infrastructure.
Wider Community Linkages	• BHF Data Science Centre, DATA CAN Hub. • Molecules to Health Records, Inflammation and Immunity, and Medicines in Acute and Chronic Care Driver Programmes. • UK LLC. • HDR Wales (SAIL Databank), HDR Scotland, HDR South West.
Dataset Priorities	• National Institute for Cardiovascular Outcomes Research (NICOR) – broader linkage and access. • NHS England SDE: CVD-COVID-UK/COVID-IMPACT, DATACAN. Priority to enhance DATACAN Data linkage – and cancer data completeness and provenance (eg Cancer Outcomes and Services Dataset (COSD)). • Devolved nations' TREs via CVD-COVID-UK/COVID-IMPACT Consortium. • Virtual Cardio-Oncology Research Initiative (VICORI). • Social data: Current opportunity in Wales, in time there will also be opportunities in sub-national NHS England SDEs.

Driver Programmes

Technology and Infrastructure Priorities	• Broader and wider linkage to NHSE datasets currently available in DATACAN and CVD-COVID-UK/COVID-IMPACT. • Enhancing disease-based cohorts: Data linkage platform under development with SAIL Databank to hold cohort study data linked to health data.
Key Opportunities	**Data Strategy**: we are keen to identify opportunities to collaborate, eg there is lots of data linkage potential across the Driver Programmes, overlapping ambitions with others such as Inflammation and Immunity. • **Data access**: We are keen to enable cross-Driver Programme access to CVD-COVID-UK/COVID-IMPACT, DATACAN. • **Technology and Standards**: Work with Infrastructure and Services Programmes.
Key Challenges	• Reorganization could impact capacity and headspace at NHS England to make rapid progress with additional datasets, linkages and data curation which are required. • Risk of delay with staff and Fellow recruitment.

How to get involved / PhD and training opportunities

• The Fellowship programme will launch in September 2023, providing opportunities for cross-Driver Programme collaboration/projects.

Contact: Cathie Sudlow, *cathie.sudlow@hdruk.ac.uk* and Helen Tatlow, *helen.tatlow@hdruk.ac.uk.*

Infrastructure and Services

Infrastructure and Services

Technology Services Ecosystem	FAIR access to population-scale data at depth and breadth is needed to enable linkage of data from many custodians, and federated analyses are needed across SDEs for many health data researchers across the UK and globally. This Pillar aims to achieve this by bringing together a UK-wide team of leading technologists, data scientists from across academia, SDE providers, industry and the NHS, all of whom are committed to the assembly of an ecosystem of services. Embedding a collaborative, federated delivery model will enable greater patient and public benefit than any single organisation can achieve in isolation, whilst still maintaining autonomy of all involved.
Trust and Transparency	To deliver HDR UK's mission to enable data-driven research that improves people's lives, patients and the public need to have trust and confidence in the safe, secure and trustworthy access to, and use of, their data. Demonstrating trustworthiness and building public confidence on a national scale in a complex data landscape is challenging, but it is vital in order for the UK to achieve its research potential. This is particularly crucial in health data research. To support this mission, this Pillar underlines HDR UK's commitment to leading both meaningful involvement and engagement with the public, and a robust, transparent, trustworthy, and streamlined governance and ethics framework.
Useable Data	The aim across the Useable Data Pillar is to support researchers to identify datasets that meet their needs, and to minimise the amount of pre-processing or curation work they must do to make it 'fit for purpose'. The work will see the building of reuseable, open, and extensible software infrastructure through the 'Data Standards', 'Phenomics and Prognostic Atlas' and 'Transforming Data for Trials' Infrastructure Programme Workstreams, which will provide support across the data-to-analysis pipeline. Together, these workstreams focus on alignment of approaches to data and metadata (including phenotypes), with the aim of developing and driving adoption of consistent standards and formats.
Capacity Building	The Capacity Building Pillar will implement a unique range of health data training resources informed by the cutting-edge science and technology developed by our Infrastructure and Services and Driver Programmes, which will advance a new talent pool that can apply advanced health technologies and deploy those advances within the ecosystem. By accelerating the early adoption of emerging data skills, it is envisioned that HDR UK can make transformational impacts in health data science training content, research practice, training delivery approaches, and outreach and connections to different communities.

Infrastructure and Services

Technology Services Ecosystem

Key Activities

- Complete a plan for community training materials.
- Launch Technology Strategy paper and webpage summary.
- Gateway mark 1 placed in maintenance mode, Gateway Mark 2 launched.
- Complete beta testing for push and pull Fast Memory Access approach.
- Promote of cohort discovery tool to potential data sources.
- Promote and train tools to support mapping to OMOP from metadata and generate Extract, Transform and Load (ETL) scripts across the community.
- Undertake a consensus approach to develop priority data platform requirements and prepare proposals to respond to suitable funding calls.

Wider Community Linkages	• HDR UK Tech Community Principles developed and implemented. • New collaborations identified for creating/interoperating with a new piece of technology within the ecosystem. • Established requirements gathering feedback process, which is documented, embedded and sharable. • Software development principles reviewed and agreed by the community. • Options appraisal documented (and shared) for the use of Observational Medical Outcomes Partnership (OMOP) mapping to support enhanced Gateway search capability at a field level Tools to support mapping to OMOP from metadata and the generation of ETL scripts across the community have been promoted and trained.
Key Opportunities	• Implementation of HDR's Technician Commitment. • Completion of testing and rollout of pre-production system for Mark 2. • Promotion of the use of OMOP as a data standard across the Alliance and collaboration with the European Health Data Evidence Network (EHDEN) to fund groups to map to OMOP, plus development of UK Observational Health Data Sciences and Informatics (OHDSI) node.
Key Challenges	• The re-engineering of the Gateway could take longer than a year and the solution developed does not meet needs of community. • Re-writing and duplication of solutions/services. • Data governance amendments not approved Risk that data custodians will not see the value of onboarding their data. Risk of a 'scatter gun' approach in attempting to meet multiple use cases. • Trying to do too much across what are currently disparate activities and plans, and a risk of multiple stakeholders engaging in differing priorities.

How to get involved

- Contact: Emily Jefferson, *emily.jefferson@hdruk.ac.uk.*

Federated Analytics

Key Activities

- Initial PhDs offered to candidates.
- Initial Driver Programme requirements gathered.
- SAIL and PIONEER to utilise TRE-FX (an application stack for Federated Activities being developed for TREs or SDEs) with Five Safes RO-Crates in a production environment.
- TRE-FX with Five Safes RO-Crates, though co-development with BC Platforms able to be a production alternative to BC Link.

Wider Community Linkages	• Transition programme aligning DARE UK into core programme. • Opportunities for international collaboration assessed. • Community established across HDR UK for Federated Analytics. • Individuals from across the Driver Programmes identified to be part of a wider network.
Key Opportunities	• Prototype of HUTCH Five Safes RO-Crates developed. • Roadmap for international engagement established.
Key Challenges	• Driver Programmes may develop their own solutions. • Balancing international opportunities with national delivery. • Uncoordinated interest from industry.

How to get involved

- Contact: Phil Quinlan, *philip.quinlan@nottingham.ac.uk* and Carole Goble, *carole.goble@manchester.ac.uk.*

Trust and Transparency

Key Activities

- Work towards building streamlined data access information governance models through the Pan-UK Data Governance Steering Group.
- Work towards data custodians adopting a single core set of questions that need to be submitted by researchers for access to data. This objective will be necessary to achieve the vision set out in the Department of Health and Social Care data access policy published in June.
- Create a set of core principles for data access in TREs, including identification of customisable controls relevant to some TREs - the principles will be developed following a benchmarking activity involving review and comparison of data access agreements for established TREs and will be mapped against the Five Safes Framework.
- Collaborate with Public Advisory Board (PAB) members to create Transparency Standards to be adopted by the Alliance and activities to promote adoption.
- Creation of a standardised Data Depositing Agreement (DDA) to streamline the contractual process between data contributors and TRE host organisations.
- Information Governance management for the set-up of the BHF DSC cardiovascular disease and diabetes cohorts TRE.
- Promote inclusivity in PPIE activities by attending three festivals that target under-represented groups over the next 18 months and evaluate the impact.
- Engage with key stakeholders to identify community groups to engage in health data-related activities to increase representation.
- Create a community that will focus on the legal challenges of international data sharing.
- Work with public members and PPIE professionals to update the HDR UK PPIE strategy for the second five years.

Wider Community Linkages	• Create a single map of all approvals required to gain access to data; an Action Force has been set up to deliver this piece of work. • Map roles and responsibilities for all parties involved in data access in TREs against UK General Data Protection Regulation (GDPR) to add clarity and consistency across contractual arrangements and obligations. This will then be turned into a functional tool to aid in division of responsibilities and to signpost which contracts will be required. • Guidance for use of the Data Access Agreement (DAA) templates will be developed including review and input from public members. • Pilot at least one community engagement event and evaluate impact in the next 18 months. • Launch pilot campaign targeting focus group regions. • Collaborate with PAB members and stakeholders interested in advancing PPIE practices, such as 'Public Engagement in Data Research Initiative' and 'Shared Commitment to Public Involvement' partners.

Key Opportunities	• Promote standardisation of data access governance processes. • Streamline data access contracting. • Increase the number of HDR UK Voices members by 100% (203) in the next year • Launch Public Engagement in Data Research Initiative (PEDRI) pilot campaign and engage at least 500 members of the public. • PAB can actively contribute to improving transparency in data use, enhancing public trust in health data. • Partner with organisations and initiatives focused on PPIE practices brings access to a wealth of expertise and resources in the field. • Broaden the participation of under-represented groups in health data research and efforts. • Incorporate diverse perspectives and expertise in the new HDR UK PPIE strategy.
Key Challenges	• There can be confusion and lack of consensus around 'Data Controllership' and many parties may be Controllers in the process of data access in TREs. For this reason, we will define responsibilities by the following roles: Data Contributor, Host Organisation, TRE Platform, TRE Service Provider, User Organisation, Approved Researcher. • Established TREs/SDEs and data custodians may be reluctant to change processes in place due to operational/resource impacts and/or lack of agreement or understanding of the benefits of standardisation. • Maintaining trust in data use while promoting access to health data can be a delicate balance, with concerns about privacy and security. • Stakeholders involved in advancing PPIE practices may have diverse interests, priorities, and goals. Balancing these interests and aligning them with HDR UK's objectives can be challenging. • Establishing clear metrics and methodologies for assessing the success of the pilot activities may be difficult. • Failing to identify the right groups can result in ineffective outreach and a lack of meaningful participation. • Striking a balance between the input from the public and the PPIE professionals involved in the development of the new HDR UK PPIE strategy.

How to get involved

• Contact: Cassie Smith, *cassie.smith@hdruk.ac.uk.*

Useable Data

Key Activities

• Maximise benefits from UK implementation of EHDEN partnerships:
• Publish a review of the current landscape and survey results.
• Increase awareness and adoption of recommendations for collection of ethnicity data.
• Map the landscape of international data standards bodies.
• Identify data priorities, informing data strategy.
• Support improvements of data quality and completeness of diversity-related data in NHS settings (in alignment with Alliance activities).

Wider Community Linkages	• Establish an OMOP CDM Special Interest Group, as a sub-group of the Alliance Data Officer Group (Inaugural meeting in the first quarter, two additional meetings planned throughout the year). • Engage new partners including NHSE Health and Disparity Unit; Race and health Observatory; Understanding Patient Data; Wellcome. • Convening the Data Officer Group, a community forum to discuss data standards, data quality, metadata, ontology and terminology.
Key Opportunities	• Map the landscape: understand and analyse the current level of adoption of the OMOP Common Data Model including desk-based research, individual interviews and a survey. • Support data strategy integrated objectives. • Identify data standards priorities with input from Driver Programme, UK and international investments, Alliance partners, HDR UK regions.

How to get involved

• Contact: Monica Jones, *m.c.m.jones@leeds.ac.uk*, and Alex Knight, *alex.knight@hdruk.ac.uk.*

Phenomics and Prognostic Atlas

Key Activities

- Develop, validate and deliver the Disease Atlas and disease phenotypes as a new body of methods and knowledge built at nationwide scale to inform research across HDR UK, clinical practice and health.

- Advance disease phenomics and the Disease Atlas through (a) hospital-wide approaches using unstructured and structured data to generate disease phenotypes in all patients and (b) using large language models (LLMs) trained in nationwide data (initially structured).

- Seek opportunities for sustainable growth of products and services from Atlas, the Phenotype Library and Foresight, through external funding and commercialisation.

- Prototype initial demonstration of value / user journeys on the Disease Atlas interactive website.

- Develop an enhanced version of Phenotype Library, holding PPA content and integrations with the PPA website.

Wider Community Linkages

- Develop the governance to assemble and reward a single national productive, interdisciplinary team to deliver common goals of PPA, as funded by HDR UK; seek synergies across HDR UK including with Data Standards, Trials, BHF DSC, Big Data for Complex Disease, and the pre-existing Atlas and CogStack/Foresight teams. We will seek to identify clear relationships, communications and lines of accountability, noting that these teams comprise people funded from different routes within and outside of HDR UK.

- Prototype development of software services/components with open APIs e.g. results service, clinical guidelines, papers service, people service, publications service, ontology service, DOI service (final quarter).

Key Opportunities

- Horizon scanning and planning how to attract additional funding and take advantage of the opportunities presented by recent advancements in LLMs.

- Draft of a series of clinical specialty papers for publication arising from Phenomics and Prognostic Atlas (PPA), eg given our close dialogue with Big Data for Complex Disease, an initial focus might be in cardiovascular and oncology specialties.

- Begin to systematically add phenotypes with molecular markers, focusing on the requirements of clinical trials (final quarter).

- Formalise core Atlas ontology, including specialties, into an international standard such as the Open Biological and Biomedical Ontologies (OBO) or similar format (third quarter).

Key Challenges

- Evolving PPA in the light of developments in LLMs: The risk is that we do not dynamically update and prioritise the aims and direction of the theme, in the light of national and international advances - eg rapid developments in the last few months in generalised medical AI may supersede, or challenge, some elements of the submitted proposal (large language models (LLM) are not mentioned at all in the QQ2 submission) Mitigation: establish a PPA Board, tasked with dynamically updating the priorities of the theme within the overall framework.

- Learning from related national and international initiatives: the risk is that we do not learn from diverse related national initiatives (eg Open Prescribing) and international initiatives (eg Global Burden of Disease). Mitigation: the PPA Board will use evidence comparing initiatives to inform decisions about priorities for the team's time.

How to get involved

- Contact: Harry Hemingway, *h.hemingway@ucl.ac.uk.*

Transforming Data for Trials

Key Activities

- The Trials Stakeholder Prioritisation Forum, working with the Alliance, will establish a forum which brings together the key stakeholders in a four-nation approach, including data custodians (eg NHS England), users of data in the trials community (eg trial teams) and users and interpreters of the findings of trials (eg the Medicines and Healthcare Regulatory Authority, MHRA).

- PPIE: this work-stream contains two elements: (1) patients and the public will be involved in developing the strategic aims and planned output of the Transforming Data for Trials program; (2) the program will identify or develop PPIE output relating to the use of healthcare systems data in clinical trials (these outputs will be accessible via the route-map).

- A thematically curated catalogue of case studies will be developed from the spectrum of trials from across the UK and international trials, if relevant.

Wider Community Linkages	- Training resources will be developed, aiming to enable trial teams to use healthcare systems data. - Data utility comparisons are methodological research studies in which the utility of study outcomes identified from healthcare systems data is assessed compared to outcomes ascertained through the traditional 'gold standard' trial methods.
Key Opportunities	- Route-map for trials using healthcare systems data: The 'route-map' for healthcare systems data which will be parallel and complement the existing, widely accessed NIHR Clinical Trial Toolkit route-map. - Knowledge transfer, exchange and mentoring -this work-stream will establish effective methods knowledge transfer between institutions in relation to use of healthcare systems data. - Demonstrating the data integrity and provenance of healthcare systems data for clinical trials.
Key Challenges	- Timely appointment to required posts. - Duplication of content with other organisations working in this space. - Ability to deliver required training (funded across five years of the programme) while supporting release (eg after 18-24 months) of the matched route-map. - Failure to engage with the key relevant stakeholders. - Insufficient staff and issues in appointing people to posts. - Data providers don't engage.

How to get involved

- Contact: Marion Mafham, *marion.mafham@ndph.ox.ac.uk* and Matt Sydes, *matt.sydes@hdruk.ac.uk.*

Capacity Building

Key Activities

- **Growing the health data learning curriculum through HDR UK's Futures Learning Platform**: This content builds on the HDR UK Futures platform that has benefitted from the wide support of the HDR UK community. We are shaping the design and delivery of innovative new training programmes to accelerate learning, build new capabilities and promote the sharing of best practice across Alliance partners.

- **Technician Commitment**: In 2022 we made the decision to sign up to the Technician's Commitment and published our action plan in June 2023 in which we described the actions we would take to support technicians both internally and externally. We will be deploying this action plan from September 2023 onwards.

- **HDR UK Turing/Wellcome PhD programme**: Our flagship four-year PhD programme aims to train the future leaders in health data science. We have recruited four of the five cohorts with the final cohort due to be recruited for an October 2024 start date. Cohort four will start their training year in October 2023.

Infrastructure and Services

Wider Community Linkages	• **Developing diverse pathways in health data science.** • **Supporting Driver Programmes, the Alliance and Infrastructure and Services**: Training champions will take up secondment positions lasting between three and 12 months, which will be embedded within Driver Programmes and Infrastructure and Services. They will identify training opportunities and requirements both within individual and across programmes and produce cutting-edge training tools to address these requirements. • **Cross-sector transition grants**: The Capacity Building Pillar will seed-fund up to 12 cross-sector transition grants to enable PhD students, HDR UK Fellows or industry-based collaborators to undertake three to six-month placements in different sectors (eg, academia to industry or vice-versa). • **Masters Scholarships**: HDR UK is partnering with charities and industry to offer master's degree scholarships to quantitative students. • **Medical School Data Science Survey:** We have issued a survey to medical schools across the UK in conjunction with the Academy of Medical Sciences and NHS England. We have recruited an intern from the Black Internship Programme to develop the analysis for this survey during the summer of 2023.
Key Opportunities	• Supporting future health data science leaders. • Black Internship Programme – The target is to recruit 750 interns over five years; 96 were recruited in 2022/24 for the Black Internship Programme and this number continues to grow year on year. We secured external funding for eight times more internships on this programme (via the EPSRC Digital Health Hubs). • Alumni Network – Our target was to have 5,000 members over five years and to date we have recruited over 350 members from core programmes (fellows, PhD students, Masters students and interns). Furthermore, network leads have been established and a programme of career building activity is underway.

How to get involved

• Contact: Sarah Cadman, *sarah.cadman@hdruk.ac.uk*

Infrastructure and Services

UK Health Data Research Alliance

Key Activities

• Become a recognised and influential network of partners to enable trustworthy use of health-related data for research and innovation through four main 12-month priorities:

  • Sustain partnership with NHS England to inform and maintain alignment with secure data infrastructure developments and structural changes.

  • Maximise benefits from the UK implementation of the EHDEN partnership.

  • Maintain and extend alignment with other key national organisations e.g. ONS, Administrative Data Research UK, Alan Turing Institute, Association of Medical Research Charities (AMRC), Association of the British Pharmaceutical Industry (ABPI) and the Bioindustry Association (BIA).

  • Extend Alliance membership to non-data custodian organisations.

Wider Community Linkages	• Grow and strengthen the Alliance to include major health-relevant data controllers, industry and other representative bodies. • Support development, implementation and adoption of recognised standards, best practice and tools to accelerate trustworthy access to, and use of, data. • Develop and implement a communication strategy to improve engagement, diversity and growth. • Taking on a convening role, we will develop and implement principles and best practice tools and policies in various streams including trust and transparency, technology and data. • Encourage recognition and adoption of standards (eg data use registers, TRE principles and the Five Safes Framework) in the UK and beyond.
Key Opportunities	• Enable Alliance members to contribute towards a UK-wide strategy to accelerate access to, and use of, data. • Maintain and extend alignment with other key national organisations including AMRC, BIA and ABPI. • Sustain partnership with NHS England to inform and maintain alignment with secure data infrastructure developments and structural changes.

How to get involved

• Contact: Paola Quattroni, *Paola.Quattroni@hdruk.ac.uk.*

Regional Networks

Regional Networks are designed to leverage expertise and forge partnerships across the regions. The geographical representation of the Network continues to be invaluable. It enables vital contributions from leading researchers and research organisations to focused, UK-wide research programmes, and opens up regional capabilities in infrastructure, data, services, the NHS, PPIE and innovation.

Each of the Regional Networks has:

- Clear leadership, which continues to build partnerships, activities and innovation across the respective regions as well as an overarching ambition to improve national capabilities.
- Well-established links with colleagues, including key local data custodians, government, the NHS and public health, localised specialist partnerships, and skills across the four nations.

The substantial differences between neighbouring nations, and regions within a nation, imply complex interactions between biological (eg genetic), environmental and social factors. Working with health data science systems in these regions will be essential if the UK is to address these inequalities.

Cambridge

Key Activities
- Industry engagement with pharmaceutical companies (eg AstraZeneca, GSK) to map out areas of potential synergy on which to collaborate via our PhD programmes.
- Embed PPIE in the full life cycle of research by engaging with the diverse participants of Cambridge-led research cohorts and seeking their input on a wide range of documents and workshops.

Wider Community Linkages	- A monthly Cambridge-hosted multi-disciplinary seminar series, which is open to the entire community. - We mobilise data resources at the regional level to complement national resources, and ensure connectivity with cognate activities, locally and nationally: the BHF Data Science Centre, EHDEN, Cancer Data Driven Detection (CD3) initiative and the East of England sub-national SDE. - We support the University of Cambridge's health data science master's programme and PhD studentships to be hosted at Cambridge will be funded through the Molecules to Health Records Driver Programme.
Key Opportunities	- Shape development of the East of England sub-national SDE. - Connectivity across the Cambridge ecosystem to ensure alignment on TRE infrastructure.
Key Challenges	- Complexity and risks of accessing and mobilising data from local / national sources. - Risk of fragmentation of data assets across different TRE infrastructure, which could lead to difficulties in data access or limit analyses.

How to get involved
- Join our monthly multi-disciplinary seminar series and help spread the word for our 2024 intake of HDR UK-funded PhD students.

  Contact: Richard Houghton, *rh12@sanger.ac.uk.*

London

Key Activities

- Inter-regional development of methods and applications of 'whole hospital' health data research to inform clinical practice, health policy, and research.
- Develop specific 'methods pilots' to generate inter-regional proofs of concept for use as building blocks of wider value across the HDR community; in addition, stimulate and incubate pathfinder projects to understand the implications of using whole-system approaches for research, clinical practice, and policy.

Wider Community Linkages	• Forge partnerships across regions to answer fundamental questions together and deliver better research and care through inter-regional engagement activities and workshops. • Incorporate novel inter-regional insights from analyses already underway in national TREs to help prioritise pilots, specify knowledge gaps and add value to whole hospital approaches.
Key Opportunities	• Focus on whole hospital data approaches as a complementary effort to other regions. • Highlight benefits of enabling access for researchers to structured and unstructured data in whole hospital systems. • Identify partner hospitals which have permissions and capabilities for whole hospital analyses, to demonstrate what is possible.
Key Challenges	• Risk of access to hospital data being withdrawn or de-prioritised, possibly mitigated by inter-regional prioritisation of research areas of value to the NHS. • Comparisons of hospitals' data may instigate unnecessary feeling of competition, which carries risks of negative views.

How to get involved

- Contact: Harry Hemingway, *h.hemingway@ucl.ac.uk*, and, for inter-regional capabilities and whole hospital approaches to health data research, Sinead Langan, *sinead.langan@lshtm.ac.uk.*

Midlands

Key Activities

- Scale of health data research activity across Midlands requires a mapping exercise to understand all Health Data Science Projects.
- The first quarterly insight sharing day is planned for September 2023.
- We are developing a PPIE strategy to ensure that patients and the public are involved in our work.

Wider Community Linkages	• Maximising opportunities for our region through the HDR UK Midlands Regional Community Platform. • Develop a training strategy to upskill the region. • Work closely with other HDR UK regions to maximise opportunities. • Connect our region with other components of HDR UK, such as Driver Programmes and Infrastructure and Services.
Key Opportunities	• Upskilling and improving the knowledge and capabilities of our workforce by learning from each other. • Continue to grow our membership base.
Key Challenges	• Key regional groups, networks and organisations are not involved in our work. • Impact, outputs and outcomes not fully understood or reported.

How to get involved

- Contact: Kevin Dunn, *k.w.dunn@bham.ac.uk.*

Regional Networks

North

Key Activities

- Ensure new discoveries using health data translate into improvements in health and wellbeing, by establishing strong links with our Integrated Care Systems (ICS), SDEs, TREs, NHS leaders, public health bodies, data custodians, charities, companies and academic groups engaged in health data research.
- To catalogue TREs, SDEs, and Federated Data Platforms (FDPs) across the north of England to provide an overall view of these assets across the region, and to facilitate data sharing across multiple systems, linking to the HDR UK Technology Services Ecosystem (Pillar One Programme).

Wider Community Linkages	• Leverage further funding by submitting collaborative regional applications and build research capacity by supporting doctoral / postdoctoral fellowship applications - the focus of applications will be reflective of the research expertise of the North, and DHSC priorities for improving health outcomes in health and social care. • Work with other HDR UK regional networks on a shared purpose, to enable access to secondary care data, host knowledge exchange activities to encourage this work, and leverage further funding.
Key Opportunities	• Shape and support NHS and local authority workforce analytical capabilities within the region through co-developed knowledge exchange activities, including joint workshops and seminars.
Key Challenges	• Potential capacity constraints on the North's Associate Directors ability to submit additional funding applications owing to clinical duties. • Stakeholders fail to realise synergy and value of integrating with the programme.

How to get involved

- Contact: Steph Robinson-Larkin, *s.robinson-larkin@liverpool.ac.uk.*

Regional Networks

Oxford

Key Activities

- Strengthen the HDR network across the region and beyond, by building stronger links with existing health data research groups and establishing new collaborations to increase efficiency and translational impact.
- Build on existing initiatives to expand the HDR infrastructure and improve secure access for approved research data that also protects the public interest.
- Continue building a strong health data science workforce by identifying and fostering new talent in internships and PhD students and supporting existing staff with their career aspirations.

Wider Community Linkages	• Remain highly engaged with NHS England and ensure the emerging new system is fit for purpose and much more efficient than previous iterations. • Recruit to a variety of schemes for internships such as UNIQ+ and the HDR UK black internship programme. • Initiate a programme of seminars and workshops to share knowledge and expertise with the health data research community.
Key Opportunities	• Establish new and enhance existing communication channels to ensure all working in health data science in our region are aware of opportunities for collaborative work. • Increase awareness, and therefore, number of users in all health data research secure data environments. • Ensure strong patient and public engagement throughout our programme of work. • Improve the student experience and success rates. • Early, mid and senior researcher applications for research funding and success rates.

How to get involved

- Contact: Eva Morris, *eva.morris@ndph.ox.ac.uk*, and Cecilia Lindgren, *cecilia.lindgren@bdi.ox.ac.uk.*

Regional Networks

Scotland

Key Activities

- Engage with Research Data Scotland to increase the diversity of accessible data at high quality to researchers; reduce latency in data workflows; introduce appropriate, responsible automation into data cleaning and analytics; and collaborate on cross-UK data standardisation.
- Encourage cross-working with the development of the HDR UK TSE (Infrastructure Pillar 1) to propagate best practice across the health data research network.

Wider Community Linkages	• Launch of the Advanced Computer Science Summit, hosted in Edinburgh under the stewardship of HDR UK: The aim is to encourage those in our HDR networks who have an interest in cutting-edge applications of computing to consider the fast-developing wave of new techniques, while forging new relationships (including industry) that could help us inhabit this new space.
Key Opportunities	• Connecting with the *five DARE UK Driver Projects* and determining how HDR UK Scotland can work with them.
Key Challenges	• Failure of key stakeholders to engage in our data management work and Advanced Computer Science Summit. • Dependencies on the TSE as a separate project with its own timescales may impact our ability to deliver across the region.

How to get involved

- Contact: Serena Tricarico, *s.tricarico@ed.ac.uk.*

Regional Networks

South West

Key Activities

- Leadership within and development of the Great Western (GW) sub-national SDE.
- Continued role in development of the NHSE national SDE and OpenSAFELY.
- Training and capacity building in health data in South West region- new Member Plymouth University to co-lead.
- Leadership and continued development of the UK LLC.

Wider Community Linkages	• Leadership roles in Big Data for Complex Disease and Medicines in Acute and Chronic Care Driver Programmes. • BHF Data Science Centre. • SAIL Databank /UK SeRP (HDR Wales) hosting UK LLC. • Work closely with other HDR UK regions to maximise opportunities.
Key Opportunities	• Bristol-based post and PhD to support Medicines in Acute and Chronic Care and Big Data for Complex Disease Driver Programmes. • Development and wider access to UK LLC. • Year 1 South West regional capacity building events.
Key Challenges	• Reduced spending power. • Potential for delay to GW SDE outside regional control. • Potential delay to PhD/staff recruitment.

How to get involved

Contact: Lizzie Huntley, *lizzie.huntley@bristol.ac.uk.*

Regional Networks

Wales

Key Activities

- Create and maximise opportunities, dissemination and outreach to further build an agile community and PPIE.
- Work with the Driver Programmes includes: Medicines in Acute and Chronic Care, Social and Environmental Determinants of Health, Inflammation and Immunity, Phenomics.

Wider Community Linkages	• Dementia Platform - Building on the dementia platform for grants in the Medicines in Acute and Chronic Care Driver Programme - Dementias Platform UK (DPUK) Data Portal. • Digital Medicines Transformation Portfolio - build on the new digitalising medicines plan from the Minister for Health and Social Services of a fully digital prescribing approach in all care settings in Wales: the Digital Medicines Transformation Portfolio - Digital Health and Care Wales (NHS Wales). • Anticipated organisation of research development groups and webinars with the other regional leads in HDR UK. • Iteratively engage and run workshops to understand requirements with potential users of the Phenotype Library across HDR UK and specialties (second quarter onwards).
Key Opportunities	• Administrative Data Research (the region to focus on bringing together administrative data and health data), eg the July workshop with the four police forces in Wales to discuss sharing of police data for linkage with health data; working with justice data (courts, prisons) and health data. • ADR Wales (ADR UK). • UKSerp - building on the expertise to host data from outside Wales (SeRP UK).
Key Challenges	• Failure to recruit researchers with appropriate skills – this could be mitigated by extending training of the interns in the Black Internship Programme if appropriate, as well as training of Masters students graduating from the Health Data Science course delivered in Swansea and affiliated with HDR UK.

How to get involved

- Contact: Alysha Morgan, *a.l.morgan@swansea.ac.uk.*

Partnership Programmes

Partnership Programmes

Hubs

The *UK Government's Industrial Strategy (2017)* recognised the untapped potential of the UK's health data as an opportunity to boost economic and health outcomes. As part of the Life Sciences *Data to early diagnosis and precision medicine challenge* (D2EDPM), HDR UK was funded by UKRI to establish the £37.5 million Digital Innovation Hub (DIH) Programme via the Industrial Strategy Challenge Fund to develop infrastructure to support the Hubs and UK Health Data Research Alliance members. The aim was to enhance routine NHS data and the UK's rich cohort, registry, and research data and make it available to industry, researchers, and innovators to use for impactful research. Following a consultation with 1,200 across four-nations, HDR UK published a *DIH Programme Prospectus* in May 2019 that set out proposals to establish three new capabilities to address challenges faced by industry and academic researchers in accessing and using health data. These were:

- The UK Health Data Research Alliance (the Alliance): an independent, legally non-binding alliance of health data providers, custodians and curators who develop and share standards, policies, and best practice to demonstrate trustworthiness in health data use for research to improve human health.

- The Health Data Research Innovation Gateway (the Gateway): a web-based platform to discover and request access to UK health datasets for research and innovation.

- Seven Health Data Research Hubs (the Hubs): UK-wide centres of excellence which focus on data curation and create the expertise, tools, knowledge, and ways of working to maximise the insights and innovations developed from health data.

The Hubs are very much centres of excellence with expertise and tools developing data to provide insights. Hubs support external users from across the NHS, academia, and industry to use data for research and innovation. Each Hub holds topic-specific knowledge, expertise and data in different domains or data types. In 2021, the UKRI Medical Research Council (MRC) supported the award of two more Hubs, *Alleviate* (pain) and *DATAMIND* (mental health), as part of a growing Hub programme. There are currently nine Hubs that provide coverage across multiple disease areas and data sources. Collectively, the Hubs work as a network to adopt similar ways of working and tackle challenges in data science through being members of the wider UK Health Data Research Alliance. Below are updates from seven of these Hubs.

Partnership Programmes

Alleviate

Alleviate Pain Data Hub has been an active Hub for over 20 months. The core activity is to transform pain data cohorts to the OMOP Common Data Model (CDM) and integrate with the Cohort Discovery Tool (CDT) on the Innovation Gateway.

The Milestone 2 assessment was well received, and the Hub showcased successful PPIE across all activities, with two PPIE leads on the project team and a pain community of more than 160 (as of August 2023). The Hub is using the expertise built from the Health Information Centre (HIC) at Dundee and the CO-CONNECT project, with a good awareness of standards and continued development of open-source tools to streamline data mapping in collaboration with University of Nottingham. The Hub has engaged extensively with the Advanced Pain Discovery Platform (APDP) community with *workshops*, national and international conference presentations (British Pain Society, APDP Annual Conference, NeuPSig, EFIC) and meeting contributions (CAPE, PAINSTORM). The Hub has delivered 20 pain research cohorts on the Gateway, knowledge sharing of the CDT across the wider pain community, an active patient group of People with Lived Experience (PWLE) and a focal point for data strategy in the APDP.

September is pain awareness month and Alleviate will share a series of video stories from our Patient and Public Involvement and Engagement members through their lived experience with pain. Alleviate will also host a webinar for the wider APDP community, showcasing some of our member's personal journeys of living with pain. We also plan to run a UK-Wide questionnaire-based survey to gather information around people's experiences with Chronic Pain.

Over the next six months, Alleviate will focus on long-term sustainability. Services will be strengthened in the coming months, with domain expertise applied to use the CDT for real-world data needs. There will be collaborations with other pain-related or data-rich communities to solve common problems. This will allow the Hub to facilitate wider APDP research community work to develop phenotypes for sharing through the library. The Hub will continue to work with the Technology Team to develop the Cohort Discovery Tool while growing the available pain relevant datasets.

Partnership Programmes

DATAMIND

The Data Hub for Mental health INformatics research Development (*DATAMIND*) aims to advance mental health research through a step change in the visibility and accessibility to NHS, administrative and longitudinal data. This will be achieved through greater discoverability, annotation, harmonisation, and advanced analytics by ensuring the inclusion of diverse groups of under-represented individuals with the greatest clinical need.

The Hub has focused on four core activity areas and has also identified four challenge areas that have guided research. These are: Children and Young People, Excluded and Underrepresented Groups, Interfaces between Physical and Mental Health, and Severe Mental Illness (SMI). Specifically building on the MRC Pathfinder community, the Hub has also been working on six identified Road Builder Innovations which are being delivered at different Milestones over the course of the project. These will potentially become Hub Core Activities over time, working across and extending further into Challenge Areas. The Road Builder Innovations are:

- Discoverable Schools.
- Discoverable excluded and under-served groups.
- Linking Physical and mental health in SMI.
- Widening availability of mental health care text analytics capabilities.
- Digitally Enhanced Trials.
- Drug discoverability.

To date, the DATAMIND team has focused on workshops and the identification of training needs for early career researchers, followed by provision of courses. The data literacy course has been developed alongside our PPIE group to develop the capacity of people with lived and living experience of mental health in Mental Health Data Science research. This online course will be released soon. There is an extensive range of tools being developed to support a range of research needs: Core Mental Health DataSet (CMHDS) collection tool; equity audit tool for clinical trials; Mental Health Text Analytics Cloud (MH-TAC); VELA tool to simplify information extraction from large, linkable, mental health data sources; and PHENOMIND for making phenotype library code lists and algorithms available for research data. The partnership with the National Institute for Health Research Clinical Research Network (NIHR CRN) is important for widespread uptake.

There are efforts to integrate the Catalogue of Mental Health Measures and Landscaping International Longitudinal Data project with the Innovation Gateway. The Welsh work on hard-to-reach groups shows real value of different types of data assets, and it is good to see that the Natural Language Processing work is being built out. The Hub established an Industry Forum to connect DATAMIND with several industries, including the pharmaceutical sector, to explore the possibility of developing work that would address mental health needs. During these discussions, various gaps in the field have been identified such as the possibility of using genomic data collected as part of a pharma-sponsored clinical trial to understand how antidepressants work and why they work better for some people than for others. This also led to the application of a Wellcome Trust call which was awarded in December 2022, contributing to the sustainability of DATAMIND. Overall, the Hub is showing excellent progress and integration with other work/investments.

Over the next six months, the Hub will continue to expand its online presence with its *website*. Two key additions that will be added to the DATAMIND website are an online data literacy course developed in collaboration with McPin, DATAMIND, and the Super Research Advisory Group (SRAG), covering topics

like patient rights, research, and risks. Additionally, the DATAMIND glossary, created with assistance from the PPIE team, will provide easily understandable explanations of essential terms in mental health research and data analysis, fostering effective communication and collaboration among stakeholders. The DATAMIND PPIE group will continue their involvement in co-producing and developing guidelines and strategies related to privacy and data sharing and other ongoing documents to aid the field of mental health and its community, including industry. The next Industry Forum will be held late in 2023 with PPIE for them to engage directly with industry partners. The next data science events are also in preparation for late 2023 and will incorporate the feedback received from early career researchers and attendees of previous workshops. They will also focus on working with HDR UK Futures and the MQ Mental Health Research charity for engagement opportunities and to develop training modules for Continuing Professional Development-style-learning. Work continues to onboard datasets through the Gateway.

Discover-NOW

Following the confirmation of funding by NHS England for the development of a London-wide SDE, *Discover-NOW* is being scaled London-wide to create a linked primary and secondary care record for the 10 plus million patients registered in the five London Integrated Care Systems (ICS), creating a truly unique SDE available for research and development. Discover-NOW also has plans to scale its operational service offer to the whole of London. A series of stakeholder interviews, public deliberations, and cross-workstream workshops will take place to ensure co-design of the new Discover-NOW service offer and underpinning commercials. The Hub aims to continue our successful collaboration with partners around AI, real world trials, retrospective analyses and population health management project, while also exploring a self-service or direct access model.

The Hub continues to be a resourceful asset to research active institutions and commercial organisations. To date, over 300 data access applications have been completed for Discover-NOW. Other work includes:

- Supporting high impact projects including the development of the London Asthma Decision Support tool (LADS). This combines asthma population, clinical care, financial and wider determinants of health data (including air quality data) from across two London ICS. This tool is being used across NWL to support clinical and financial decision making for precision care planning from practice to ICB level and has been submitted for a HSJ transformation award.
- Investigating survival and health economic outcomes in heart failure diagnosed at hospital admission versus community settings found that diagnosis of heart failure through hospital admission continues to dominate and is associated with significantly greater short-term risk of mortality. This work was *published in BMJ Health and Care Informatics.*
- The NWL Health Research Register, which consents patients for contact about future research opportunities, has achieved 80,000 sign-ups, which is increasingly being used for commercial clinical trials and is recognised as a quicker way of finding and referring patients.

Meanwhile, a priority for the business development team is to engage with current partners and reach out to additional organisations in the health technology, pharmaceutical, consultancy, Contract Research Organisation, and academic spaces, to not only make them aware of the potential plans and offers of Discover-NOW, but to derive insights around the needs and requirements of these key partners for delivering high impact health projects. The Hub has become an alliance partner with the Paddington Life Sciences Group and have ringfenced NHSE SDE funding to create an important collaboration with the Kings College AI Centre around imaging data. Based upon recent contracts, potential use cases include the deployment of a medical history model to determine risk of future health complications, patient identification and recruitment for a cardiovascular metabolic trial, retrospective analysis of a CVD patient pathway, effectiveness of a medical device supported with AI, and a medicine profiling and optimization study.

Partitnership Programmes

PIONEER

At *PIONEER*, access to detailed, individually linked health data, supported by expert consultancy and highly trained staff, is transforming healthcare for our population. The PIONEER Hub continues to facilitate groundbreaking research and innovation in acute care. By providing researchers with access to comprehensive and diverse datasets, the Hub has enabled discoveries that drive the development of new treatment modalities, care pathways and medical technologies. There is now a scalable, interoperable system that can be federated to deliver high agility regionally, nationally and internationally. PIONEER is a model that the team designed and own.

PIONEER has achieved remarkable success by serving over 80 requests across eight sectors, including four from industry partners. These achievements have been recognised and supported by prestigious funders, resulting in grant awards totalling >£29 million. PIONEER has also made significant contributions to data accessibility, offering 45 platinum-rated datasets through the HDR UK gateway. Moreover, the Hub has secured funding to ensure its financial sustainability for the next two years, ensuring continued innovation and support for the healthcare community.

PIONEER has always been actively engaged in collaboration with numerous organisations, as well as patients and public. Over the past six months they have focussed on partnerships established via the HDR UK Alliance, namely the Regional Delivery Group working to deliver impactful research across the country. PIONEER has continued to participate in several large multi-centre grants and projects, addressing crucial areas such as multi-morbidity, polypharmacy, and the DARE TRE-FX grant, which explores technical solutions to the federation of analytics. They have continued to consult with patients and the public on a number of health research projects, including co-production of a patient leaflet for four acute hospitals, as well as a script of a patient-facing video explaining federated analysis. The next six months will see PIONEER initiating studies as part of the HDR UK second five years Medicines in Acute and Chronic Care driver enabling further collaboration across primary and secondary care.

Furthermore, the team are working on projects with the Patient Safety Research Collaborative that brings together NHS trusts, universities and private business to evaluate how digital tools can support clinical decision-making and reduce risks of harm for expectant mums and anyone in need of emergency treatment. They are also carrying out work on the Acute Care theme for the NIHR Birmingham Biomedical Research Centre, which combines world-class strengths in immunology and inflammation research and extensive experimental medicine infrastructure. The next six months promise an exciting continuation of collaborative efforts to advance acute care and improve patient outcomes.

NHS DigiTrials

NHS DigiTrials is a health data hub alumnus continuing to go from strength to strength and has recently released a new and improved version of its feasibility tool. This includes a range of accessibility and usability enhancements which have been tested with users. The NHS DigiTrials Feasibility Self Service tool enables researchers to determine if there are enough suitable people for their clinical trials in a matter of minutes. Researchers can run queries using filters such as disease diagnoses, hospital interventions, and primary care medicines, within a secure environment. The Hub continues to develop our other services to help trialists communicate with their participants and use routine healthcare data to follow-up their cohort.

If you would like to hear more or think the Hub can help you with your research you can email the team at *enquiries@nhsdigital.nhs.uk*. In the meantime have a look at their blog with more detail of what the tool can do: *Clinical trials get faster - NHS Digital.*

Partnership Programmes

INSIGHT

INSIGHT, the Health Data Research Hub for Eye Health, is the world's largest ophthalmic bioresource, with over 25 million retinal images and associated clinical data. An NHS initiative led by Moorfields Eye Hospital in partnership with University Hospitals Birmingham NHS Foundation Trust, INSIGHT makes routinely collected eye data available for approved research to advance development of new healthcare treatments and diagnostic technologies for eye disease and systemic diseases such as diabetes and dementia. The INSIGHT team and its partners are applying advanced analytics, including machine learning and artificial intelligence, to the Hub's datasets for patient benefit. New data from Moorfields Eye Hospital and University Hospitals Birmingham is added on an ongoing basis. For more information *visit the INSIGHT website* and follow INSIGHT on Twitter *@INSIGHTeyehub* for updates.

Gut Reaction

Gut Reaction is a unique, secure data resource designed to facilitate academic and industry research in Inflammatory Bowel Disease (IBD), working with the IBD community to improve treatment options and patient outcomes through safe, transparent and responsible use of patient data.

Gut Reaction is the world's largest repository of high quality, consented and linkable data supporting research to change the lives of people living with IBD. The goal during the initial funding period was to integrate health and research data from patients recruited to the NIHR IBD BioResource who have consented for information from their health records to be held in a research database. The specific aims are:

1. To transfer health records, including digital pathology and digital imaging data, from NHS Trusts, and HES data from NHS Digital, to AIMES, where clinical and phenotypic data are currently held.

2. To integrate these data with genomic data held at the University of Cambridge High Performance Computing (HPC) Service in a secure environment in Microsoft Azure, where they can be analysed anonymously, with the capability to re-identify to allow contact and recall of individuals.

The Hub has benefitted from industry interactions (e.g. GSK and AstraZeneca). AstraZeneca wish to conduct a fully funded data refresh across the whole BioResource, in exchange for data access. Gut Reaction has continued its work as part of the NIHR Bioresource. Working with partners, collaborators and the IBD community, they will continue to support research that makes a difference to those living with IBD. Gut Reaction are currently reviewing how data is housed, together with the need to supply this under licence to various requestors in different environments. There have been plenty of learnings from UK BioBank and the expectation is that there will be a large programme of work over the next one to two years.

DATA-CAN

DATA-CAN continues to work towards its strategic ambition of facilitating access to high quality cancer data from patients across the four devolved nations of the UK. The Hub continues to work with, and support, clinical, academic, charitable and commercial partners, offering expert advice, guidance and analytic expertise on routine datasets collected nationally or within individual cancer centres. In partnership with NHS England, DATA-CAN had helped develop an SDE containing both cancer-specific and general healthcare datasets on all cancer patients diagnosed in England. Current work to evaluate this data is underway through an evaluation of the impact of the COVID pandemic on cancer referrals, diagnosis, treatment and outcomes. This follows up on their previous work on the initial impact of the pandemic on cancer services and cancer patients, where they were the first to highlight the disastrous impact of the pandemic on cancer in the UK. This work was the catalyst for a *pan-European study* that showed that 100 million screening tests were not performed and up to one million cancer diagnoses may have been missed due to the negative impact of the pandemic.

DATA-CAN works with cancer centres across the UK to identify cohorts of patients with a specific diagnosis to describe treatment pathways and outcomes. This work, often performed for commercial partners, supports a deeper understanding of routine care or provides real-world evidence to support funding approvals for new cancer therapies within the UK. DATA-CAN continues to expand this network actively seeking new partners to enhance the value of this offer. DATA-CAN aims to deliver this work through the adoption of the OMOP common data model and is evaluating privacy preserving technologies including federated analysis to further enhance this work. DATA-CAN continues to work with a range of commercial partners to support the evaluation and implementation of novel technologies aiming to enhance the collection, curation and use of routine health care data in this work. This includes work linking structured data to unstructured letters and reports, digital imaging and pathology, detailed molecular and genetic phenotype and most recently patient reported outcome data.

Our PPIE representatives remain critical to DATA-CAN's mission and continue to guide and influence our strategy ensuring that the work performed delivers value to patients and the NHS more widely. DATA-CAN received the HDR UK Impact of the year award for its work on evaluating continuous versus intermittent cetuximab treatment for colorectal cancer. The evidence that this work generated was critical to *NHS England approving a policy change* to allow treatment breaks for patients with colorectal cancer. DATA-CAN will work closely over the next five years with HDR UK's Big Data in Complex Diseases Driver Programme to maximise the use of the national-scale datasets highlighted above, aiming to enhance early diagnosis, better treatment and improved outcomes.

BREATHE

BREATHE was a unique collaboration driving the use of health data in research to transform respiratory health.

BREATHE's legacy lives on through the establishment of three respiratory data registries in England, Scotland and Wales. The groundwork to create a fourth in Northern Ireland is underway, tying in with the activities of the Inflammation and Immunity Driver Programme. In each of these nations, BREATHE collaborated with Trusted Research Environments (TREs) and data providers to create cohorts of patients with chronic respiratory diseases, specifically: asthma, Chronic Obstructive Pulmonary Disease (COPD) and Interstitial Lung Disease (ILD). BREATHE worked with the Clinical Practice Research Datalink (CPRD) in England, DataLoch in Scotland, and the Secure Anonymised Information Linkage (SAIL) Databank in Wales. The Northern Ireland registry will be built in collaboration with the Honest Broker Service and Queen's University Belfast.

All registries provide a baseline harmonised set of criteria and clinical coding for how asthma, COPD, and ILD should be characterised in routine health records, contain research-ready data related to patients' demographics, diagnoses, condition events (e.g. ongoing GP or hospital care) and medications, and also provide the facility to link to wider records pertaining to comorbidities, other conditions, and healthcare history.

The DataLoch Respiratory Registry, currently covering South-East Scotland residents within the NHS Lothian catchment, holds additional information on Cystic Fibrosis and Wheeze (a common respiratory symptom). In England and Wales, the processes of linking further conditions to the registries is relatively straightforward, subject to governance and dependent on the inclusion criteria of specific research studies.The registries have been expertly curated with clinical input and are harmonised as far as possible (across multiple clinical coding sets and agreed individual-level characteristics), making it easier to conduct pan-UK analyses. A paper detailing the methodology used in the cohort construction and harmonisation processes has been submitted for publication. Once published, the curated script and codes used to create the asthma, COPD and ILD cohorts will be made available, in collaboration with the team who created the cohorts. This will allow for equivalent cohorts to be curated elsewhere, enabling curation and linkage within and to wider records held by the data custodians that control them.

Another, separate paper is also in progress. Using data from England, Scotland and Wales, this epidemiology study aims to generate baseline prevalence and incidence of asthma, COPD and ILD across the UK, within the context of the curated cohorts. It will also include asthma data from Northern Ireland.

Partnership Programmes

BHF Data Science Centre

The *BHF Data Science Centre (BHF DSC)* is a *partnership* between HDR UK and the British Heart Foundation and sits within HDR UK.

We work with a wide range of partners including patients, public, clinicians, researchers and NHS organisations to help them carry out research using health data into the causes, prevention and treatment of all diseases of the heart and circulation (such heart attacks, stroke and vascular dementia). We do this to ensure new advances in treatment and care for diseases of the heart and circulation get to the patient as quickly as possible. The centre started in January 2020 and has initial investment from the BHF for £10 million over the first five years. The BHF DSC has the following *thematic areas*:

- Structured data - this work is driven by the *CVD-COVID-UK/COVID IMPACT Consortium* which has enabled an England-wide electronic health record (EHR) resource within the NHS England SDE.

- Unstructured Data.

- Personal Monitoring Data.

- Computable Phenotypes.

- Data enabled clinical trials.

- Enhancing cohorts.

- Diabetes Data Science Catalyst.

Community-wide opportunities include joining the *CVD-COVID-UK/COVID IMPACT Consortium* to enable rapid access to NHSE SDE data and national TREs in Scotland and Wales, collaborative opportunities with BHF DSC and SAIL Databank to *enhance cohort study data* by linking to EHR.

Partnership Programmes

DARE UK

The *DARE UK* programme is funded by UKRI as part of its *Digital Research Infrastructure* portfolio of investments, which supports the development of a coordinated vision for digital research infrastructure in the UK. DARE UK is a pan-UKRI, cross-domain programme - its scope covers all types of sensitive data, including data about education, health, the environment and much more. There is growing consensus (including from *Phase 1 DARE UK recommendations*, the *Goldacre Review*, the *UK Health Data Research Alliance TRE Green paper*, and the *DHSC Data Saves Lives* policy paper) that all sensitive data should only ever be accessed and analysed by researchers within a TRE. Central to the DARE UK programme's ambition is to enable the development of a national interoperable network of secure digital research infrastructures or TREs, laying the foundation for a next-generation ecosystem of TREs for advanced data research for the public good, using a cloud-first design philosophy.

Delivery of Phase 1 of the programme began in July 2021 and is being delivered with joint oversight from HDR UK and ADR UK. Phase 1 is an extensive programme of community engagement that is an essential foundation for developing a clear vision of the needs of different research communities, and to address the interests and concerns of the public around the use of sensitive data for research. It will lead to a community co-designed blueprint for an interoperable network of next generation TREs alongside a model for delivery of Phase 2 which will begin to iteratively build, test, and establish this vision. As Phase 2 approaches, with several key outputs from Phase 1 in development such as the initial version of the *Federated Architecture Blueprint* and the outputs of the portfolio of *Driver Projects*, the DARE UK programme is considering how best to deliver the programme's ambitions in collaboration with the various sensitive data research communities and stakeholders, while incorporating the full range of earlier outputs of Phase 1 (for example the *Phase 1 DARE UK recommendations*, the portfolio of *Sprint Exemplar Projects*, and *DARE UK Public Dialogue* amongst others).

Important for this community to consider are the kinds of inter-disciplinary, cross-domain scientific use cases that could be realised through a national interoperable network of secure TREs and to collaboratively begin to consider the opportunities for co-delivery of components of the DARE UK vision - for example through involvement with *DARE UK Community Groups* - in the lead up to Phase 2 of the programme.

HDRuk
Health Data Research UK