

THE RISE OF NEW ENGLISH SECURE DATA ENVIRONMENTS

Rachel Denholm

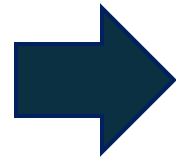
With thanks to Jonathan Sterne, Venexia Walker, Charlotte James, Andy Boyd, OpenSafely, and BHF Data Science Centre

Traditional access to routinely collected data

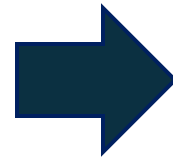
Data sharing process



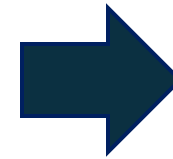
Researcher proposes a project



Project-specific IG/ethics/contracts



Once approved, copy of the data sent to researcher



Data analysed on their organisation's computer: should be deleted once analyses completed

- **Insecure:** detailed data are inherently disclosive, no control over analyses done.
- **Inefficient:** no incentive or framework for sharing code or codelists
- **Opaque:** analyses not open or reproducible

The NHS will no longer permit sharing/dispersal of unconsented data



The image shows the cover of a policy paper from the Department of Health & Social Care. The cover features a blue header with the title 'Secure data environment for NHS health and social care data - policy guidelines' and a sub-header 'Policy paper'. Below the title, it says 'Updated 23 December 2022'. The main body of the cover is white and contains the first point of the policy: '1. Secure data environments will be the default way to access NHS Health and Social Care Data for research and analysis'. Below this point, there is a paragraph explaining that secure data environments must be adopted by organizations hosting NHS health and social care data for research and analysis, as they improve data privacy and security, helping to build public trust.

A review of
of State for

**Better
Safe
Data
and**

69
April 2022

 Department
of Health &
Social Care

Policy paper

**Secure data environment for NHS health
and social care data - policy guidelines**

Updated 23 December 2022

**1. Secure data environments will be the default way to
access NHS Health and Social Care Data for research and
analysis**

Secure data environments must be adopted by organisations hosting NHS health and social care data for research and analysis. These environments have features that improve data privacy and security, which will help build public trust in the use of their data.

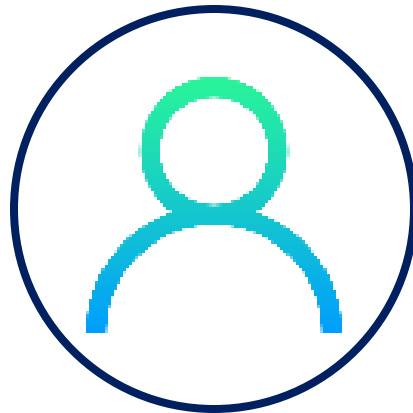
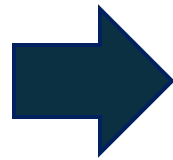
Health and

What is a secure data environment (SDE)? (= trusted research environment (TRE))

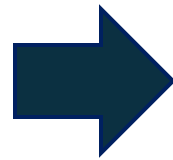
Data access process



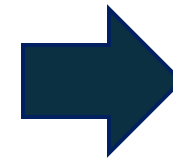
Accredited researcher
proposes project



Request approval from
one advisory group



Access via a remote
portal. Cannot copy or
download data. Outputs
exported only after
checking by an
independent researcher



End of project, access removed

History of Secure Data Environments in the UK



NHS England national programme

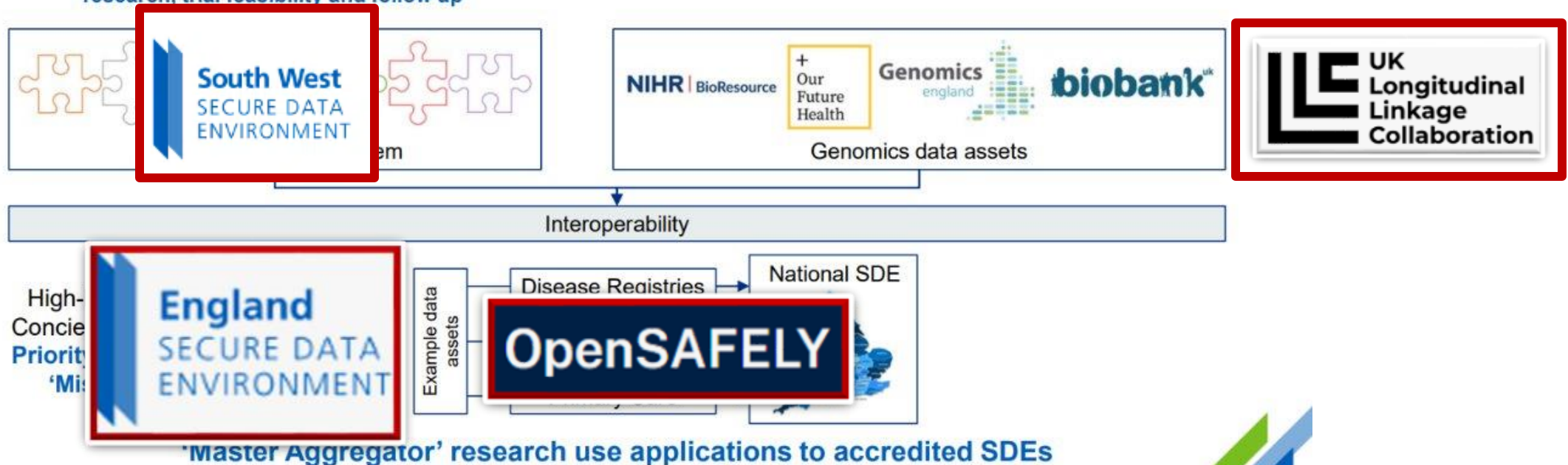
Data for R&D
PROGRAMME



An interoperable network of SDEs

Granular | Agile | Multi-modal | Real-time |
Connected to services | Data controller buy-in
AI development, real world studies, translational
research, trial feasibility and follow up

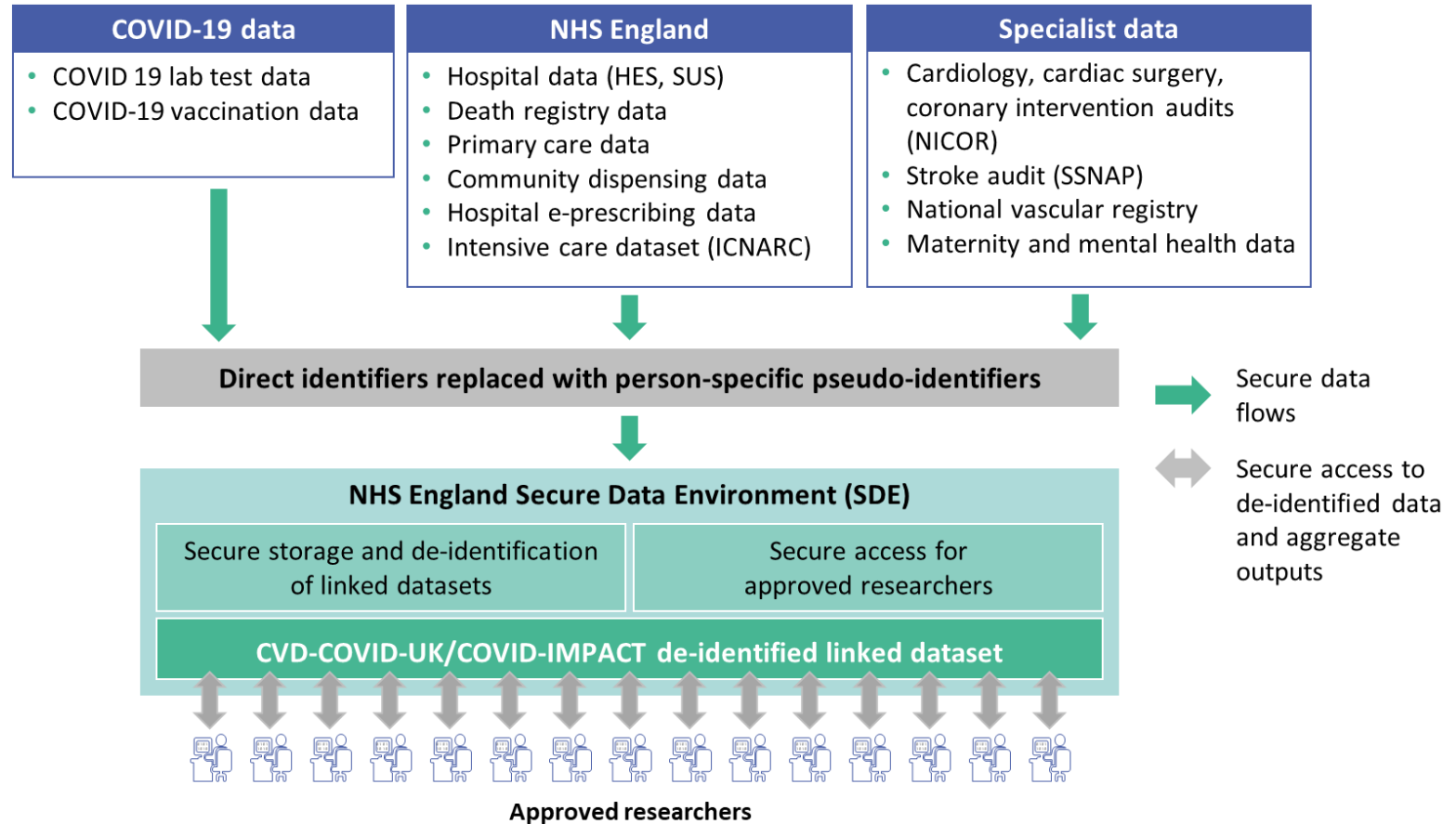
High-value | Enhanced with linkage |
Advanced compute
Clinical and research use-cases



'Master Aggregator' research use applications to accredited SDEs










NHS England's Secure Data Environment

- Linkable person-level records accessed within NHS England's Secure Data Environment
- Enables whole population insights on the English population of 57 million people
- Comprehensive information on characteristics and health outcomes
- Includes all age groups, ethnicities, geographic locations, socioeconomic, health and personal characteristics
- All datasets updated monthly



BMJ 2021; 373 doi: <https://doi.org/10.1136/bmj.n826>

CVD-COVID-UK/COVID-IMPACT Consortium in numbers (26/08/24)

- | | | |
|--|--|---|
|  >400 members |  >50 institutions |  >70 analysts |
|  3 national TREs |  73 datasets |  68 projects |
|  >70 active studies |  19 preprints/
submissions |  21 publications |

<https://bhfdatasciencecentre.org/areas/cvd-covid-uk-covid-impact/>



Secure analytics platform for NHS electronic health records

OpenSAFELY delivers research across over 58 million people's health records, always respecting patient confidentiality

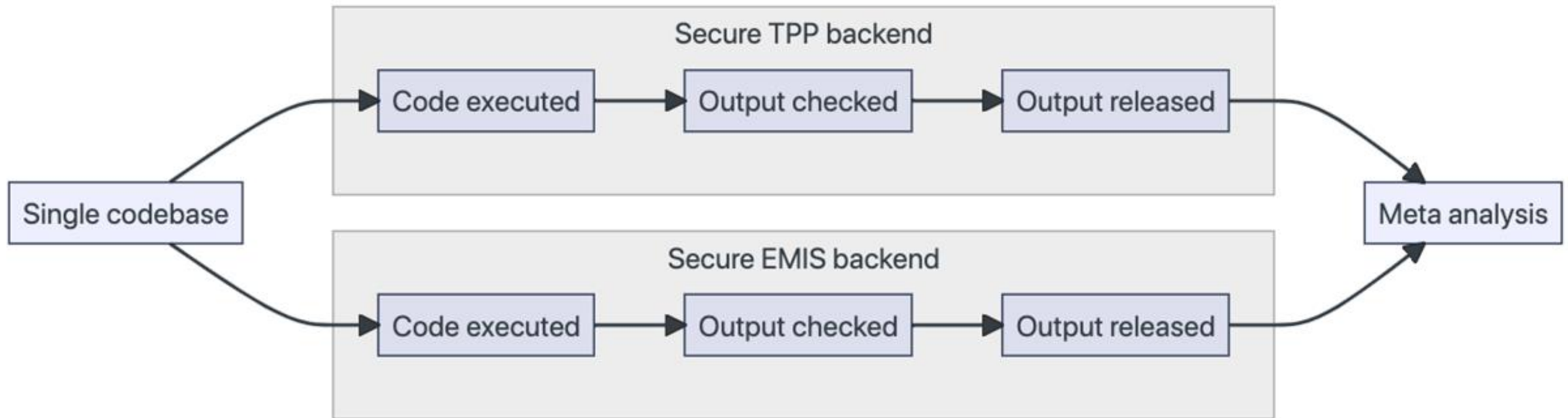
[What is OpenSAFELY?](#)

[How do I know my data is safe?](#)

WORKING IN PARTNERSHIP WITH



- OpenSAFELY is an open-source software platform that can be built on top of any database.
- It is installed inside the data-centres where GP data already sits
- All code (platform and users') is shared on GitHub
- Over 150 projects from 22 different organisations
- 86 publications and 14 preprints to date



Data: Electronic Health Records



OpenSAFELY



Primary care	- Embedded with primary care software system	- COVID-19 General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR)
Secondary care	- Hospital admissions - Emergency attendances	- Hospital Episode Statistics
Deaths	- Registered deaths	- Civil Registrations of Deaths
Medicines	- High Cost Drugs	- Medicines Dispensed in Primary Care (NHSBSA)

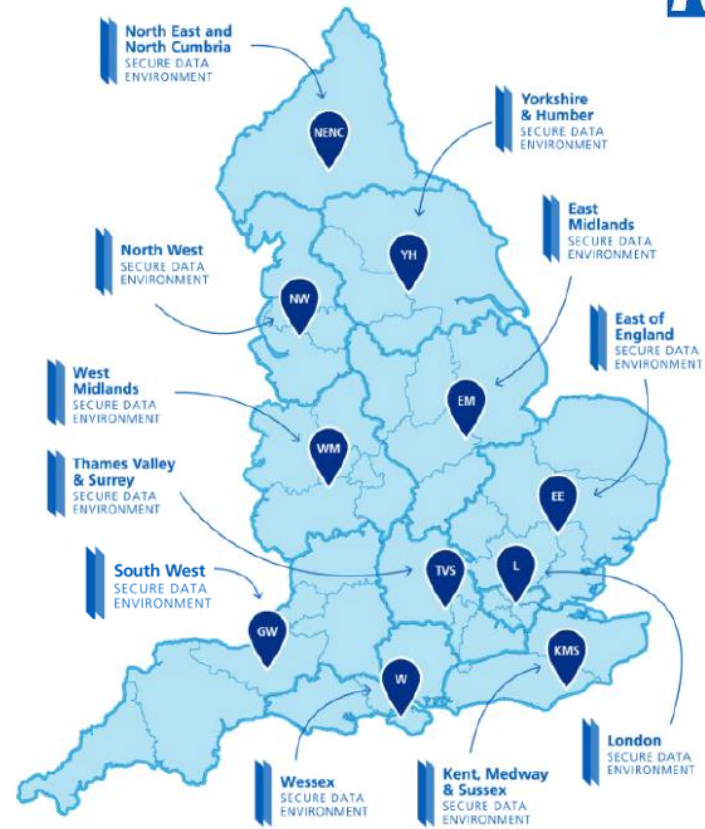
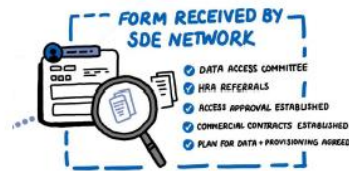
Data: primary care



Active Records	TPP: 24M With EMIS: 55M	≈55M	Gold (Vision): 3m AURUM (EMIS): 13.8M
Freq	Real time	Fortnightly	Monthly
Population coverage	Regional variation	Whole population Alive 1 st Jan 2020	Regional variation
Code list coverage	All	Extract	All

NHS England Sub National SDE network

Research
Secure Data
Environment
NETWORK

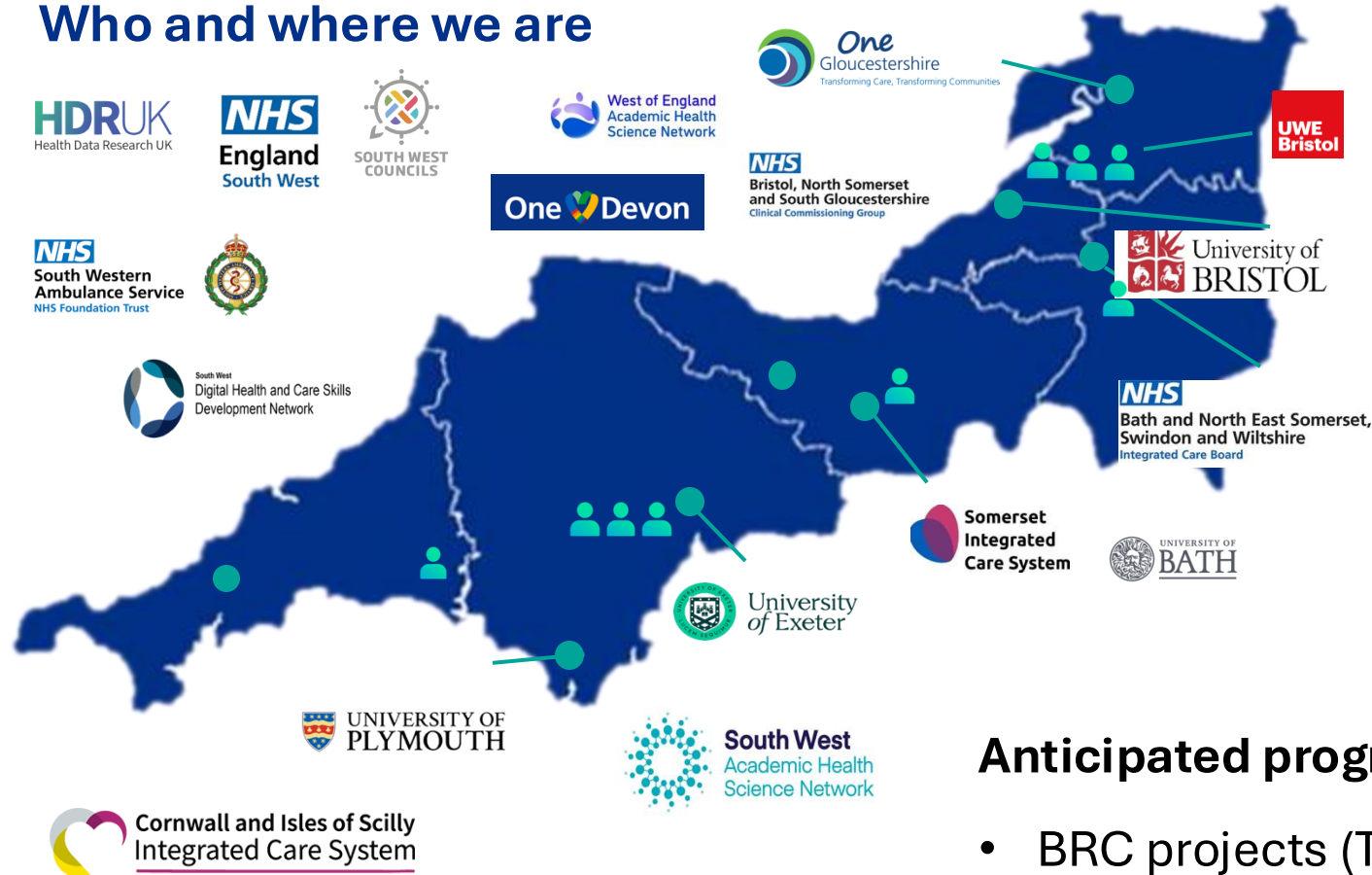


Research
Secure Data
Environment
NETWORK
Research Powered
by Data

HDRUK
Health Data Research UK
Health Data Research
Innovation Gateway

England
SECURE DATA
ENVIRONMENT
National coverage

Who and where we are



Progress

- Ethical approvals
- Innovative data access model agreed
- SDE procured and being built

Anticipated progress

- BRC projects (TDS SWSDE seed corn funding)
- Wider partnership with NHS data controllership model

Linking Data for Future Research Discoveries

- Deidentified data from >20 Longitudinal Population Studies **≈ 2 million participants**
- Linked to administrative data
 - NHS England (soon NHS Wales and NHS Scotland)
 - Department for Work and Pension and HMRC
 - Geospatial data
- 28 projects approved
- >50 potential users
- 16 Universities requested letters of support

Machine learning and AI methods in SDEs

Early days....

The image shows a screenshot of the DARE UK website. The header features the DARE UK logo on the left, followed by navigation links: "About us", "How we work", "Involving the public", "News and events", and "FAQs". A "GET IN TOUCH" button is positioned to the right of these links, along with search and accessibility icons. The main content area has a dark background with the text: "We're on a mission to improve lives by enabling better, faster, and more trustworthy sensitive data research in the UK." To the right of this text are several colorful geometric shapes and two circular images showing data visualizations. At the bottom, there is a small green circular icon on the left and a footer containing the text: "7 NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK".

DARE UK About us How we work Involving the public News and events FAQs GET IN TOUCH

We're on a mission to improve lives by enabling better, faster, and more trustworthy sensitive data research in the UK.

7 NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK

Machine learning and AI methods in SDEs

Challenges	Solutions
Compute	Access to local/national HPC
SDE researcher skillset	Capacity building
Information governance	Methodological development – synthetic data
Disparate data	Federated learning
Disclosure control	SAFE models
Patient/public perceptions	Outreach and engagement

Work led by Charlotte James

THANK YOU FOR LISTENING!

ANY QUESTION

r.denholm@bristol.ac.uk

Electronic Health Record Group