



Digital

Access to data from NHS Digital for biomedical research

**Bioinformatics for Biobusiness: Accessing
biomedical data for data analysis**

Presented by:

Michael Chapman - Director of Research & Clinical Trials

June 2022



Acknowledgements

Find Recruit Follow / Data for R&D

- Centre for Improving Data Collaboration, NHS England Transformation Directorate
- Science Research and Evidence Directorate, Department of Health and Social Care
- National Institute for Health Research
- UK Research & Innovation

NHS DigiTrials Health Data Research Hub

(Oct 19-Jul 21)

- Health Data Research UK
- University of Oxford
- Microsoft
- IBM

Covid-19 TRE development

- Health Data Research UK
- British Heart Foundation Data Science Centre

NHS DigiTrials Co-development Panel

4,000 staff across NHS Digital



This work uses data provided by patients and collected by the NHS as part of their care and support.

#datasaveslives

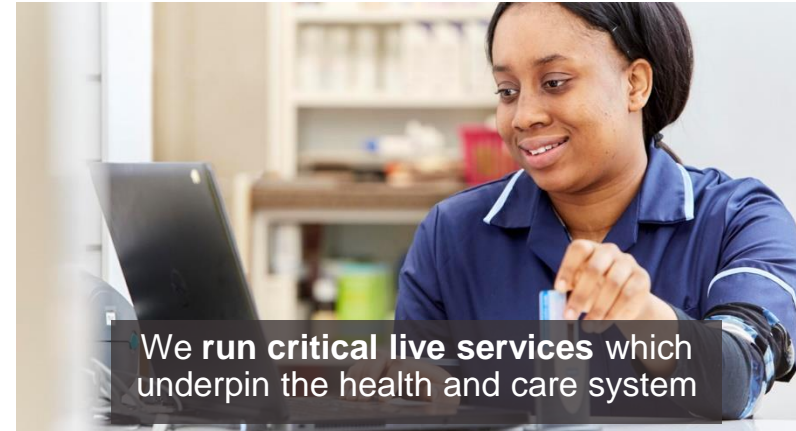


NHS Digital is the national digital, data and technology delivery partner for the NHS and social care system



We support the NHS and the people that it serves

We believe in improving people's lives through data and technology. To help create better health and care services and solutions, we deliver products, platforms and services, and share national data and insight.



Researchers and planners use data from NHS Digital for discovery and insight

We publish 300 official health and care data sets every year, supporting life sciences research, helping local and national planning and enabling accountability



Over 500 billion health and care data points are stored in our records, from Accident & Emergency Admissions to Zero Bed Days



Accessing data from NHS Digital

DATA ACCESS REQUEST SERVICE



Data disseminated to approved users to improve health and social care



Secure Data Environment

Secure access to linked, de-identified health data



NHS DigiTrials

Data enabled clinical research

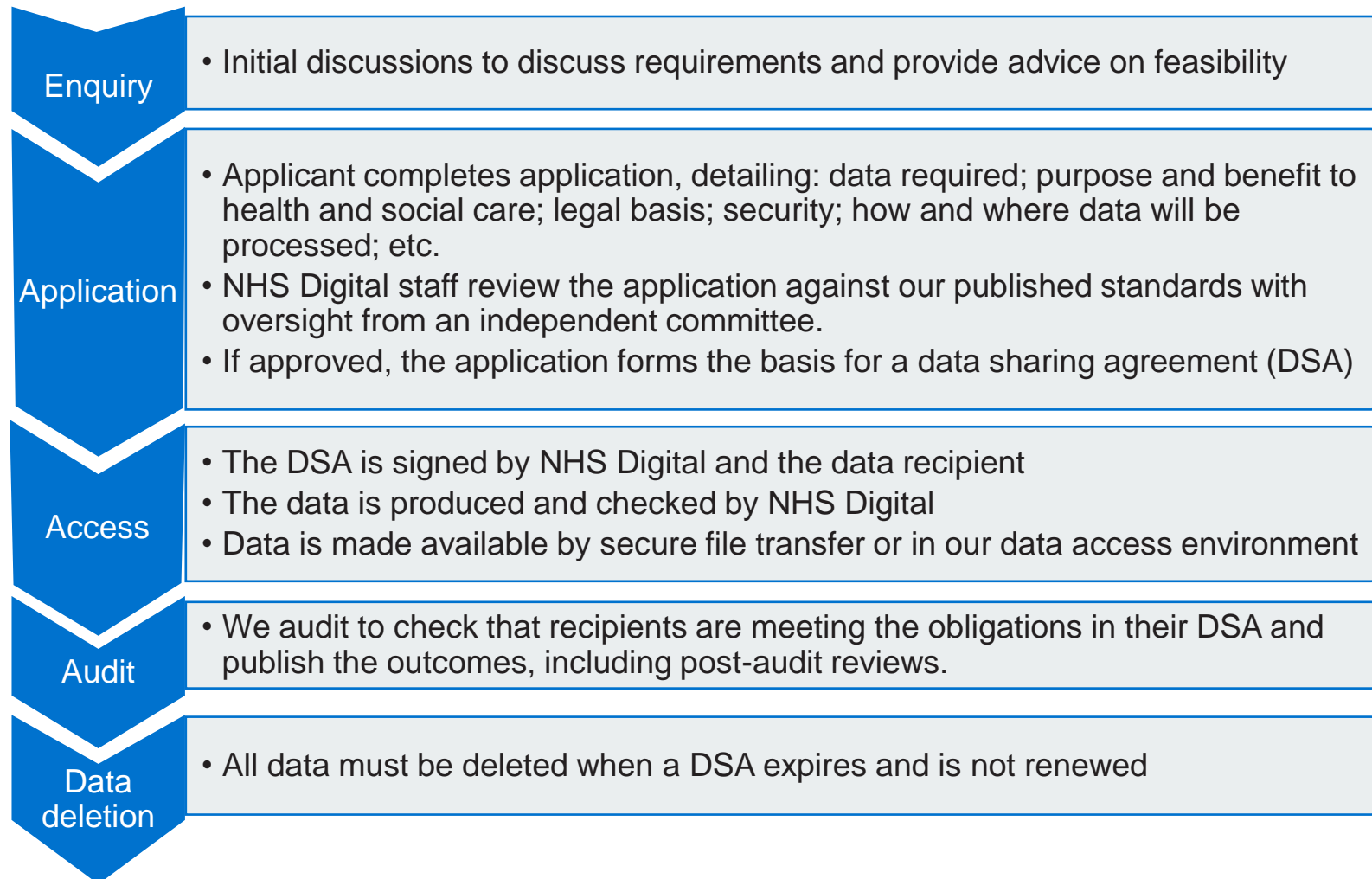
OUTCOMES

COMMUNICATION

RECRUITMENT

FEASIBILITY

Requests to access data must meet published standards. Details of data sharing agreements and audits are public



NHS Digital's Data Uses Register (Applicant Organisations)

Select an Applicant Organisation from the list on the right, and then click the button in the bottom left corner to view the Data Sharing Agreement(s) (DSA) associated with it.

Applicant Organisation

Search

Click on an APPLICANT ORGANISATION and then click [HERE](#) to view the DSA(s)

Applicant Organisation

- 2020 DELIVERY LTD
- 3M UNITED KINGDOM PLC
- ADDENBROOKE'S HOSPITAL
- ASTRAZENECA AB
- ASTRAZENECA UK LIMITED
- BANGOR UNIVERSITY
- BARNSELY METROPOLITAN BOROUGH COUNCIL
- BARTS AND THE LONDON SCHOOL OF MEDICINE AND DENTISTRY
- BARTS HEALTH NHS TRUST
- BATH AND NORTH EAST SOMERSET COUNCIL
- BIRMINGHAM CITY COUNCIL
- BIRMINGHAM WOMEN'S AND CHILDREN'S NHS FOUNDATION TRUST
- BLACKBURN WITH DARWEN BOROUGH COUNCIL
- BLACKPOOL BOROUGH COUNCIL
- BOLTON METROPOLITAN BOROUGH COUNCIL
- BRACKNELL FOREST COUNCIL
- BRIGHTON AND HOVE CITY COUNCIL
- BRISTOL CITY COUNCIL
- BRITISH THORACIC SOCIETY
- BUCKINGHAMSHIRE COUNCIL
- BUCKINGHAMSHIRE HEALTHCARE NHS TRUST

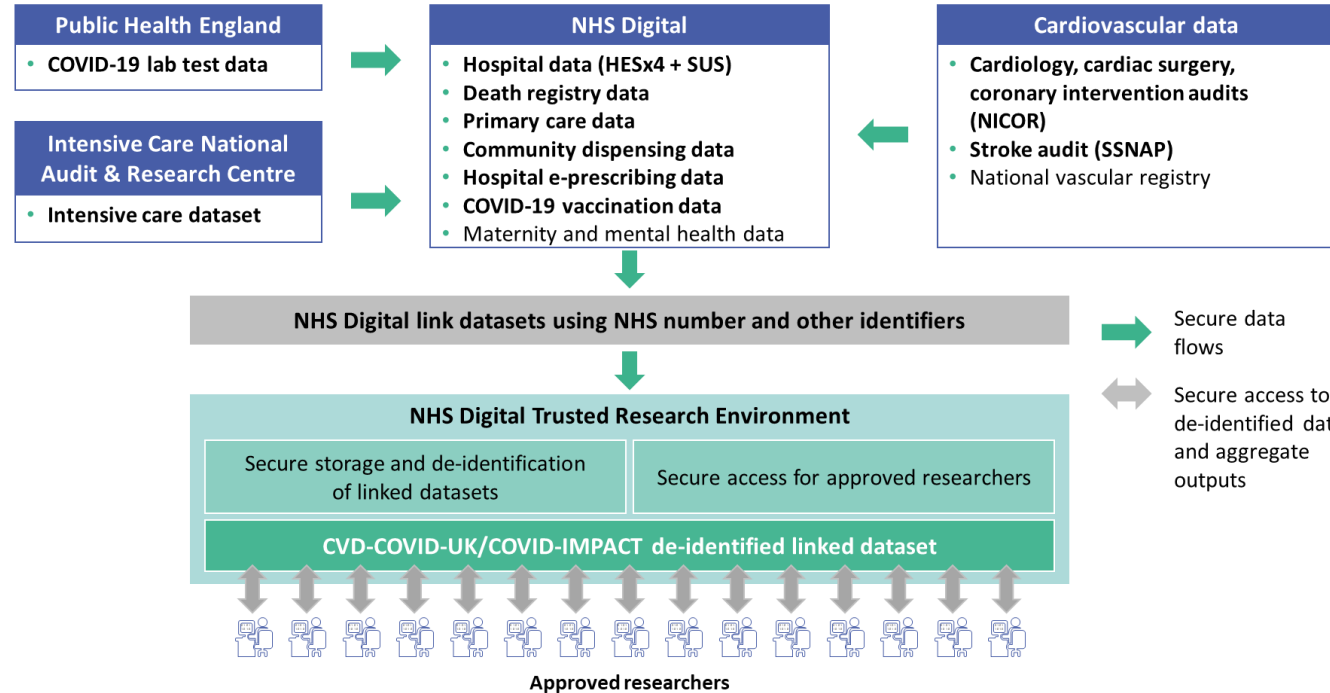
Purpose Statements

Objectives for Processing	Processing Activities	Expected Output	Expected Measurable Benefits	Yielded Benefits
<p>DARS-NIC-604847-S4B5L-v0.4</p> <p>DSAs, Site UK Ltd and King's College London (KCL), as part data controllers, are requesting access to pseudonymised research level data in order to carry out follow up analysis based on a cohort of patients who are being recruited to a clinical trial called 'NHS Galien', following being invited to participate if they are found to meet the requisite eligibility criteria.</p> <p>Notes that all members of the team carrying the trial are based at the Cancer Prevention Trials Unit (CPTU) at King's College London. For the sake of consistency, and as King's College London is listed as a data controller, we are referring to King's College London throughout the registration, this encompasses the team at CPTU.</p> <p>Background, Purpose and Rationale behind NHS Galien: A new Multi-Cancer Early Detection (MCED) test has been developed that can detect many types of cancer from a single blood sample. This test is called Galien™ and this trial aims to find out whether it is better at detecting cancer early, compared to other tests that are currently used. The purpose of NHS Galien is to determine the clinical utility of MCED based test for individuals in a general screening population in a real world setting. This includes relevant before this trial is that MCED is a novel screening paradigm, and assessment of the use and impact of test results is necessary to enable appropriate clinical practice. This includes the first randomised, double blind, controlled trial specifically powered to assess clinical utility of a MCED test.</p> <p>Registration to the trial is being added by NHS Digital (Site UK) using a secure application, DARS-NIC-604847-S4B5L-v0.4, and is being placed in the Data Access Request Service (DARS).</p>	<p>Patient Identifiable Data (PID) referred to in section 3a (above) will be stored in the Clinical Records Management System (CRMS) hosted on Amazon Web Service (AWS) UK. NHS Digital record level pseudonymised data referred to in section 3a (above) will be stored in a secure S3 bucket (S3 is a type of the storage system in a cloud environment based on Amazon Web Service (AWS) UK).</p> <p>METHODOLOGY</p> <p>1. The PID referred to in the cohort will be securely transferred from the CRMS as an encrypted dataset (S3) to AMES, Management Services Limited's FTP server. From there, King's College London associates the cohort into AMES' Trusted Research Environment (TRE) and transfers to NHS Digital in the form of a partial over Health and Social Care Network (HSCN). Data is stored on the AMES Trusted Research Environment (for a maximum of 12 weeks for confirmation of successful receipt by NHS Digital).</p> <p>2. King's College London will provide a cohort of approximately 140,000 records (in total), including a pseudo Study ID, plus the identifier, NHS Number, Date of Birth, Sex and Post Code along with four other variables (date of diagnosis, date of application to NHS Digital via a Secure Electronic File Transfer Service (SEFT) or other secure, NHS Digital approved file transfer mechanism).</p> <p>3. An encryption key is sent from the King's College London team to the CRMS system for encrypting the dataset, allowing for the participants identified in the dataset to be flagged for following by NHS Digital. Only the King's College London team retains the ability to decrypt the dataset.</p>	<p>All outputs will be aggregated with overall number suppression as per the HES analysis guide or according to the specific data sets suppression policies.</p> <p>The results of the NHS Galien trial outputs will be published in peer reviewed publications.</p> <p>The results of the NHS Galien trial outputs will be published in peer reviewed publications and used by NHS Digital in their decisions around implementation of the test within the NHS. The results analysis, which will include analysis of a key secondary objective, is expected by the end of 2023 or beginning of 2024. The primary objective result is expected by the end of 2023 or beginning of 2024. Planned dissemination activities will aim to enable the engagement with the scientific and policy making communities to ensure that knowledge developed by the research can benefit these communities to help the NHS detect cancer early, where it is easier to treat.</p>	<p>The use of the dataset referenced herein will be used for the primary and secondary objective analyses of the NHS Galien trial. If the trial is successful, it is anticipated that there will be future health benefits to individuals within the NHS.</p> <p>Multi-cancer early detection (MCED) testing is a novel screening paradigm, and assessment of the use and impact of the test requires robust clinical standards of care to be necessary to establish clinical utility and ensure integration into clinical practice. This will be the key benchmark/considered trial to be UK clinically proven to assess the test performance and clinical utility, including harms and benefits, of a multi-cancer early detection test when used in population screening. Due to the scale of this study (involving approximately 140,000 participants in total to enroll) and the speed with which participants will be enrolled and the results analysed, local data collection for primary and secondary objectives analyses is not practical and may compromise already achieved NHS priorities. As such, the study analyses depend entirely on central NHS datasets.</p> <p>In its Long Term Plan (LTP) published in January 2019, NHS Digital set an ambition to achieve a significant shift in the proportion of cancers diagnosed at an early stage (stage 1 or 2). Patients diagnosed early, or stages 1 and 2, have the best chance of curative treatment and long term survival. This study aims to determine whether the MCED test can provide a meaningful contribution towards achieving that ambition, which would have substantial impacts on cancer treatment and the survival of those patients within the NHS. If the study is successful, it is hoped that NHS will adopt the test.</p>	

COVID-19 Secure Data Environment

Minimum viable product supporting academic, NHS, charity and industry users

COVID-IMPACT



DATA-CAN

Timely and secure access to de-identified cancer data, enabling important COVID-19 related research on rates of cancer referrals, diagnosis and treatment



Effectiveness of the Oxford-AZ COVID-19 vaccine, and if there is a causal association of thrombosis with thrombocytopenia syndrome



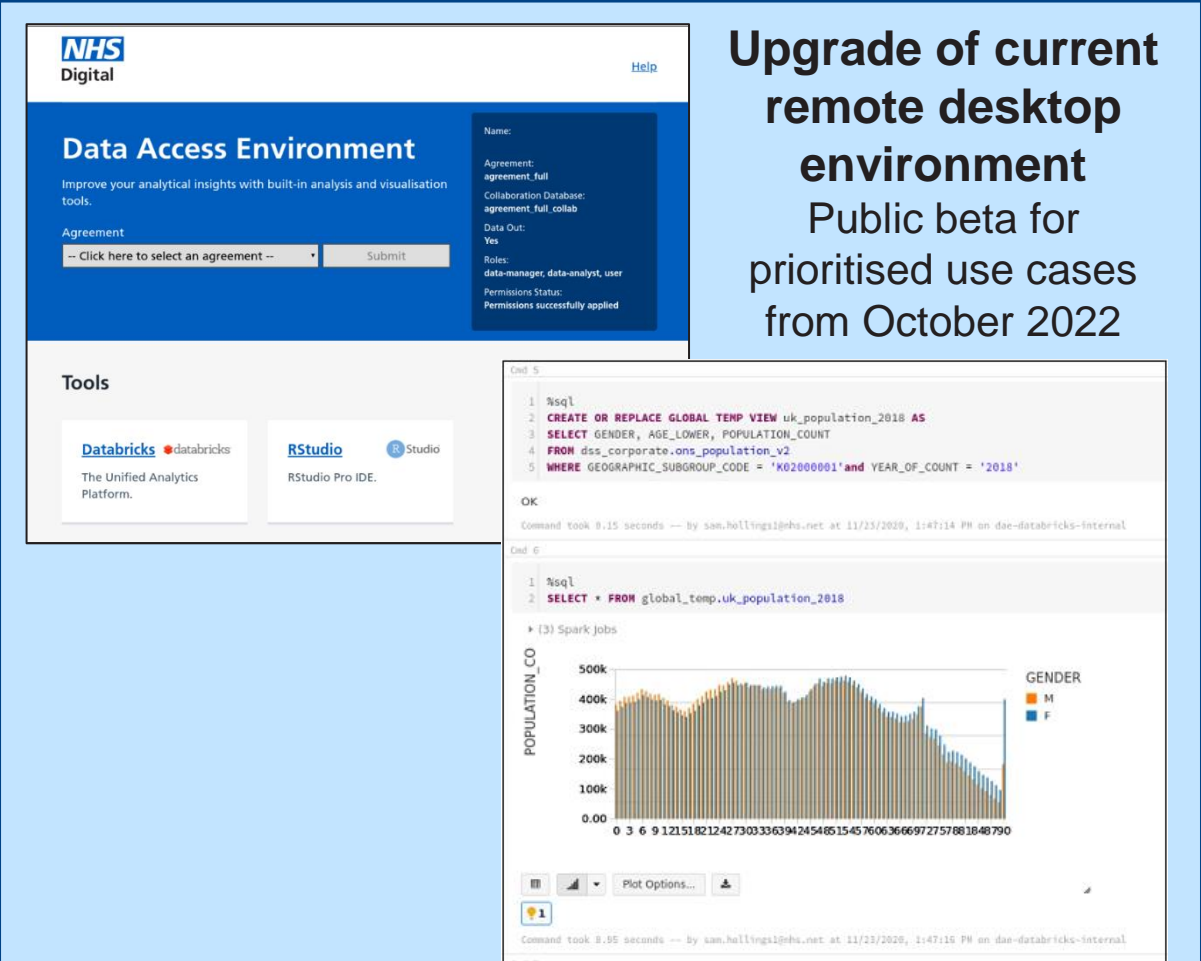
Analysis to inform policy and strategy around COVID-19 pandemic response and recovery



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



The secure data environment will become the default route for unconsented access to data from NHS Digital



NHS Digital

Data Access Environment

Improve your analytical insights with built-in analysis and visualisation tools.

Agreement: **agreement_full**

Collaboration Database: **agreement_full_collab**

Data Out: **Yes**

Roles: **data-manager, data-analyst, user**

Permissions Status: **Permissions successfully applied**

Tools

- Databricks** - The Unified Analytics Platform.
- RStudio** - RStudio Pro IDE.

Upgrade of current remote desktop environment

Public beta for prioritised use cases from October 2022

```
1 %sql
2 CREATE OR REPLACE GLOBAL TEMP VIEW uk_population_2018 AS
3 SELECT GENDER, AGE_LOWER, POPULATION_COUNT
4 FROM dss_corporate.ons_population_v2
5 WHERE GEOGRAPHIC_SUBGROUP_CODE = 'K02000001' and YEAR_OF_COUNT = '2018'
```

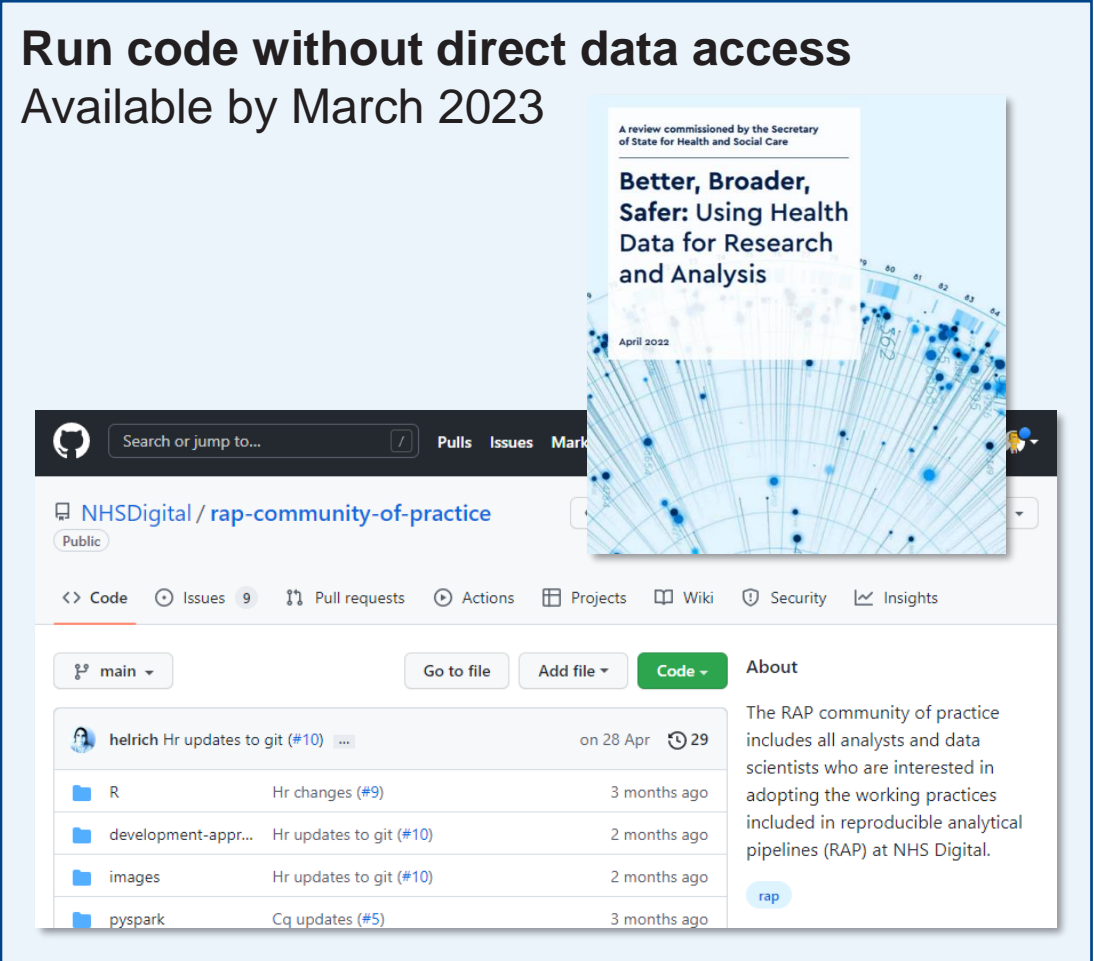
Command took 0.15 seconds -- By sam.hallings@nhs.net at 11/23/2020, 1:47:14 PM on dae-databricks-internal

POPULATION_CO

GENDER

- M
- F

Command took 8.95 seconds -- By sam.hallings@nhs.net at 11/23/2020, 1:47:16 PM on dae-databricks-internal



Run code without direct data access

Available by March 2023

Better, Broader, Safer: Using Health Data for Research and Analysis

A review commissioned by the Secretary of State for Health and Social Care

April 2022

NHSDigital / rap-community-of-practice

Public

Code Issues 9 Pull requests Actions Projects Wiki Security Insights

main

Go to file Add file Code About

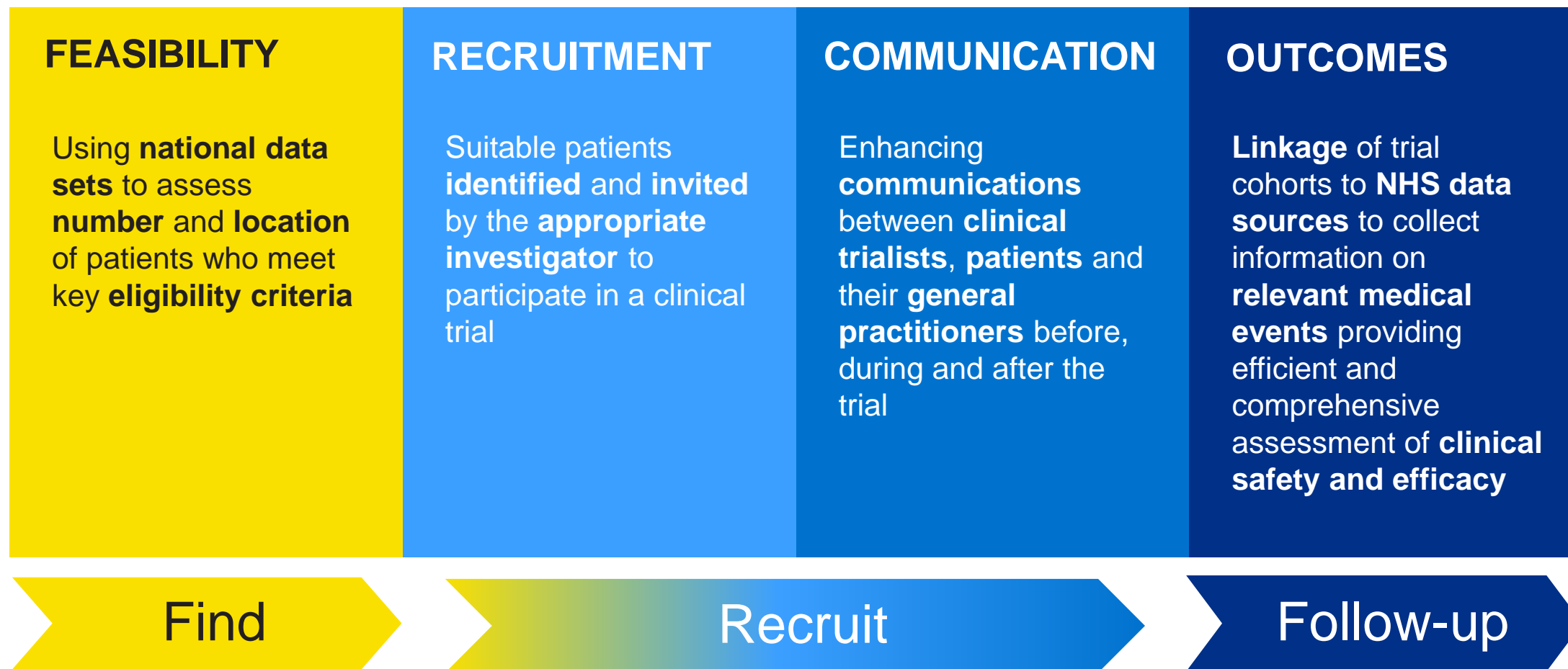
File	Commit	Time
R	Hr changes (#9)	3 months ago
development-appr...	Hr updates to git (#10)	2 months ago
images	Hr updates to git (#10)	2 months ago
pyspark	Cq updates (#5)	3 months ago

The RAP community of practice includes all analysts and data scientists who are interested in adopting the working practices included in reproducible analytical pipelines (RAP) at NHS Digital.





NHS DigiTrials – proposition





NHS DigiTrials supporting critical health research

- **RECOVERY trial:** Identifying effective treatments that have saved an estimated 1m lives globally
- **PRINCIPLE and PANORAMIC trials:** more than doubled recruitment rates and supported safe prescribing for virtual trials.
- **Vaccine trials:** Novavax completed recruitment in 2 months with 55% volunteers recruited via DigiTrials
- **GenOMICC:** discovery of genes which increase susceptibility to life-threatening COVID-19
- **NHS Galleri:** on track to recruit 140,000 participants with highly representative demographics who will be followed up through routine datasets





“As someone of black African origin I feel it is so important for clinical research to be representative of the population so that treatments can help people from all backgrounds. Services like DigiTrials give anyone who is clinically eligible the opportunity to participate, regardless of their background.”

~ Dolapo Ogunleye

NHS DigiTrials Co-development Panel



Where next

The future is here, it's just not evenly distributed

Data for R&D Vision

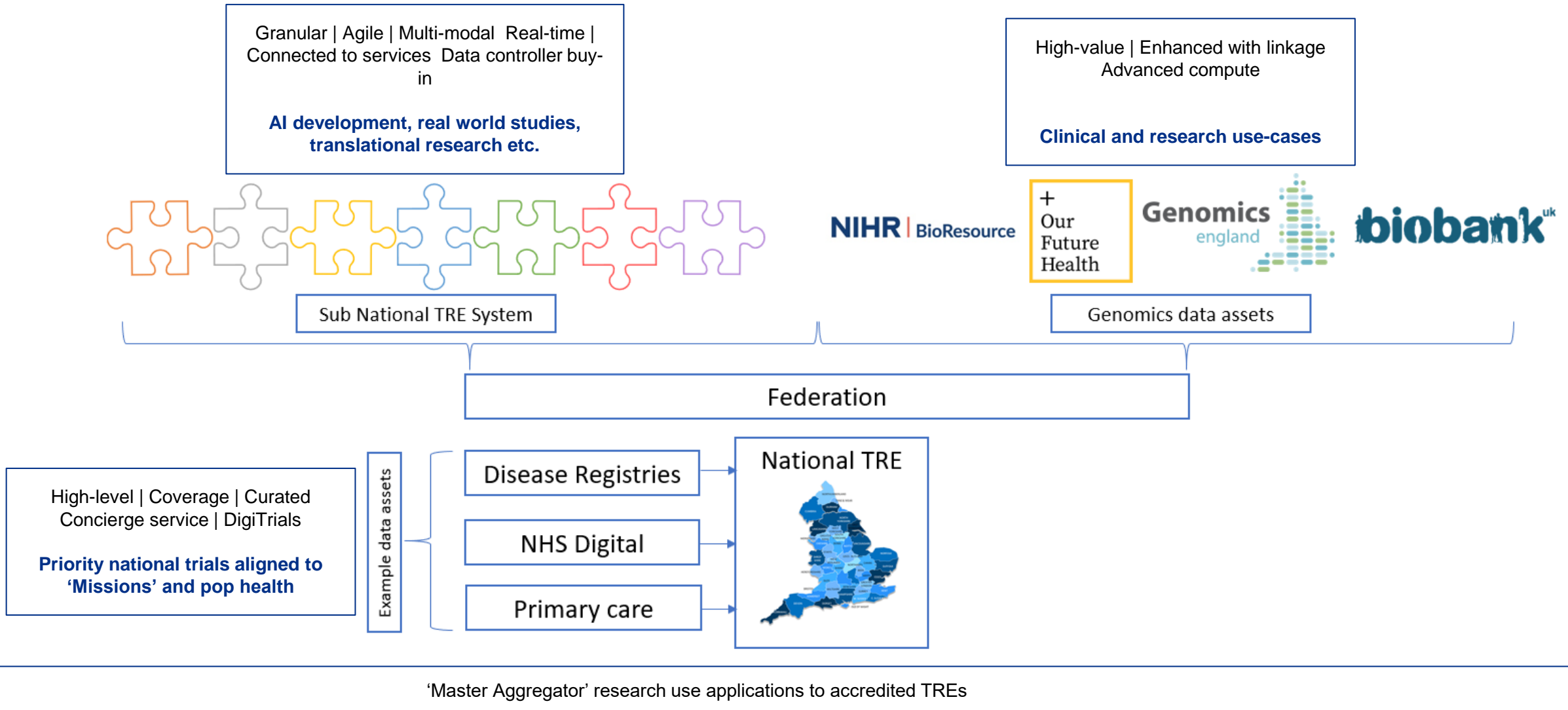
By 2025, have a world leading
NHS-wide health data research infrastructure
that enhances patient care, sustains the NHS
and supports innovation

Data for R&D objectives



- Develop an **England-wide ecosystem of NHS owned and managed infrastructure** to allow researchers access to secure, high-quality, linked datasets, by:
 - **Expanding the scope, capacity and capability** of a national health data infrastructure platform
 - **Developing a network** of sub-national/regional linked health data infrastructure platforms and services
- Enhance **patient and public support and enthusiasm** for use of NHS health data to drive research and development
- **Increase clinical trials capacity** by underpinning clinical trial operations with NHS health data, improving the efficiency of clinical trial set up; and increasing the volume of people recruited to participate in clinical trials
- Expand the ability for researchers to **access a range of genomics datasets through linkage** of sources.
- **Generate returns to the NHS** by applying commercial principles consistently and coherently across the programme to generate value for the NHS.

An interoperable system of TREs





NHS
Digital

Thank You



@nhsdigital

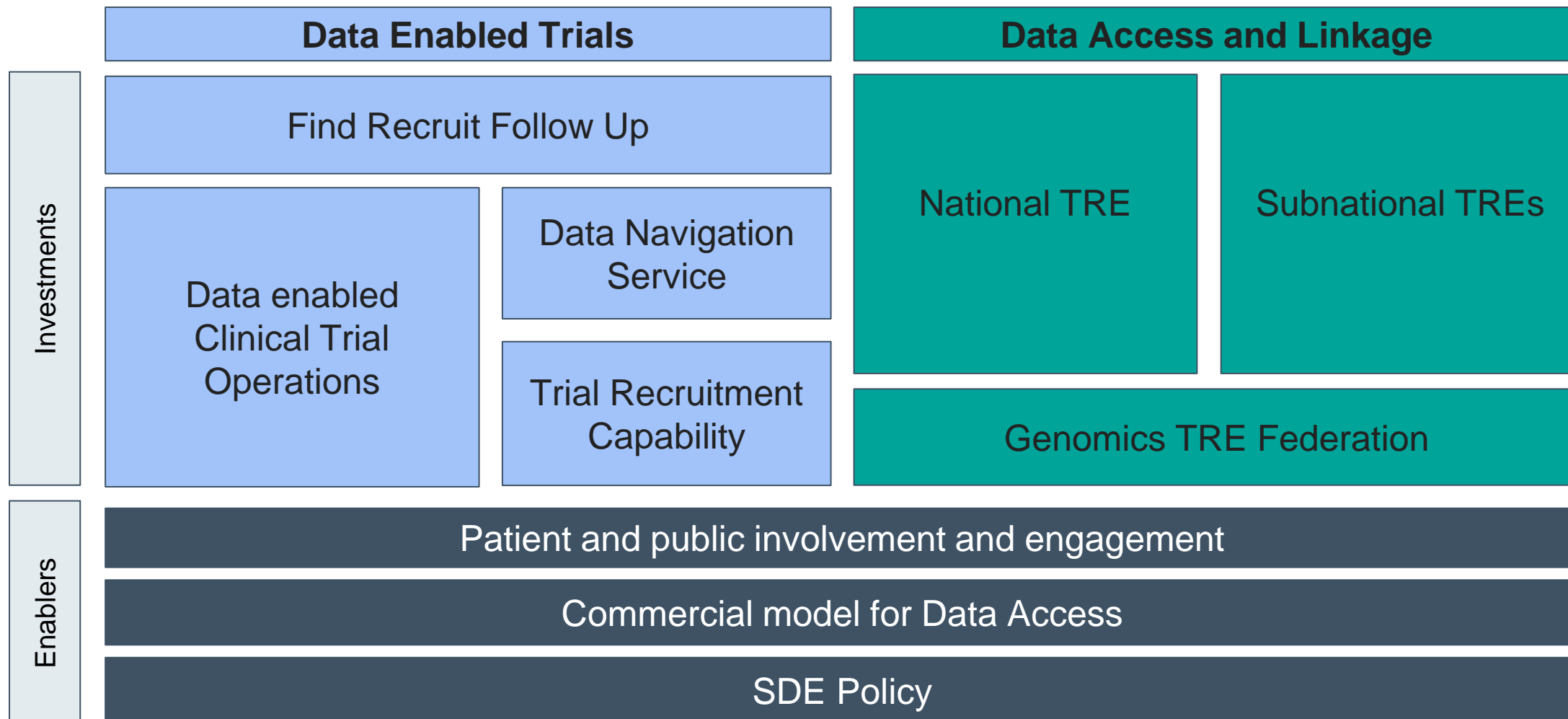


company/nhs-digital



digital.nhs.uk

Programme on a page



Policy principles

11 high level guidelines that will guide our approach to SDEs. Published in recent Data Strategy

Technical capabilities specification

Defines technical capabilities required of SDEs, including security, functionality, access and data management.

Roadmap

A guidance document that explains the transition to SDEs. How the SDE guidelines, tech spec and accreditation should be implemented and over what timeline.

Accreditation framework

Details Information Governance standards that SDEs must adhere to, and how this will be assessed and monitored.



Engagement approach: PPIE



‘Stacked’ throughout the programme, including:

- Learning from others
- PPIE lead within Programme
- Quantitative research
- Representation on governance structures
- Co-design panels
- Citizens juries
- Public deliberations
- National communications
- Engagement with National Data Guardian, Medical Research Charities, and patient lobby groups

