The 100,000 genomes project and beyond

Tim Hubbard

Professor of Bioinformatics, King's College London

Head of Genome Analysis, Genomics England

Associate Director, Health Data Research UK London Site

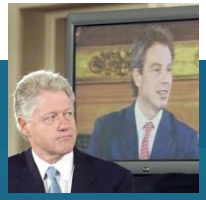
X-omics festival 2020

Monday April 6 2020 @ Nijmegen-NL

The first human genome sequence

- 26th June 2000 - Cost $3.2 billion
- 100,000 Genomes at Millennium Prices - Cost $320 trillion

Cost per genome



The 100,000 Genomes Project

Announced by the former Prime Minister in December 2012 – An Olympic Legacy

Announced by the former Prime Minister in December 2012; Genomics England formally launched by then Secretary of State for Health in speech during NHS 65th Anniversary Celebrations, July 2013

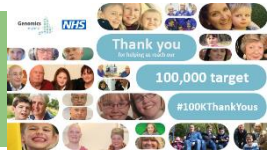
Opening of new Sequencing Centre in 2016

CMO's Generation Genome and the Life Sciences report in 2017

NHS England

Commissioning of new NHS Genomic Medicine Service October 2018

Reached goal of sequencing 100,000 genomes in December 2018



Recommended targets

- 2013 - Professor Dame Sally Davies (CMO) established a **Strategic Priorities Working Group** for the Project - chaired by Professor David Lomas (UCL)
- Recommended rare diseases, certain cancers, and infections
- Areas where they believe the introduction of genomic technology will have the greatest benefit for patient health

100,000 genomes project

Announced end 2012; Genomics England created 2013

- Primarily a treatment project
  - NHS transformation project
- All whole genome sequencing (clinical grade >30x)
  - Rare disease (3 genomes: affected individual and parent)
  - Cancer (2 genomes: normal tissue/tumour tissue)
- Mission
  - Improve Health of individual NHS patients
  - Create legacy of infrastructure, human capacity and capability in NHS
  - Stimulate wealth generation in the Economy
  - Enable large scale genomics research

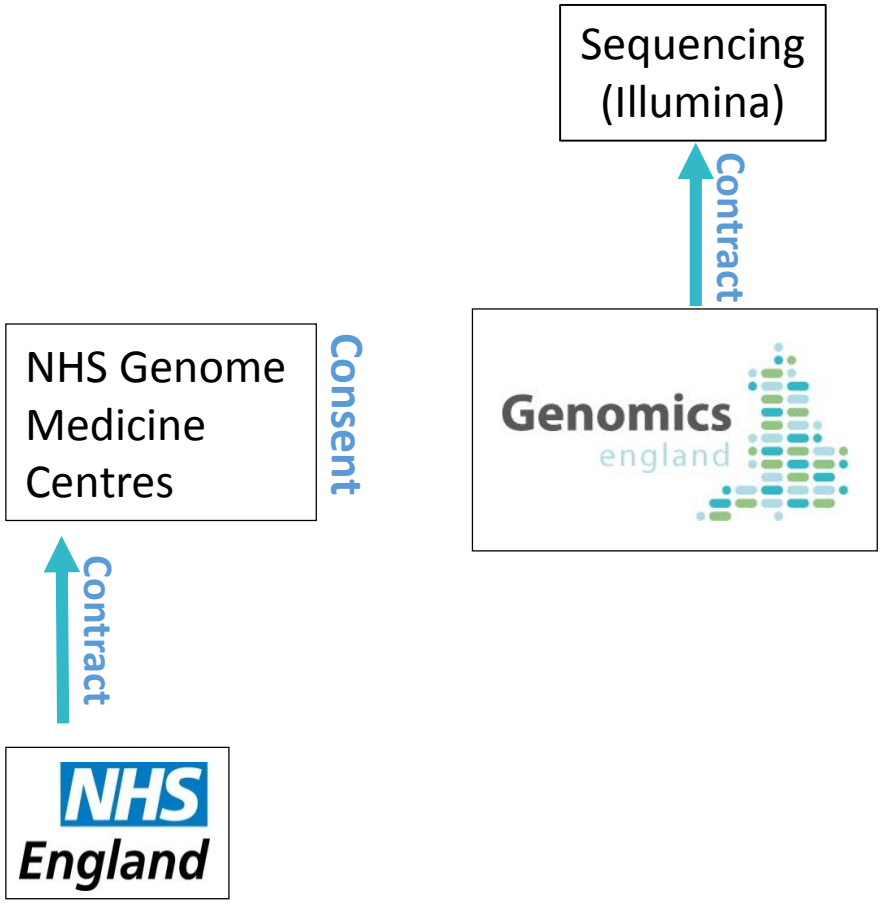
100,000 genomes project

Three phases towards sustainability

- Pilot (2014)
  - Through Biomedical Research Centres
- Main Programme (2015-2018)
  - Through Genome Medicine Centres
- NHS Genome Medicine Service (2018-)
  - Through NHS testing directory:
    National Genomic Information Service (NGIS);
    Genome Laboratory Hubs

Result: sustainable framework for genomic medicine embedded in NHS for clinical care and research

Genomics England - Clinical

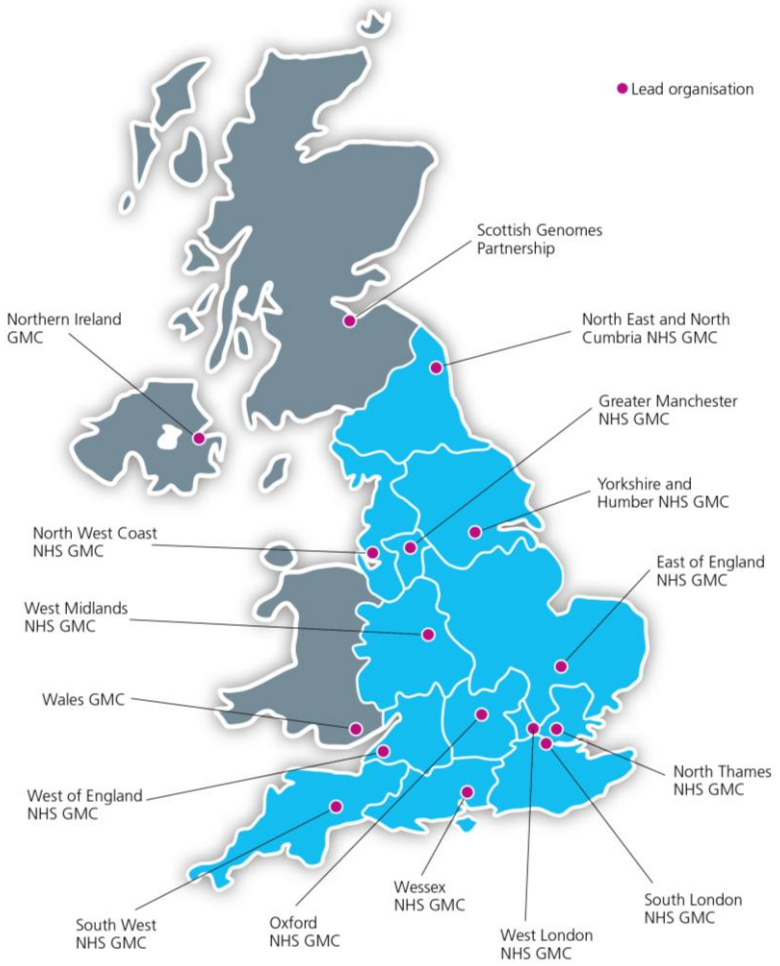


NHS Firewall

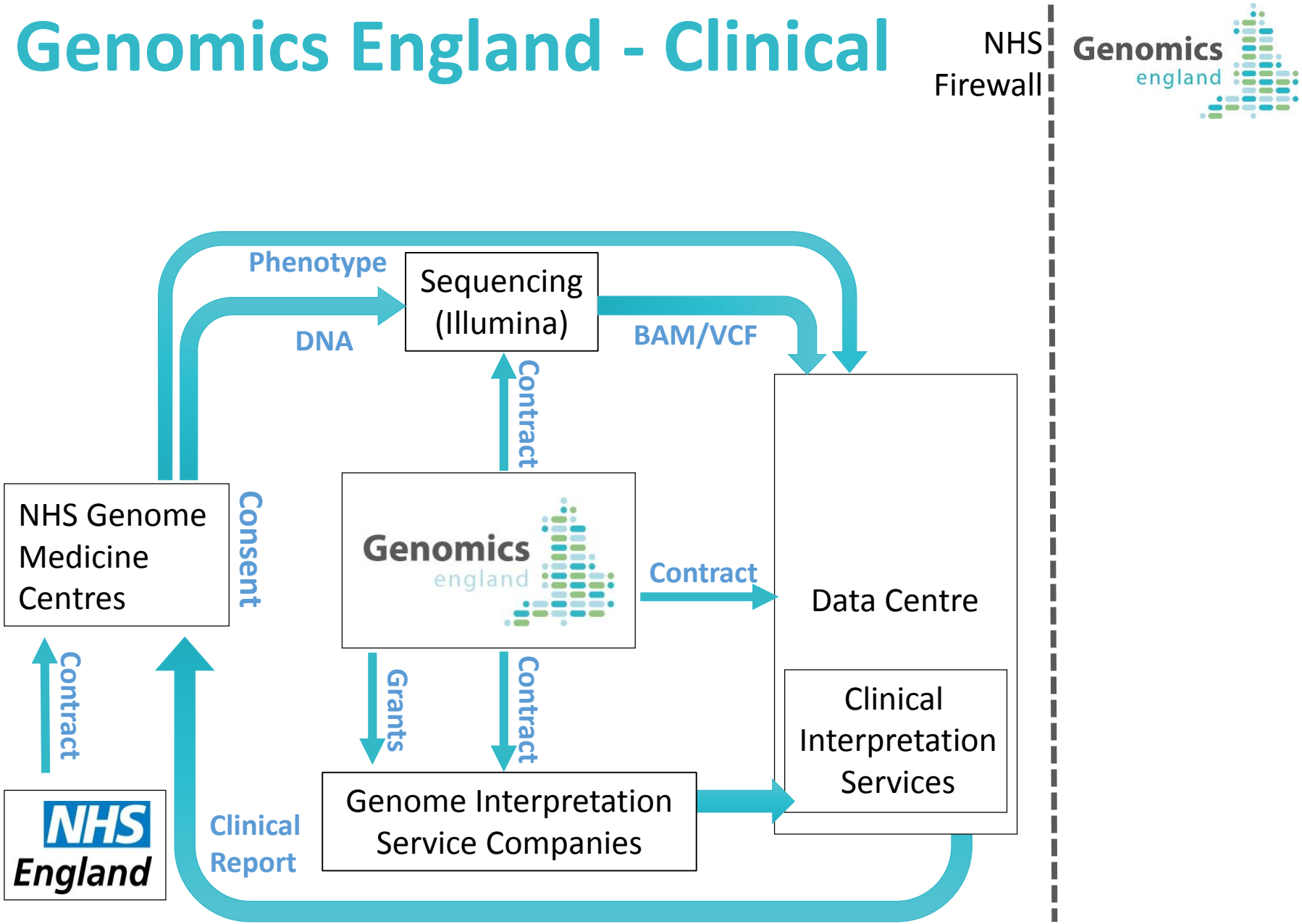
Genomics england

NHS Genomic Medicine Centres

- 13 Genomic Medicine Centres covering England
- Joined by NHS in Scotland, Northern Ireland and Wales
- Responsibilities:
  - identifying and recruiting participants
  - clinical care following results



Genomics England - Clinical



What are we telling participants?

- Information about a patient's main condition
- Information about additional 'serious and actionable' conditions (optional)
- Carrier status for non affected parents of children with rare disease (optional)

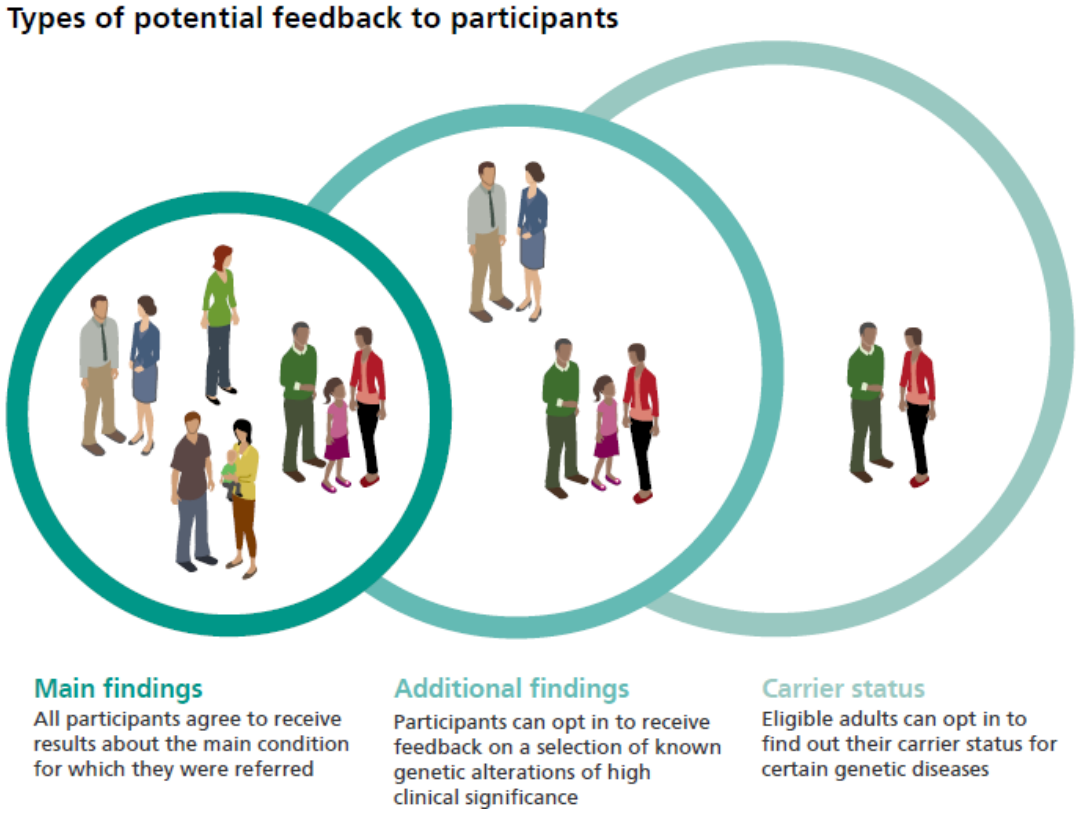


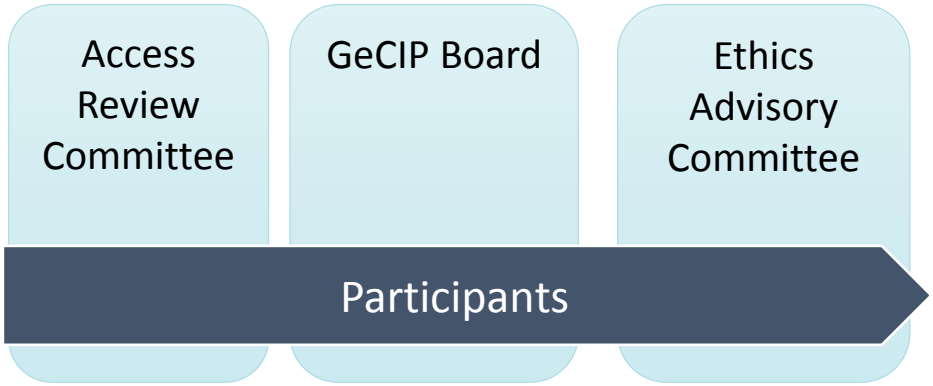
Image courtesy of Health Education England

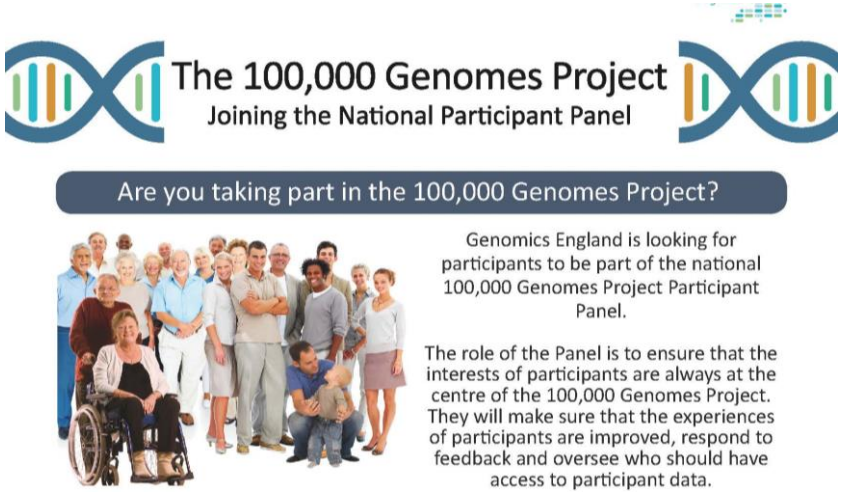
Patient involvement - the National Participant Panel

Role of the Panel is to ensure the interests of participants are always at the centre of the 100,000 Genomes Project.

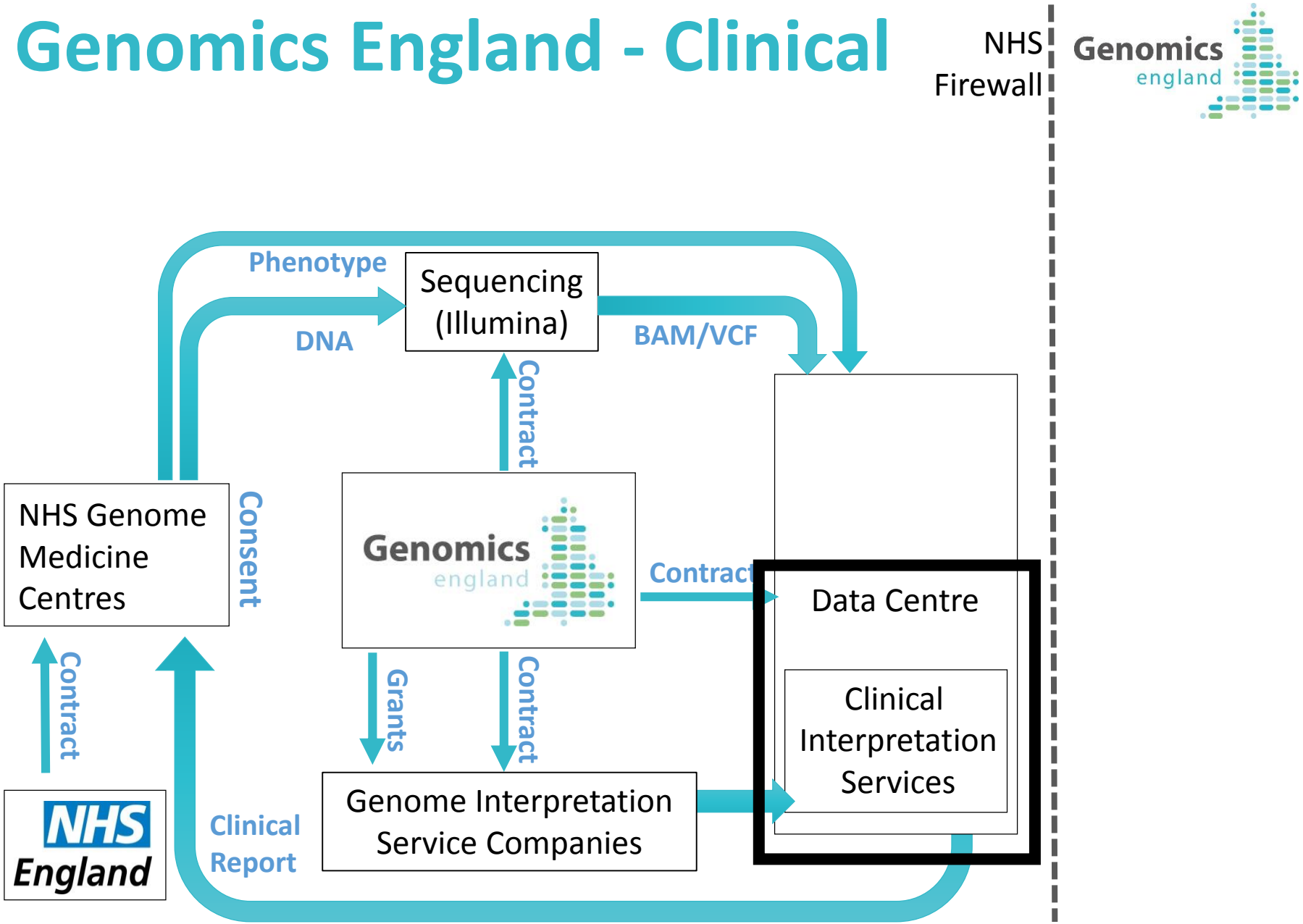
They do this by:

- Making sure experiences of participants are at the heart of the project
- Responding to feedback.
- Overseeing who should have access to participant data

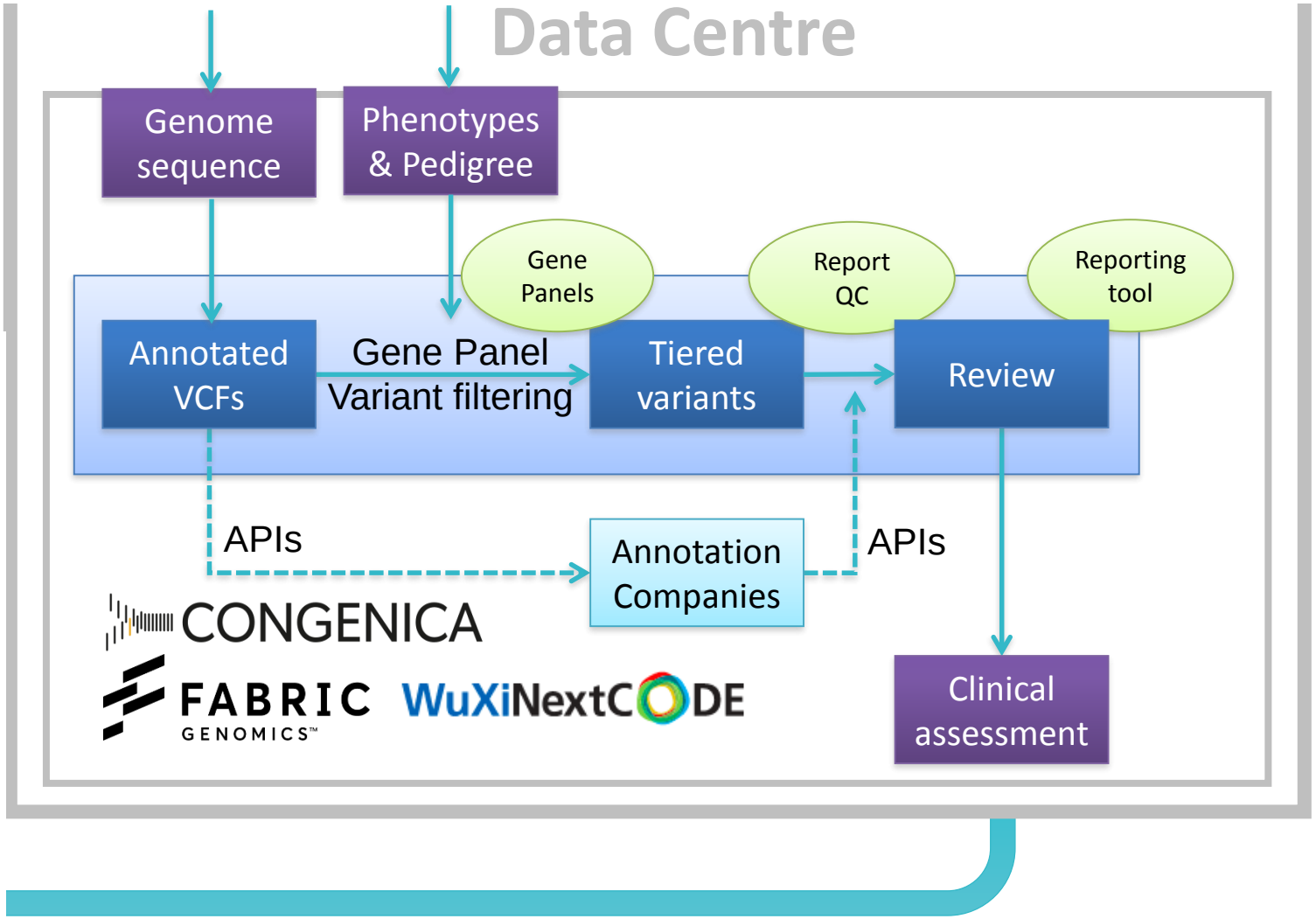




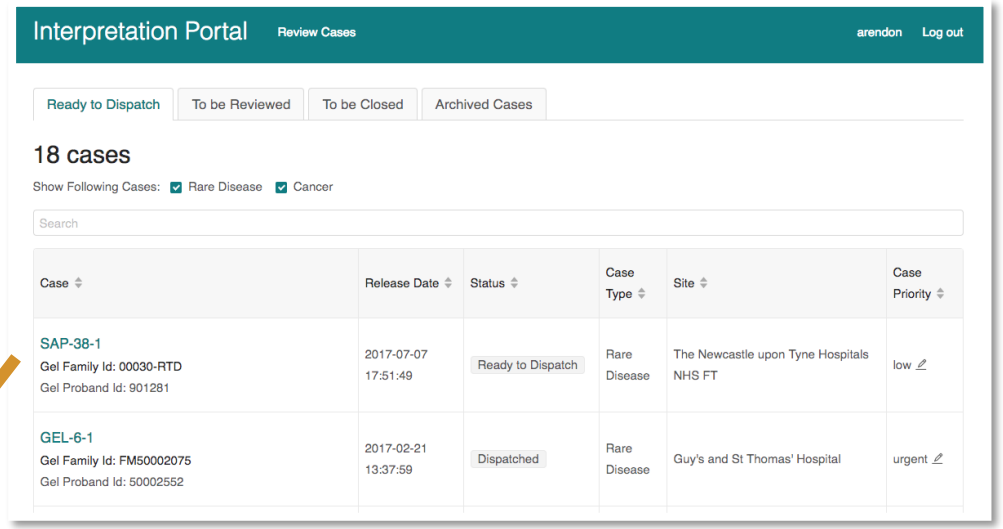
Genomics England - Clinical



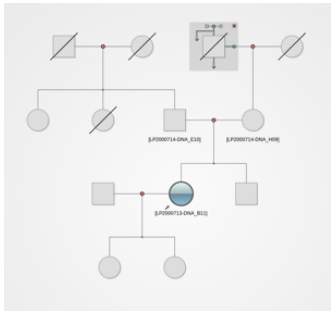
Scalable rare disease diagnostics



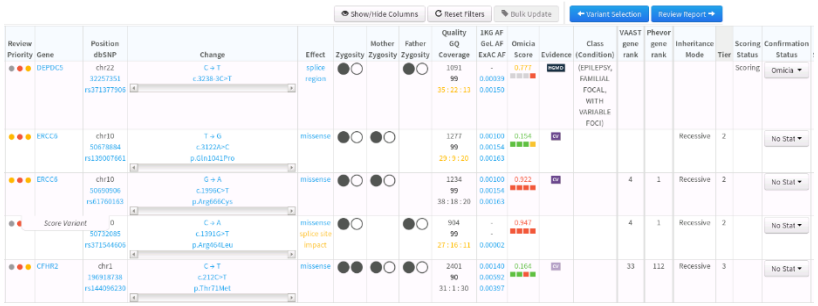
Reporting back to the NHS



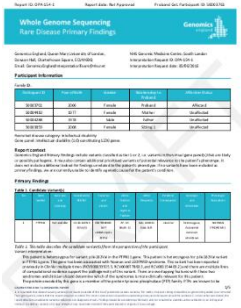
1. View family pedigree



2. Review variants and close case



3. Download the report



Progress to date (100,000 Genomes Project)

Figures as at 06/09/2019

Genomics england

Samples

122,941
Samples collected and received at the UK Biocentre

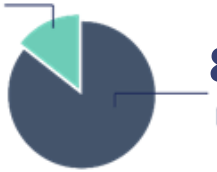
36,868 cancer

86,073
rare disease

Genomes

116,268
Genomes sequenced

32,720 cancer



83,548
rare disease

+406
genomes
since last month

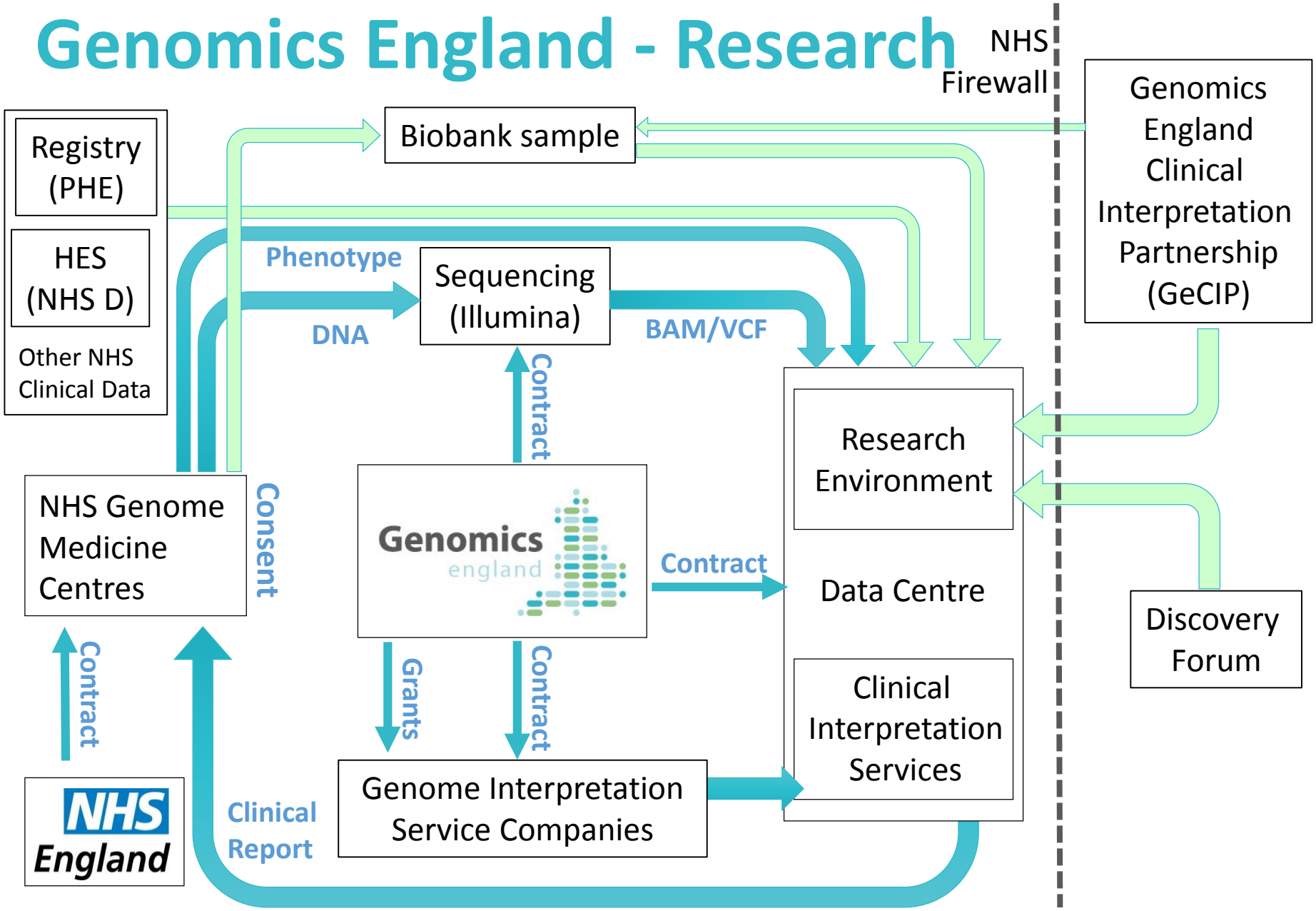
Analysis and Results

Results for **98,056** genomes sent to NHS GMCs

Equivalent to **28,047** cancer genomes and **70,009** rare disease

- 20-25% actionable findings for Rare Disease
- ~ 50% cancer cases contain potential for a therapy or a trial in our report

Genomics England - Research



Genomics England Clinical Interpretation Partnership in numbers

(As of 17th July 2020)

GeCIP launched in **June 2014** @ Wellcome Trust

3388 researchers world-wide

431 registered research projects

131 Project publications

24 Successful grant applications

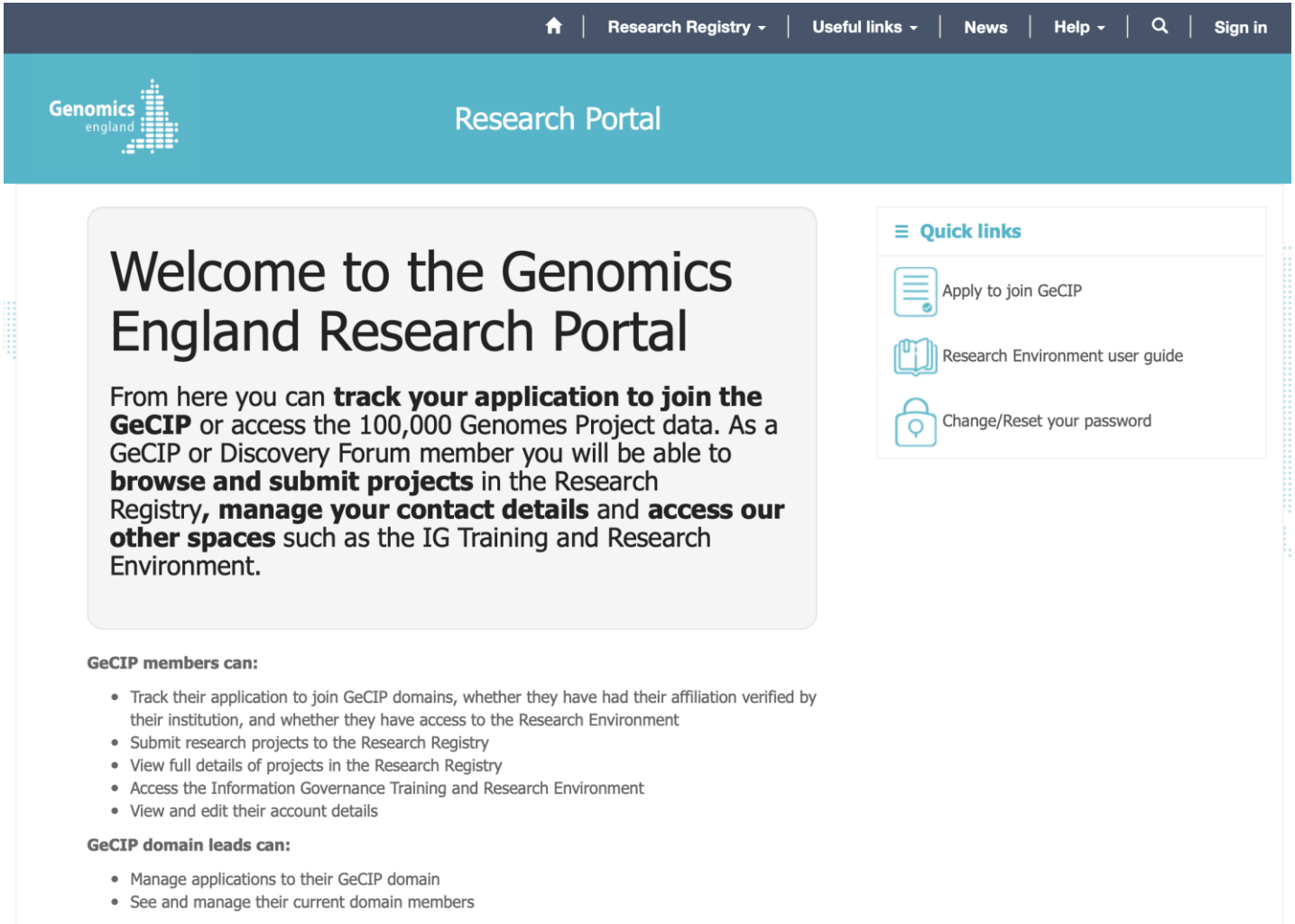
£28.9 Million total awarded

175 institutions with signed Participation Agreement

1089 researchers with access to data

GeCIP Research Portal

https://research.genomicsengland.co.uk



Genomics England Research Environment at a glance

Data and documentation

Genomes (BAM and VCF) in Isilon share

Clinical data in LabKey

Confluence

- data release notes
- user guides
- airlock
- live issues

Tools and analysis

Virtual desktop interface provides GUI and security

LibreOffice for document editing

R and Rstudio for data analysis

Internet browser: access to whitelisted sites

Command-line tools and HPC cluster for large-scale analysis

Collaboration

Domain-specific and shared storage for files

Social media platform for communication

Research registry:

- promote collaboration
- enforce publication moratorium

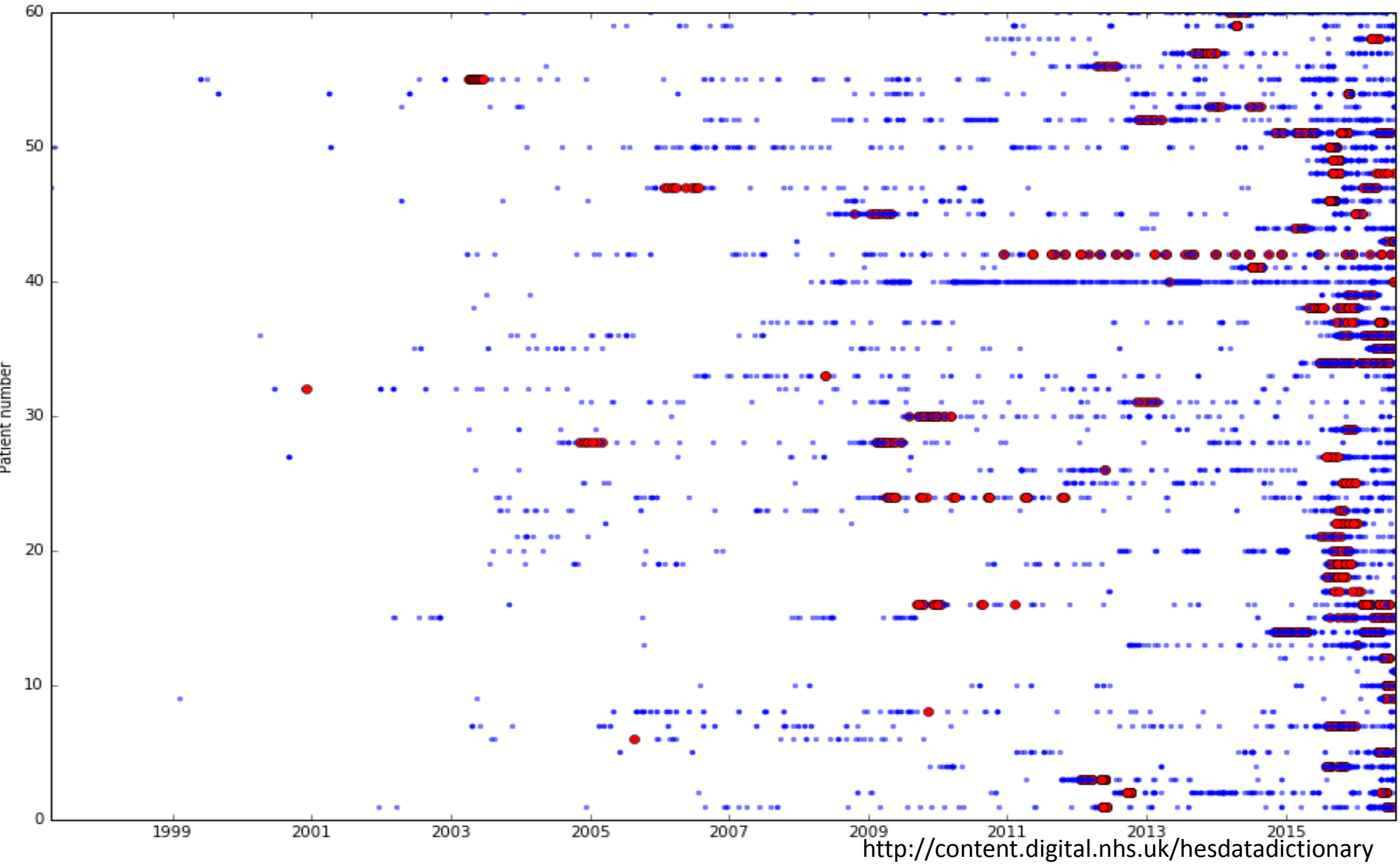
Data in our Research Environment

10th release: September 2020

Genomes	**111,232 genomes** • **37,224 Cancer** • **74,008 Rare Disease**	**Primary clinical data**	**89,256 participants** • **17,339 Cancer** • **71,800 Rare Disease**
Secondary data	• Hospital Episode Statistics (HES) • Patient Reported Outcome Measures (PROMs) • Mental Health Services Data Set (MHSDS) • Uncurated SACT (chemotherapy) data		• Office for National Statistics (ONS) – mortality data and cancer flagging • COVID-19 diagnosis data for 100,000 Genomes Project participants
Clinically interpreted data & QC	• **33,827 families** with Tier 1, 2 and 3 variants from interpretation pipeline • **20,032 families** with GMC exit questionnaires • **61,138** tiered and quality checked rare disease genomes; **31,590** quality checked cancer genomes	**Quick view tables**	• COVID-19 diagnosis data • Real time uncurated SACT data • Aggregated gVCF dataset • *de novo* variant dataset for 13,836 trios • PHE Mental health data

Awaiting – Primary Care, Prescribing Data

Hospital Episodes delving deeper 1997-2005
Previous treatment, 61 patients care pathways



http://content.digital.nhs.uk/hesdatadictionary

Opportunities for GeCIPs

- Interpret cases where CIPs (Clinical Interpretation Providers) currently fail
- Develop clinical applications against stored WGS
  - Phamacogenics; Polygenic Risk Scores
- Improved interpretation algorithms
  - machine learning; artificial intelligence
  - using whole genome; predicting variable penetrance
- Experimental investigation of function of variants
  - Is it really the cause? How does it function?

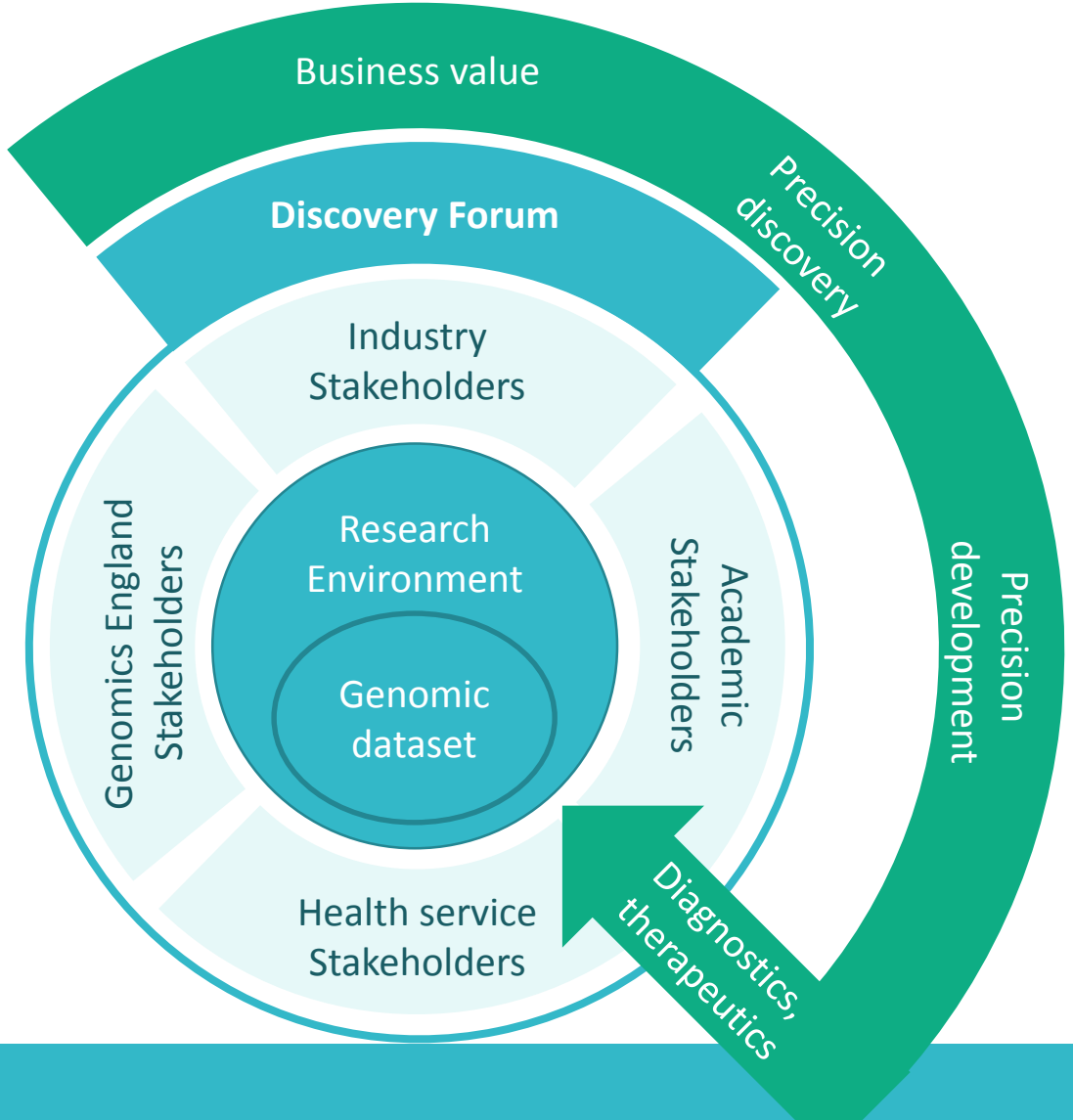
Bioinformatics / Analysis

Lots of progress still required

- Better alternatives to current reference genome
  - Graph genomes, incorporating all human haplotypic variation
- Better ways to handle large numbers of WGS, VCF
  - GEL using OpenCGA (Hadoop), Hail (spark)
- Better interpretation algorithms
  - Non coding variants, SVs, CNVs, Variable penetrance etc.
- Better sequencing technology
  - Long reads

The Discovery Forum

A driver of translational research



- **Exploring** the business value of genomic medicine data.
- **Connecting** industry stakeholders to the Genomics England community.
- Providing a **gateway** to our Research Environment and dataset.
- Leading to **discovery** and development of precision methods, diagnostics, and therapeutics.

The 100,000 Genomes Project in numbers

Genomics england

Over **100,000** genomes

Over **97,000** patients and family members

21+ Petabytes of data.
1 Petabyte of music would take 2,000 years to play on an MP3 player.

13 Genomic Medicine Centres, and **98** NHS Trusts within them were involved in recruiting participants

Around **5,000** NHS staff (doctors, nurses, pathologists, laboratory staff, genetic counsellors)

Over **3,000** researchers and trainees

5 million genomes aspiration announced in October 2018

On the 2nd October, Matt Hancock the Secretary of State for Health and Social Care, announced an ambitious vision for genomic healthcare in the UK…

> Expansion of the 100,000 Genomes Project to **one million whole genomes** sequenced by NHSE and UK Biobank in the next five years

> From 2019, the NHS will offer **whole genome analysis for all seriously ill children with a suspected genetic disorder, including those with cancer.** The NHS will also offer the same for all adults suffering from certain rare diseases or hard to treat cancers

> An **aspiration to sequence 5 million genomes** in the UK within the next five-year years

The strategic aspiration: 5 million genomes over 5 years aspiration is system wide

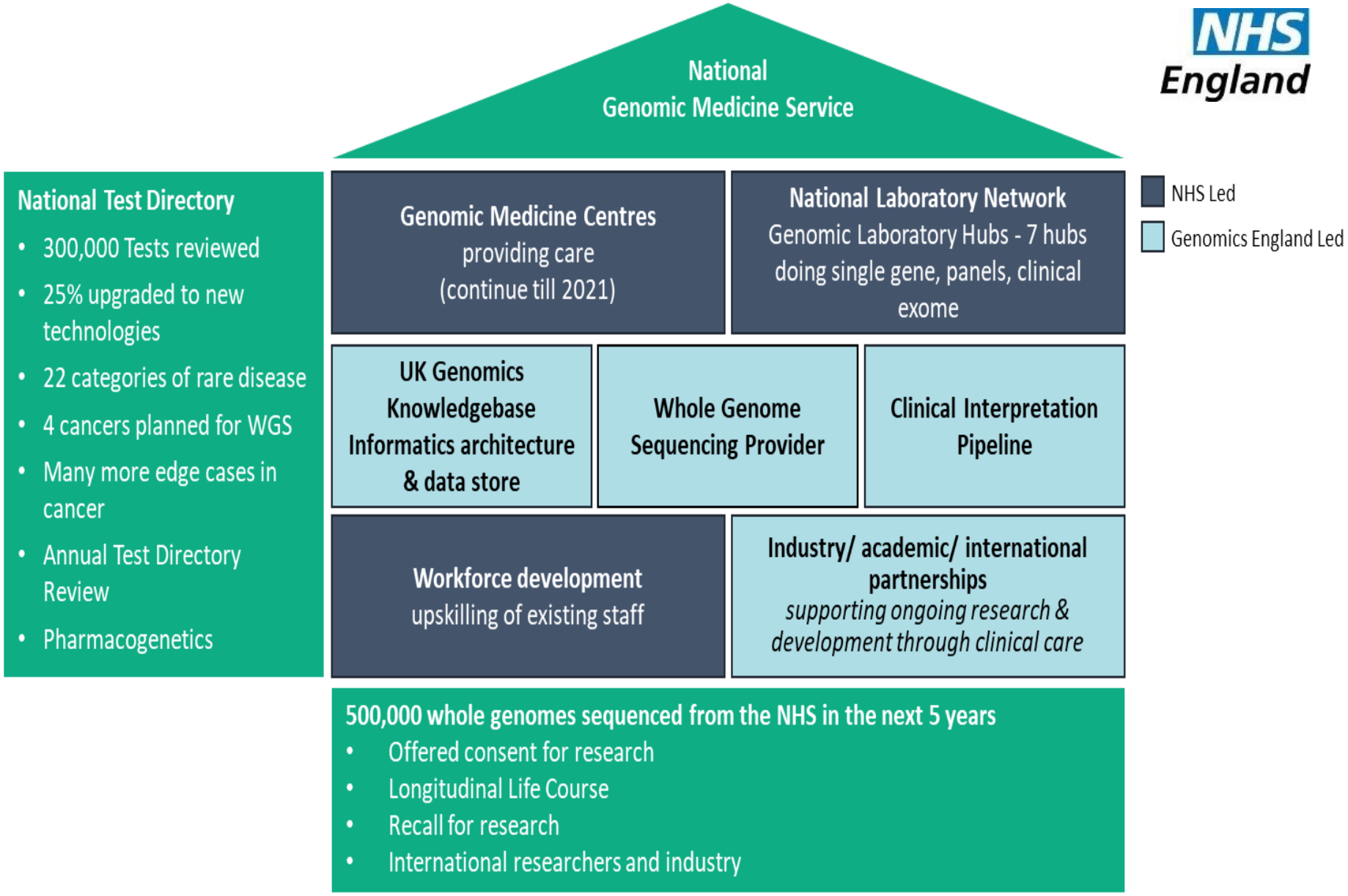
- 500,000 whole genomes through the **Genomic Medicine Service**
- Plus up to 1.5 million through other genomic tests

- 500,000 whole genomes through upgrade of **UK Biobank** sequencing

- Whole genome sequencing of **strategic cohorts** focusing on unmet needs and long term value to the healthcare system

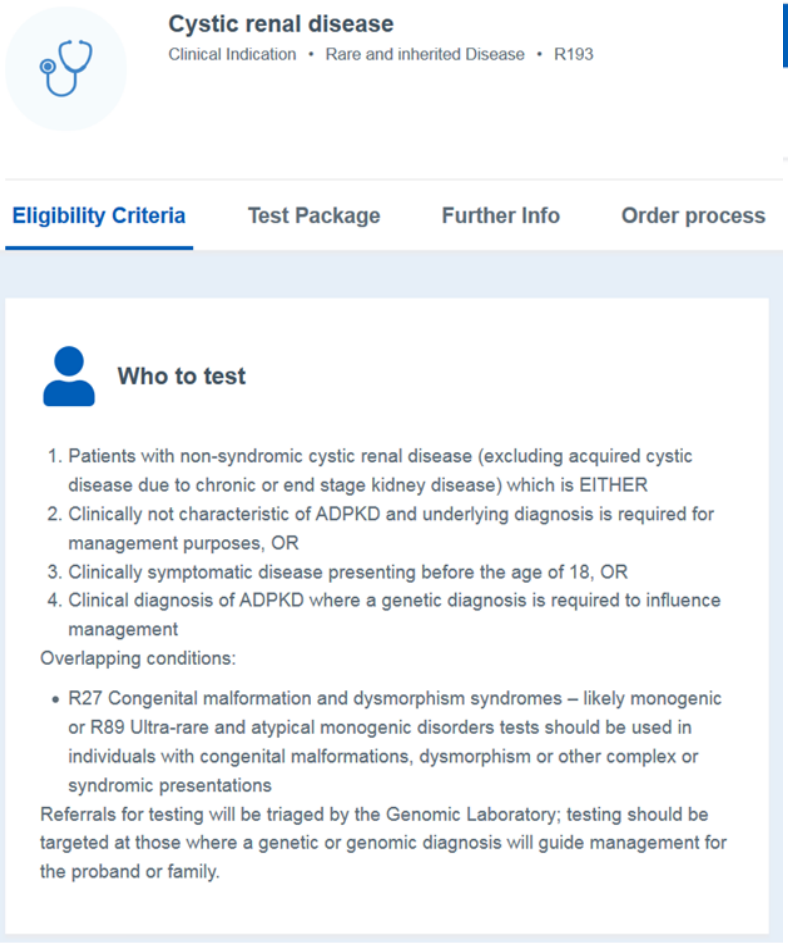
- Evaluating the benefits of whole genome sequencing of **newborns**

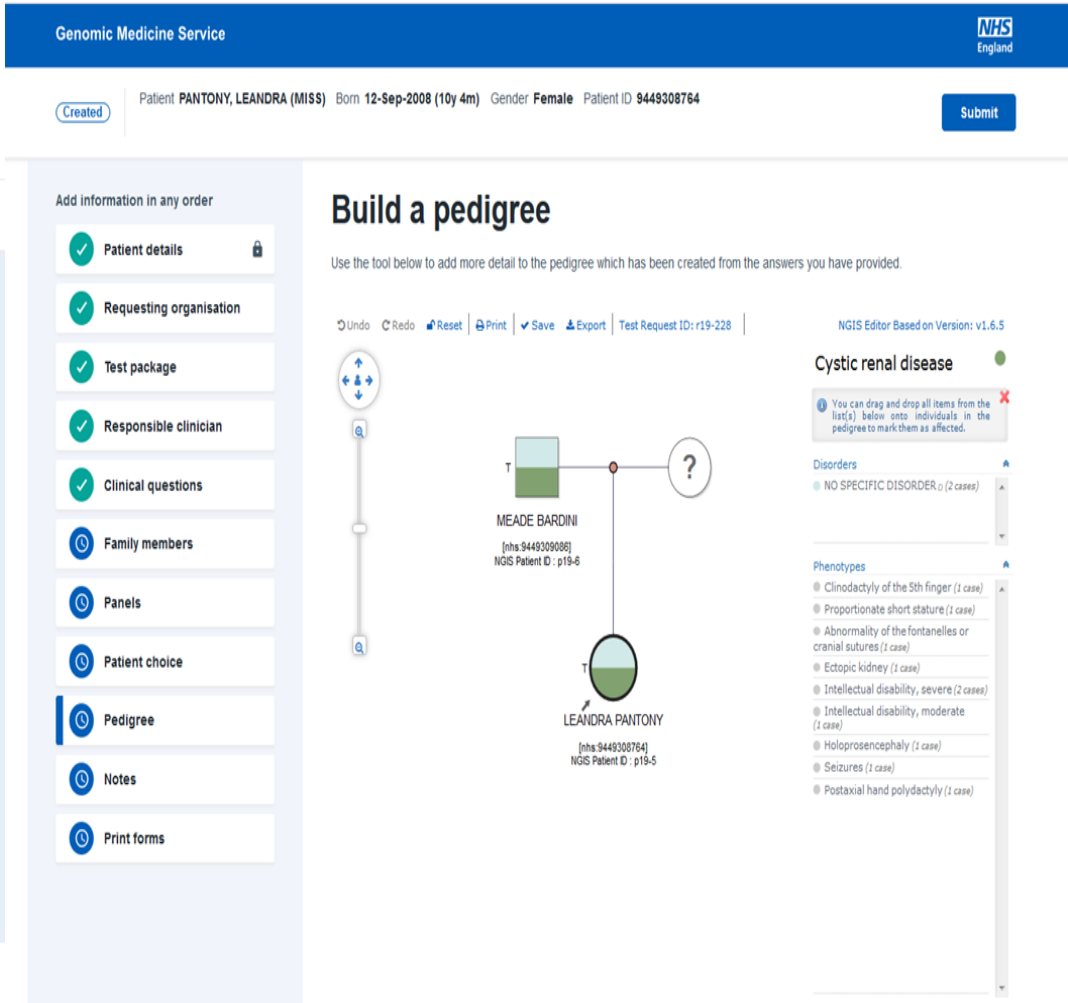
- 1 million+ accelerated detection of disease cohort


National
Genomic Medicine Service
NHS England
National Test Directory
• 300,000 Tests reviewed
• 25% upgraded to new technologies
• 22 categories of rare disease
• 4 cancers planned for WGS
• Many more edge cases in cancer
• Annual Test Directory Review
• Pharmacogenetics
Genomic Medicine Centres
providing care
(continue till 2021)
National Laboratory Network
Genomic Laboratory Hubs - 7 hubs doing single gene, panels, clinical exome
NHS Led
Genomics England Led
UK Genomics Knowledgebase Informatics architecture & data store
Whole Genome Sequencing Provider
Clinical Interpretation Pipeline
Workforce development
upskilling of existing staff
Industry/ academic/ international partnerships
supporting ongoing research & development through clinical care
500,000 whole genomes sequenced from the NHS in the next 5 years
• Offered consent for research
• Longitudinal Life Course
• Recall for research
• International researchers and industry

National Genomics Informatics System

Find Patients, explore and order tests





GenOMICC - CoG-UK Covid-19 Human Whole Genome Programme

NHS

ISARIC-GenOMICC NHS Framework 70+ Intensive Care Units
Severely Ill COVID-19 Patients -
Consented via an app, on admission or deferred consent
Dedicated human sampling or Unused routine samples (EDTA, plasma, serum)

Public Health England

Liverpool, Glasgow, & Edinburgh Centres/ NHS GLHs
DNA & multi-omics Repository

Sequencing Centre
Wellcome Trust £27m
Illumina Partnership

Genomics England Trusted Research Environment

Identifiable - NHS access only
- Immunodeficiency diagnosis

De-identified-Researcher access
- Baseline Clinical Data
- Life-course datasets
- Linked whole genomes
- Other omics data
- 100,000 Genomes WGS Dataset
- Analytical Tools

Genomics england

NHS Digital

Primary Care
Hospital episodes
Other registries
Mortality data
Patient entry

Intensive Care (ICNARC) Registry

Fire wall
Individual level data stays in safe haven

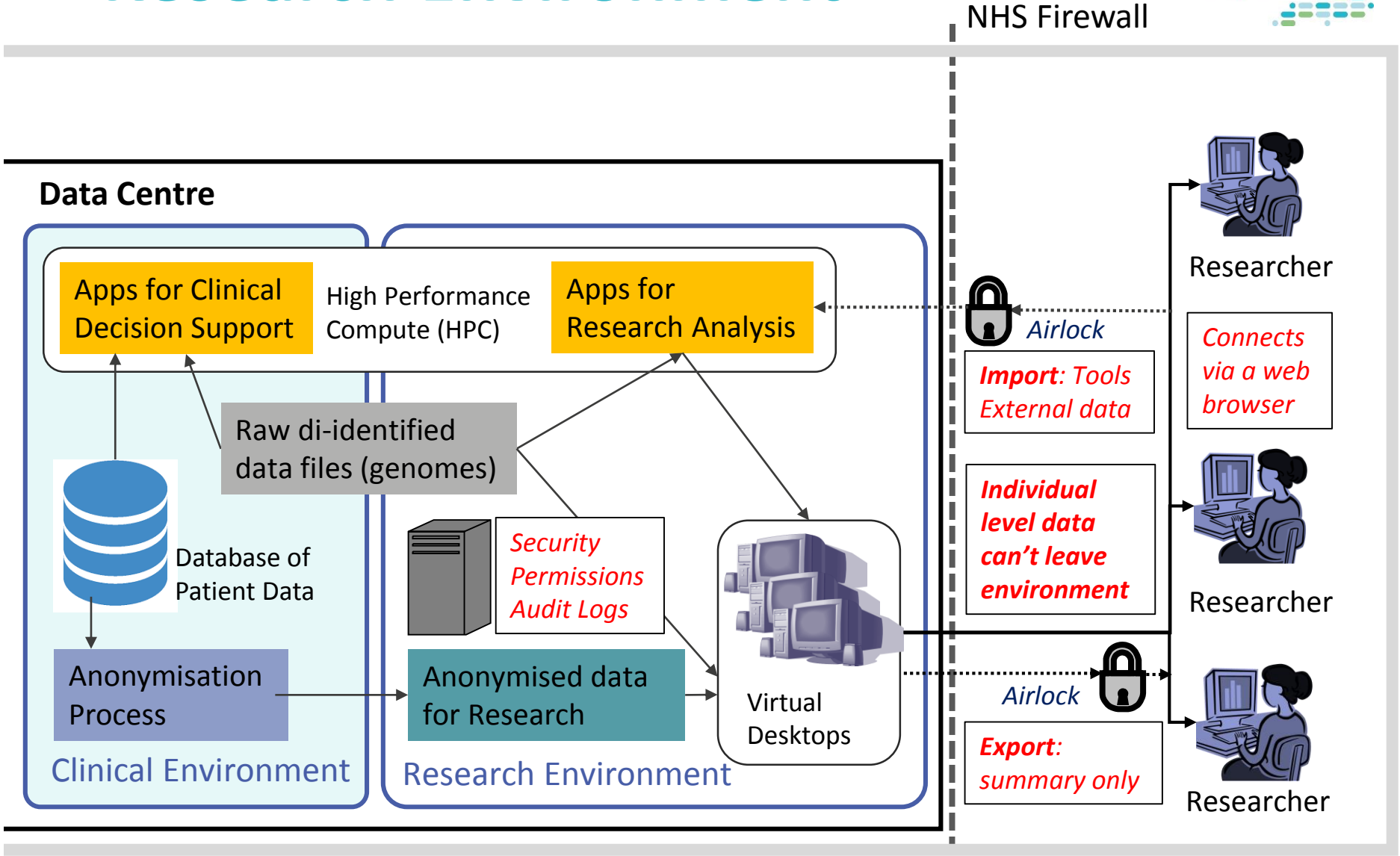
Only summary results pass outside For publication

ISARIC/GenOMMIC, National & International academics & industry
Extant coalition of 3600 researchers from 33 countries, industry

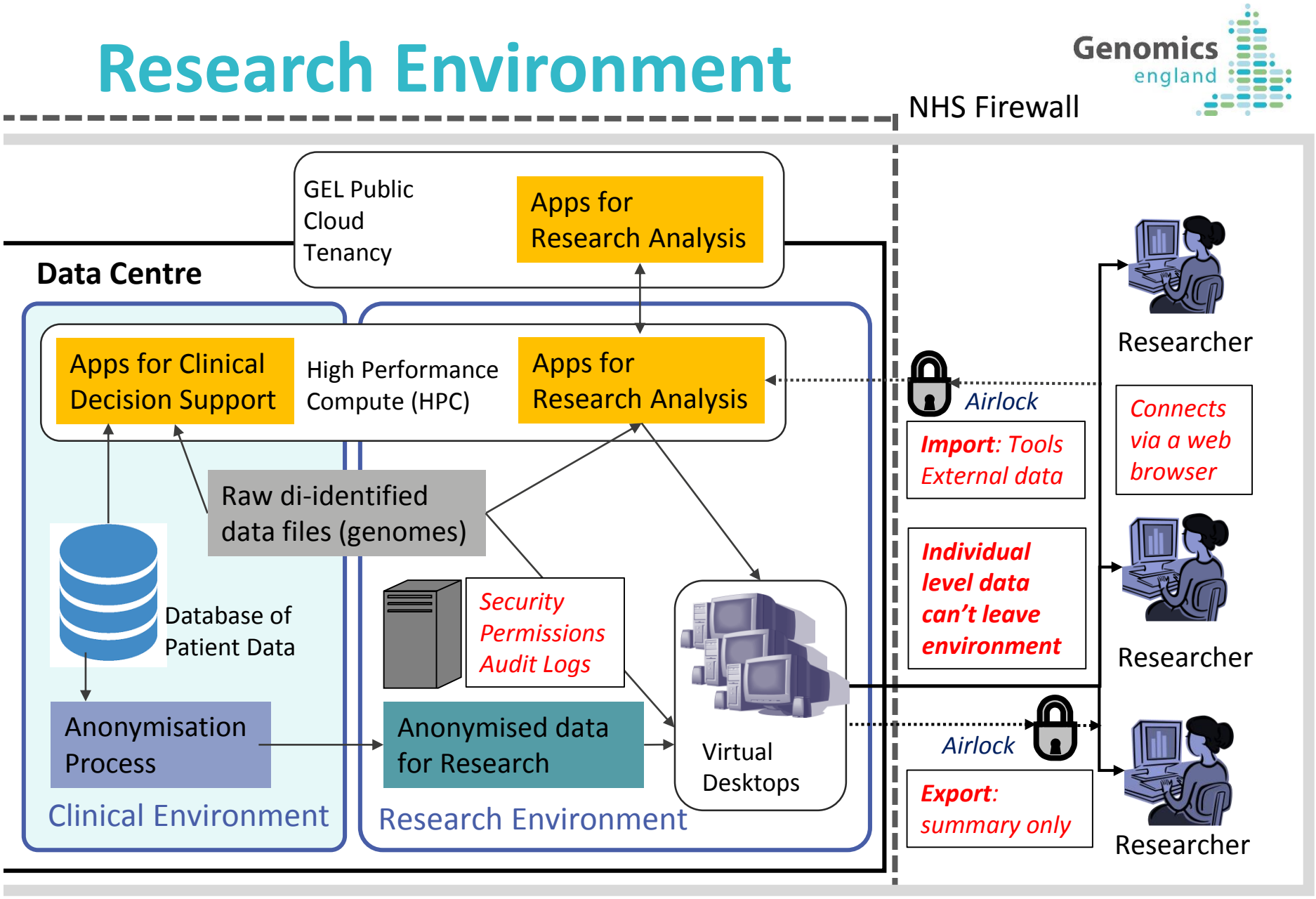
HDRUK Health Data Research UK

Genomics england

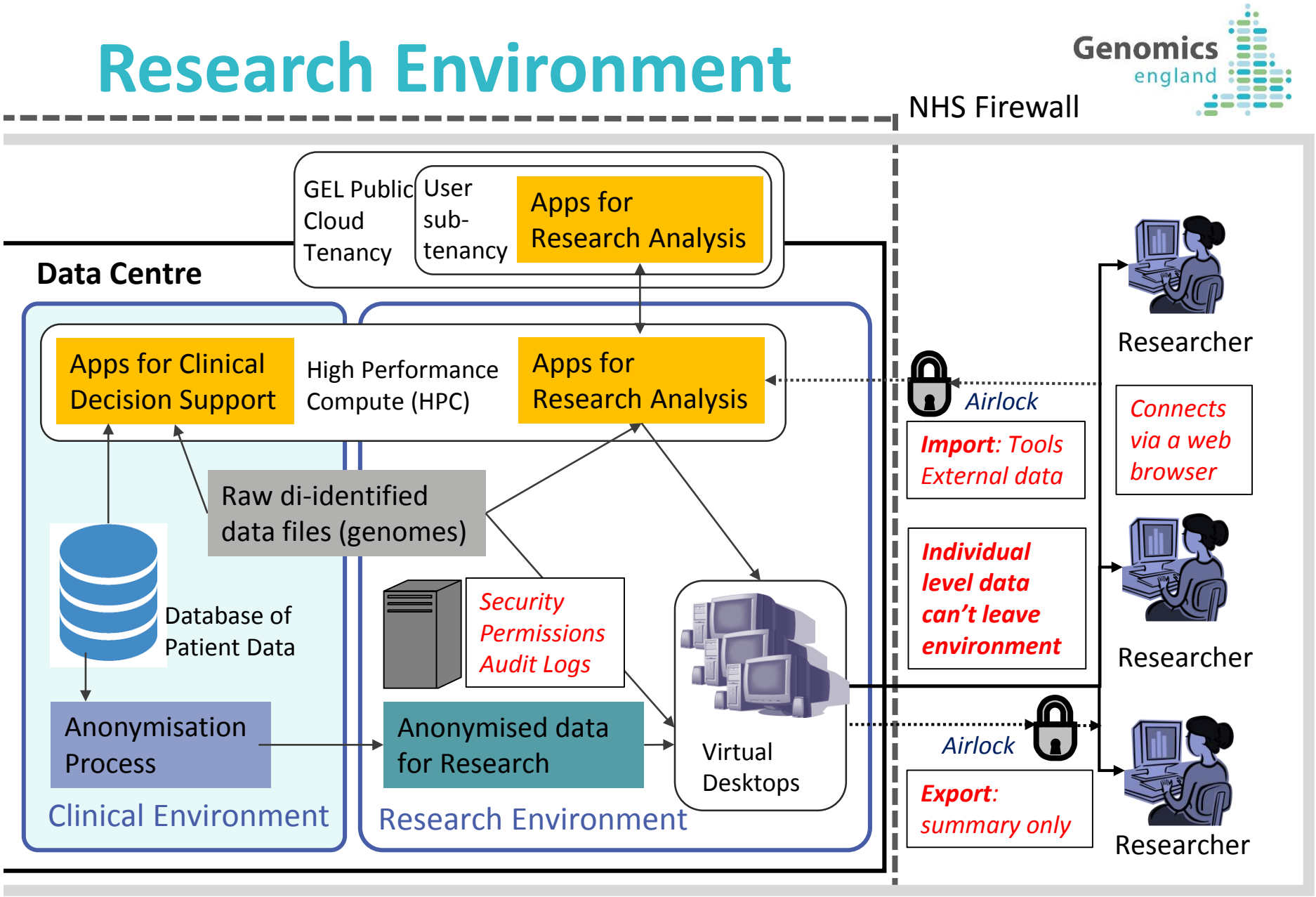
Research Environment



Research Environment



Research Environment



Research Environment evolution

Existing Research Environment

Data Asset: 100,000 Genomes Project

>108,000 genomes from >87,000 participants

Research Environment v1.0

>3,300 academic researchers, 175 institutions in **Genomics England Clinical Interpretation Partnerships (GeCIPs)**

20-25% actionable findings for Rare Disease

~50% Cancer cases with potential for therapy or trial

COVID Research Environment

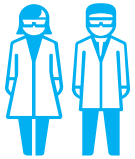
Data Assets: 100,000 Genomes Project and COVID-19

GenOMICC

Additional 35,000 genomes from severe and mildly affected COVID-19 +ve individuals

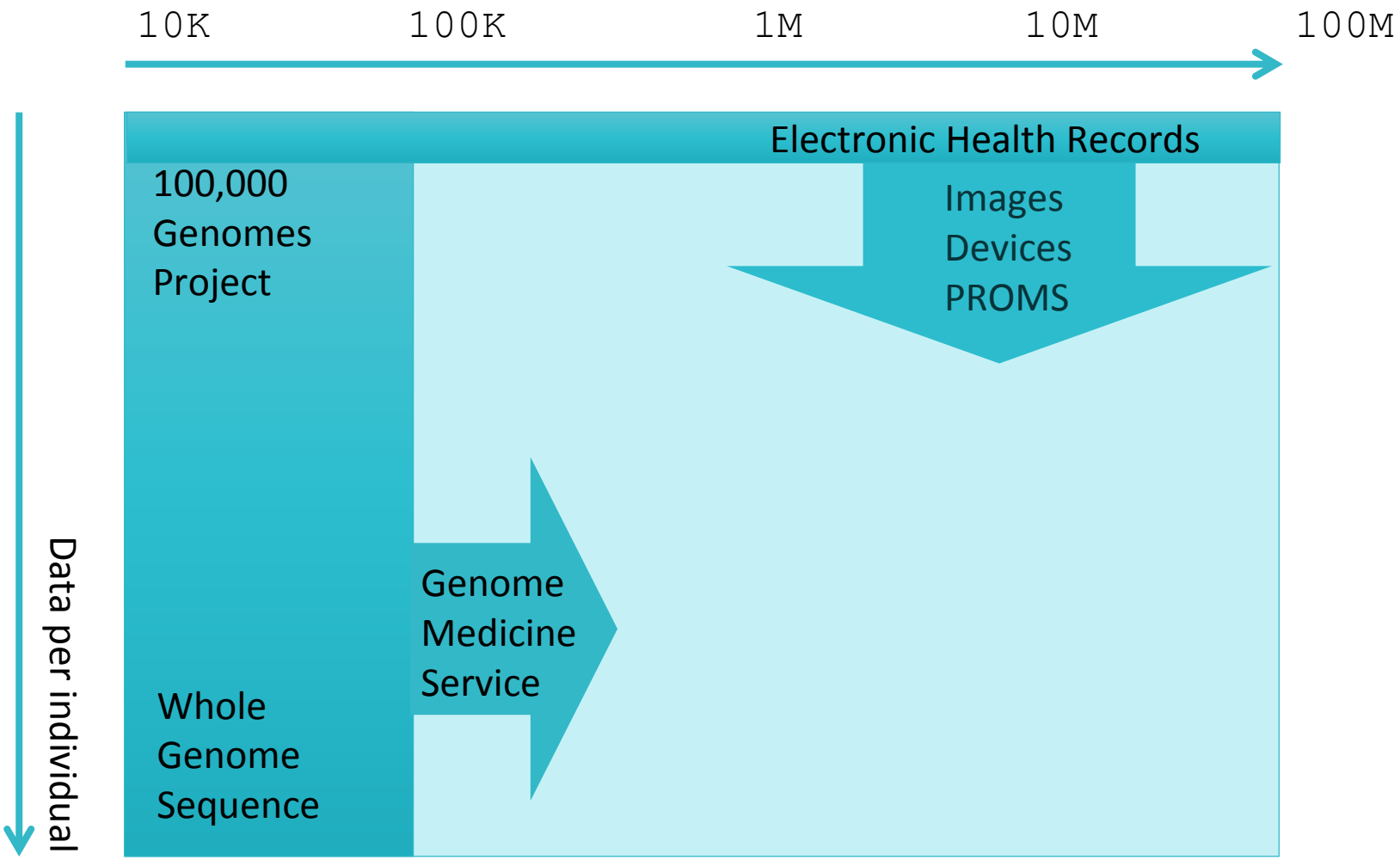
Newly established COVID-RE in partnership with AWS and Lifebit

aws lifebit



Collaborative academic and commercial **COVID-19 Consortium**

Expanding health data sets



A new national Institute for health data science

HDRUK
Health Data Research UK

History: Launched in April 2018 with selection of six initial sites

Mission: make game-changing improvements in the health of patients and populations through research and innovation.

How: Apply cutting-edge data science approaches to clinical, biological, genomic and other multi-dimensional health data to address the most pressing health research challenges facing the public

Funding: Medical Research Council, the British Heart Foundation, the National Institute for Health Research, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, Health and Care Research Wales, Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, and Wellcome.

https://www.hdruk.ac.uk/

HDR UK triple aim

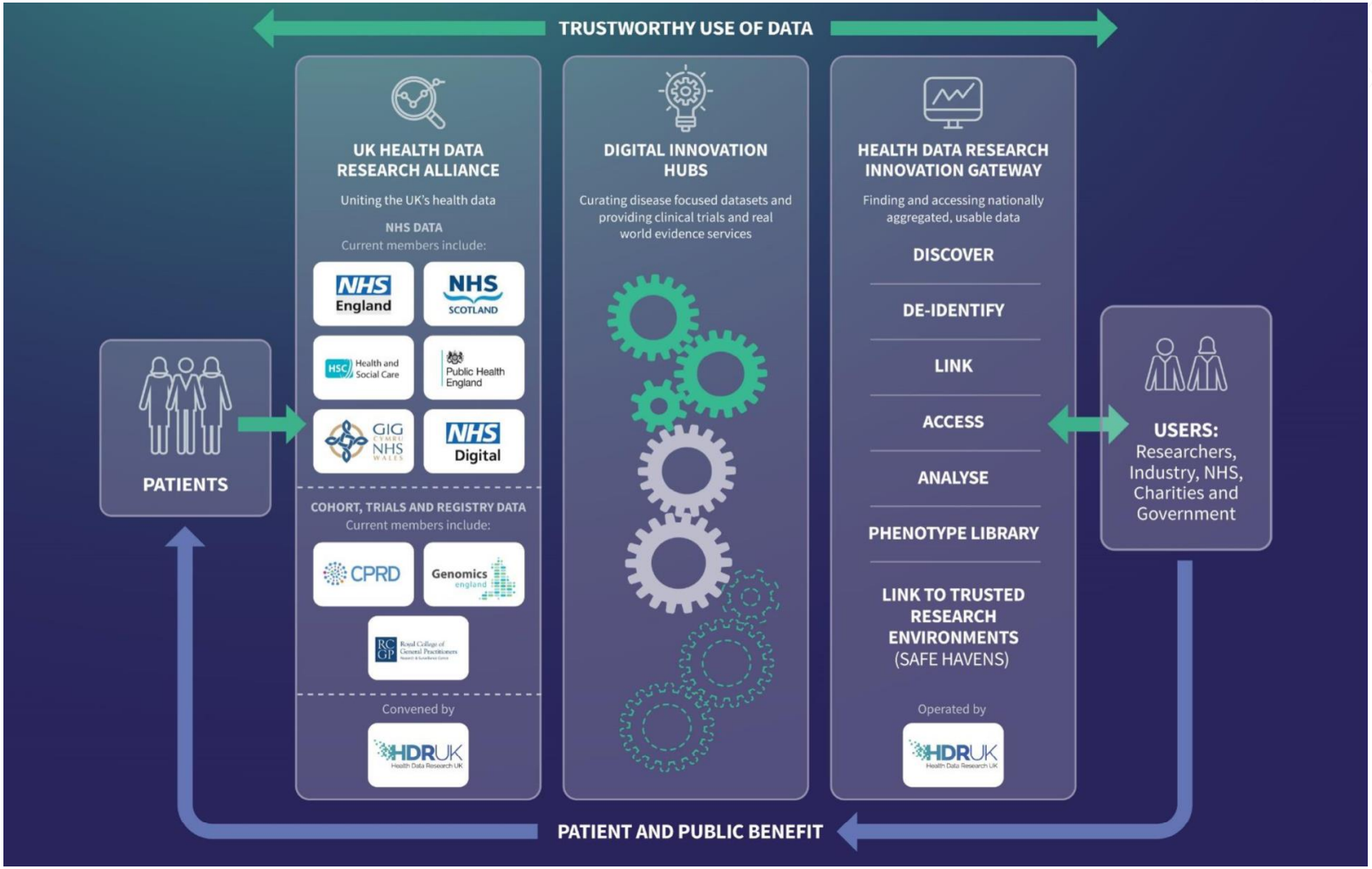
HDR UK
Health Data Research UK

Scientific programmes	Training the next generation	UK wide expert research data services
Integration of data science with biomedical and health science expertise to perform ground-breaking research, with an initial focus on data analytics, precision medicine, 21st century clinical trials and modernising public health.	To develop novel approaches to research training and mentorship to foster a cadre of health data science researchers, on a substantial scale.	Development and delivery of cutting-edge technologies and trusted research platforms that acquire, store, represent, and process large, multi-dimensional research data.

Trustworthy use of data

We will work in partnership with the public, funders, social scientists and legal/ethical experts to champion the trustworthy use of data.

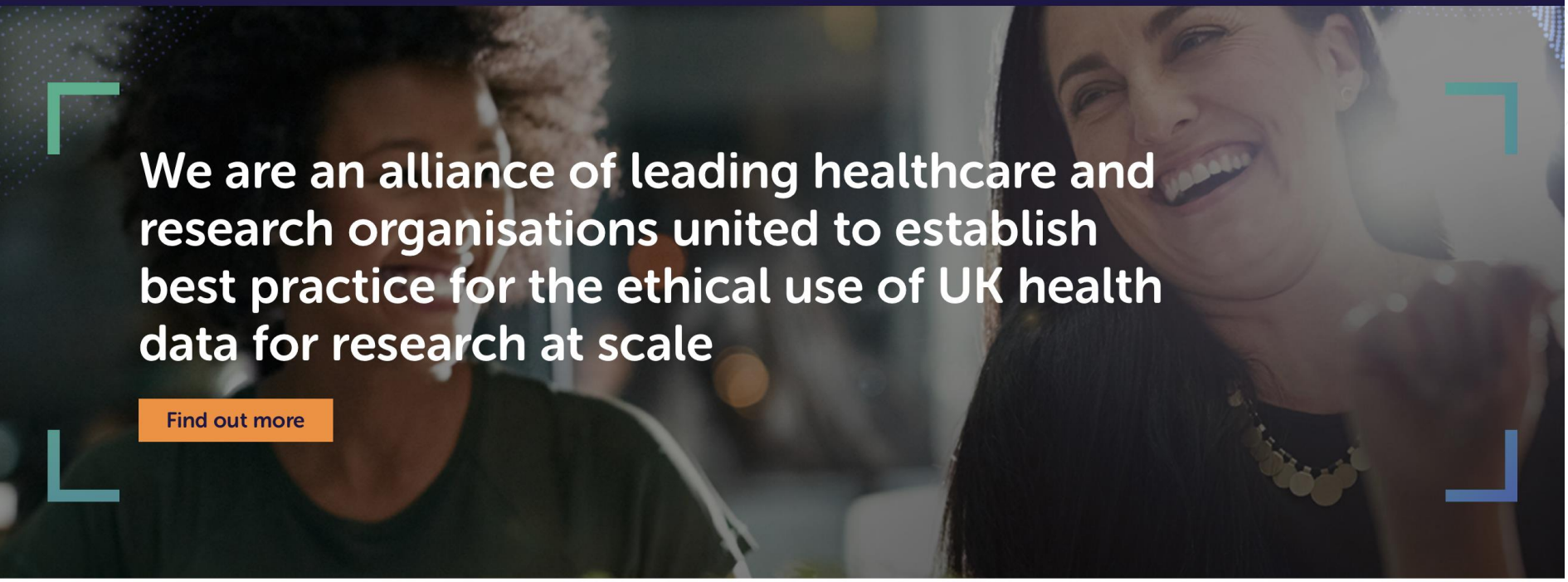
Infrastructure:
Uniting the UK's health data



UK Health Data Research Alliance

Search

About Members Work News & Events

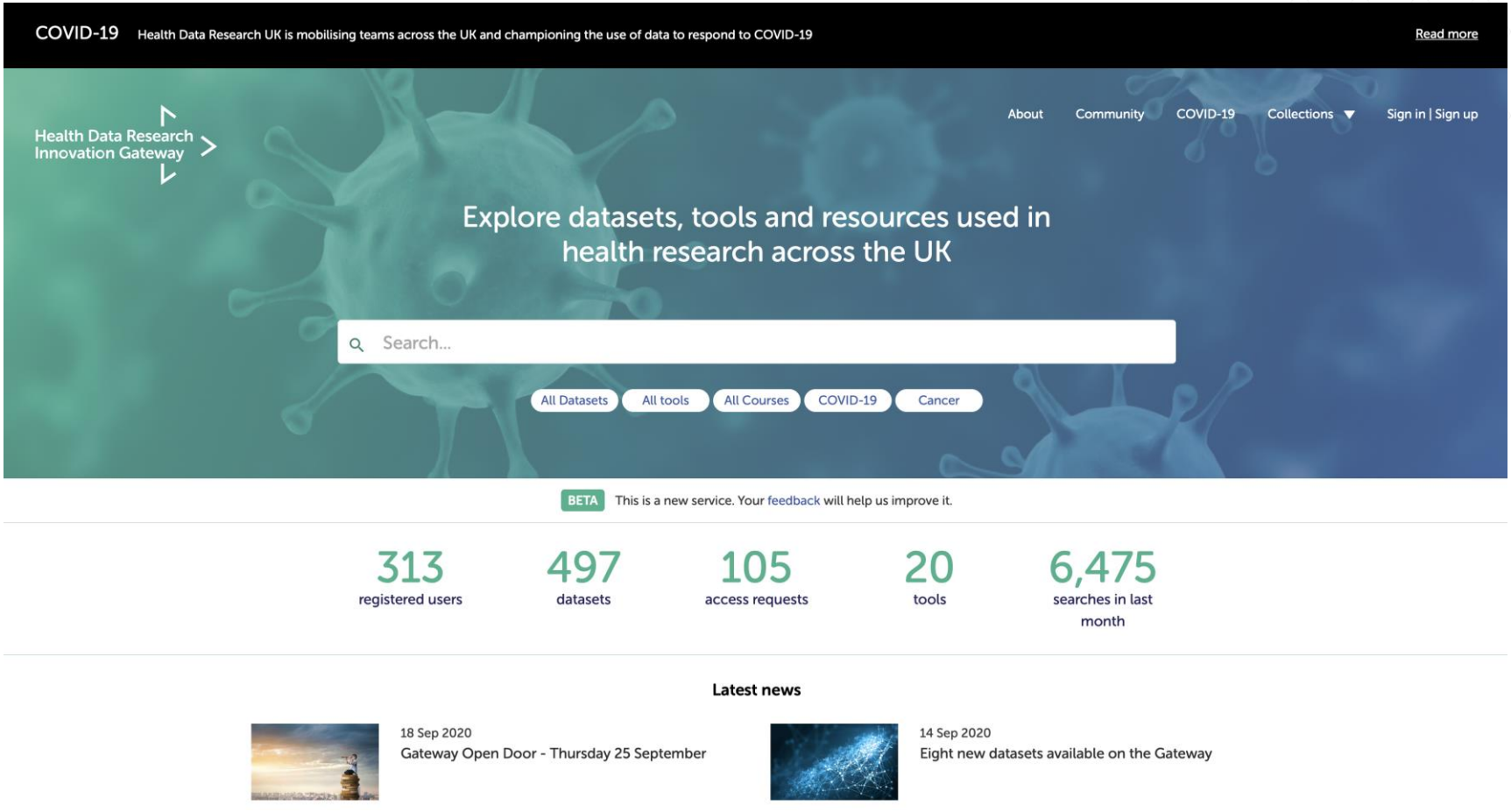


What is the UK Health Data Research Alliance?

By combining expertise and a shared commitment to work collaboratively, the Alliance helps researchers to answer some of the most difficult questions and address the most important health challenges faced in the UK. The Alliance is convened by Health Data Research UK.

https://ukhealthdata.org

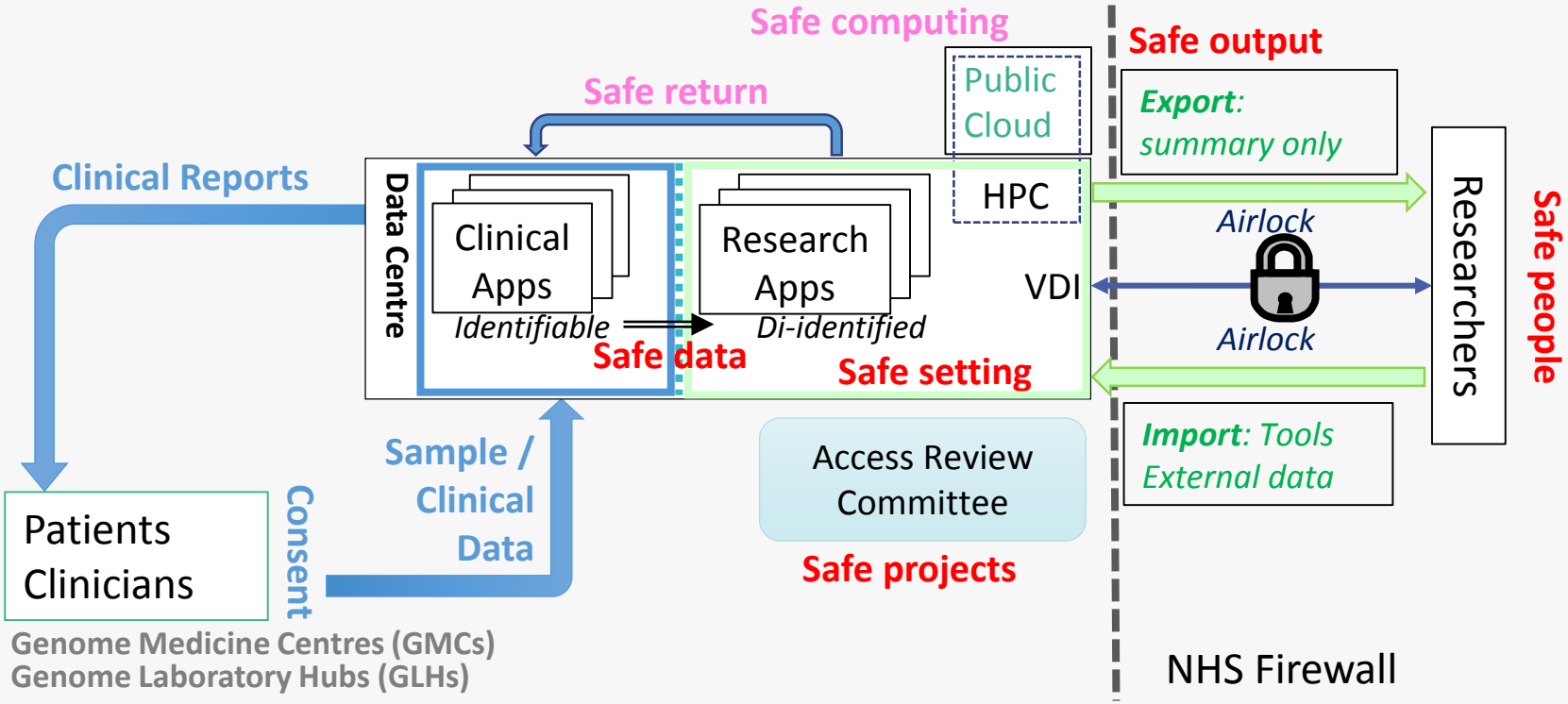
Heath Data Research Innovation Gateway



https://www.healthdatagateway.org

Research Environments for 100,000 genomes project
as a Trusted Research Environment

HDRUK
Health Data Research UK



News story

Next steps on the journey to Trusted Research Environments

31 July 2020

Project
Aligning approach to Trusted Research Environments

The UK Health Data Research Alliance Board agreed to move beyond the 'Green Paper' phase of the Trusted Research Environment (TRE) programme at its meeting on 23 July 2020.

We are now pleased to publish our updated paper which is the result of a collaborative community effort. There was agreement in the concept and an aspiration to work towards TREs in the future, but also understandable concerns regarding how quickly TREs can meet different researchers' needs.

We are now entering the more detailed design and delivery work set out in the next steps of the paper. The six areas identified for further work cover:

https://ukhealthdata.org/news/next-steps-on-the-journey-to-trusted-research-environments/

Summary

- 100,000 Genomes Project completed
- 1 million whole genomes in 5 years
  - 500,000 NHS Genome Medicine Service
  - 500,000 UK Biobank Cohort
- GeCIP and Discovery Forum research activity

- COVID response
  - 35,000 whole genomes of affected individuals
  - COVID Research Environment, leveraging public cloud

- Health Data Research UK
  - UK HDR Alliance & Innovation gateway
  - Genomics England Research Environment template for Trusted Research Environments

Acknowledgements

- 100,000 genomes project patients and their families, Genomics England Team, NHS England, Health Education England, Public Health England
- Genomic Medicine Centres in England, Scotland, Wales and Northern Ireland, UK CLL Consortium, CRUK, RCPath, NHSE, DoH, Biobank UK, NIHR BioResource Rare Disease, DDD, NIHR Translational Research Collaborative
- Health Data Research UK; Imperial College, London School of Hygiene and Tropical Medicine, King's College London, Queen Mary's University London and University College London; OneLondon
- Wellcome Sanger Institute, European Bioinformatics Institute, King's College London.