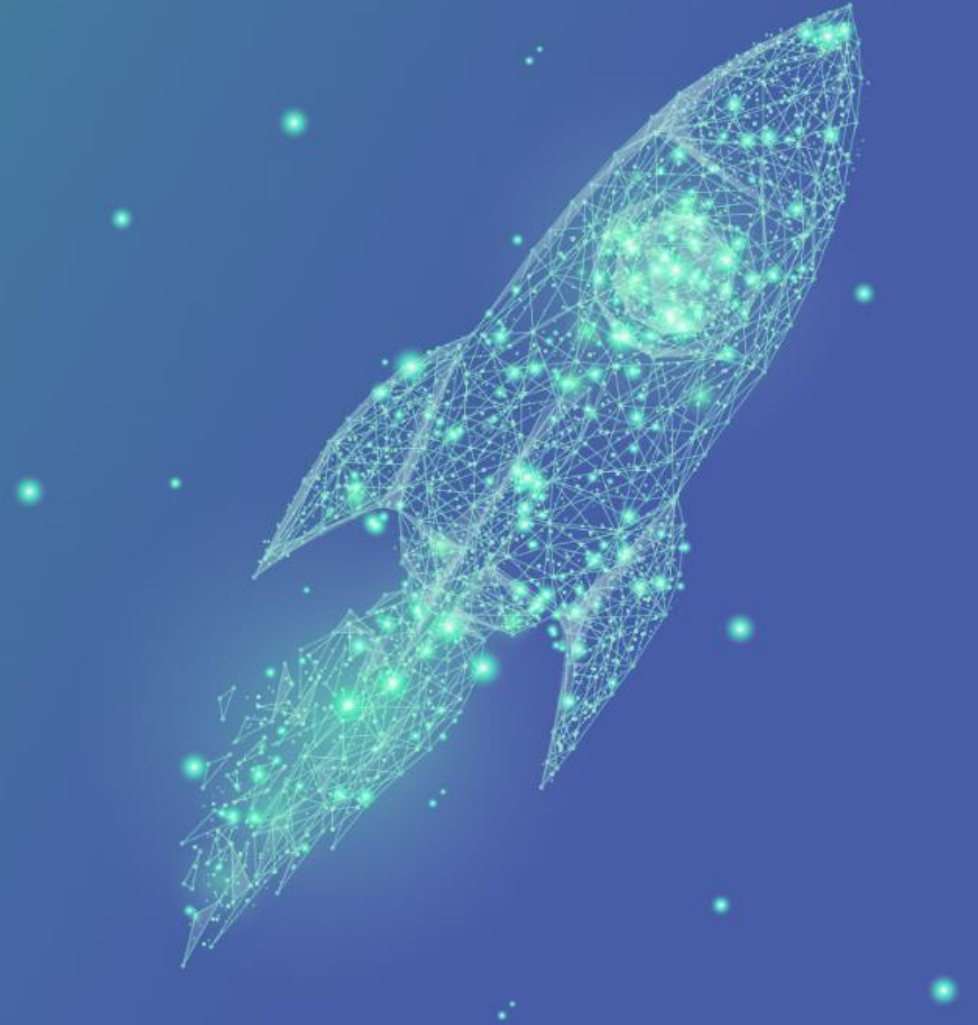




UK Health Data Research Alliance Symposium

4 February 2020

@HDR_UK | #hdralliance





Welcome!

The Inaugural UK Health Data Research Alliance Symposium

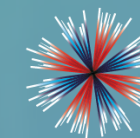
Andrew Morris, Director, Health Data Research UK

@HDR_UK | #hdralliance



UK Health Data
Research Alliance

HDRUK
Health Data Research UK



**INDUSTRIAL
STRATEGY**

UK Research
and Innovation

**Our mission is to unite the UK's health data to enable
discoveries that improve people's lives**

It's a complex environment

Establishing processes for large-scale health data research at scale



Interoperability: to work across systems with no additional effort

An alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale.

Knotty Issues

- Information governance data access*
- Data quality*
- Sustainability and commercial models*
- Meaningful public engagement*
- Technology strategy and standards*
- Inter-operability*

Design of the Alliance

Inspired by global collaborations

- No single organisation possesses all the information required to address every health and well-being issue facing society.
- Timely access to data across organisations requires:
 - Expertise
 - Trusted governance
 - Interoperability



Inspired by...



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.



W3C[®]



“As a neutral UK cooperative, we will be defined by our community of data controllers across the UK”.



Why is this important?

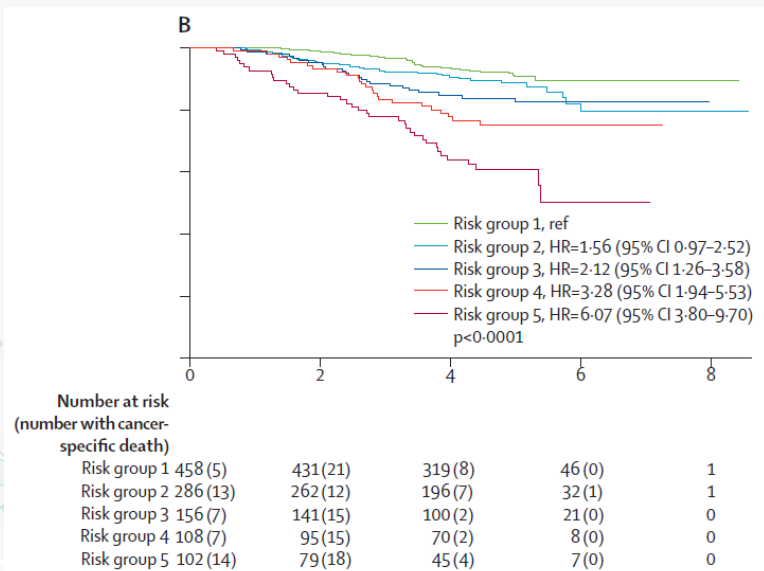
A Case Study



Deep learning for prediction of colorectal cancer outcome: a discovery and validation study

Ole-Johan Skrede*, Sepp De Raedt*, Andreas Kleppe, Tarjei S Hveem, Knut Liestøl, John Maddison, Hanne A Askautrud, Manohar Pradhan, John Arne Nesheim, Fritz Albrechtsen, Inger Nina Farstad, Enric Domingo, David N Church, Arild Nesbakken, Neil A Shepherd, Ian Tomlinson, Rachel Kerr, Marco Novelli, David J Kerr, Håvard E Danielsen

- 12 000 000 image tiles from patients with a distinctly good or poor disease outcome
- Four cohorts used to train a total of ten convolutional neural networks,



Editorial

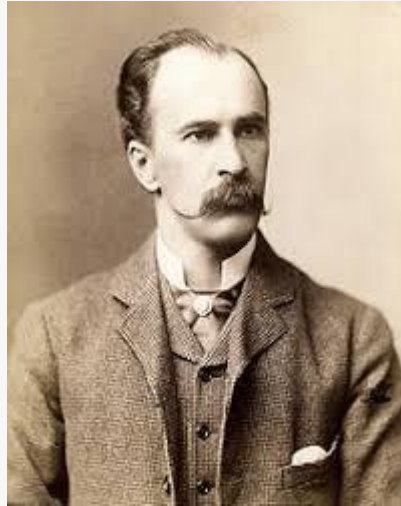
Leaving cancer diagnosis to the computers



- If AI systems are to be developed and used widely, the use of diverse population data will be critical.
- Sharing data between institutes and regions is becoming widely accepted as a necessity to AI research,
- Pledges from NIH, US National Cancer Institute, Wellcome, and the Bill & Melinda Gates Foundation, which will require data gathered from any funded project to be made available to the scientific community.

Aims of the Day

- Build the community engaged with the development and use of the UK health data research infrastructure
- Celebrate and showcase achievements to date across Alliance, Gateway and Hubs
- Gain inspiration from analogous developments and international perspectives
- Help to shape priorities and plans for the future



The best preparation for tomorrow
is to do today's work superbly well

William Osler

July 12th 1849-December 29th 1919

Thank you for listening!

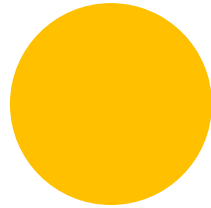
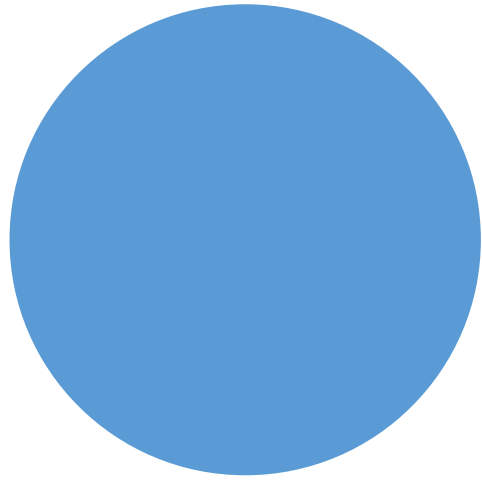


Engaging and involving patients and the public in uniting health data to enable discoveries that improve people's lives

Angela Coulter, Chair, Public Advisory Board, Health Data Research UK

@HDR_UK | #hdralliance





Engaging and Involving Patients and
Public in Uniting Health Data to Enable
Discoveries that Improve People's Lives

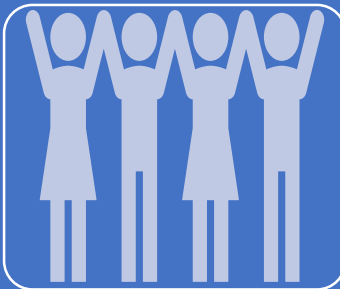
Angela Coulter
HDR UK Public Advisory
Board
@acpatient



Having a clear plan to
inform and involve people



Building confidence about
privacy and security



Demonstrating public
benefit

Core Principles for Engagement (DATA-CAN)

- We will ensure patients are included in our work
- We will ensure our engagement is representative of the diverse, UK wide population
- We will operate according to core principles of clarity, transparency and involvement
- We will act on what our community says is important
- We will enable and empower people to be involved, engage in and influence our work

Safe Projects,
Safe People,
Safe Data,
Safe Settings,
Safe Outputs





Using patient data is vital to improve health and care for everyone

There is huge potential to make better use of information from people's patient records. Data is vital for your individual care, and to improve health, care and services across the NHS. The information can be used to help understand more about disease, develop new treatments, monitor safety, plan NHS services and evaluate policies.

Examples:

Individual care

- Making sure your whole care team, from GP to hospital to care home, has up-to-date information to help give the best and safest care
- Helping people to manage their condition, from psychosis to asthma, using smart phone apps

Understanding disease

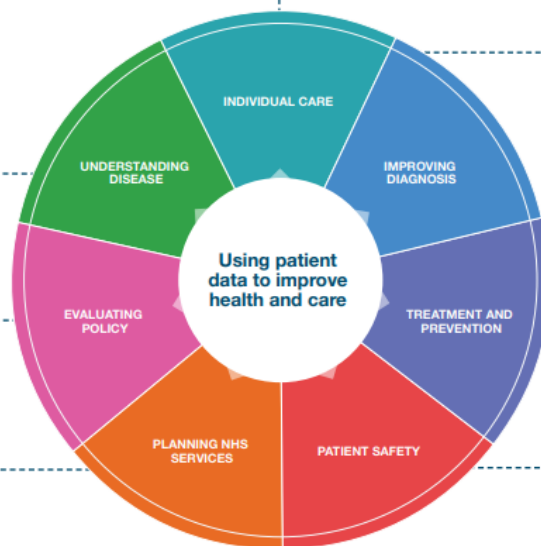
- Exploring why diabetes rates vary between ethnic groups
- Finding out about the biological changes involved in Parkinson's

Evaluating policy

- Comparing cancer survival rates between countries
- Showing the impact of a smoking ban on reducing premature births

Planning NHS services

- Finding out how many people have arthritis to inform planning
- Improving the delivery of kidney dialysis services



Diagnosis

- Demonstrating that bowel cancer screening is effective
- Finding new ways to identify early warning signs of dementia

Treatment and prevention

- Testing a treatment for lung disease in the real world
- Investigating the benefits of giving statins to men who haven't had a heart attack

Patient safety

- Monitoring the safety of whooping cough vaccination during pregnancy
- Checking the safety of hip replacements

Demonstrating Public Benefit

Understanding Patient Data



Overview: Health Data Research UK Infrastructure

Caroline Cake, COO & Deputy Director, Health Data Research

@HDR_UK | #hdralliance



Innovation Gateway

Gerry Reilly
Chief Technology Officer
Health Data Research UK

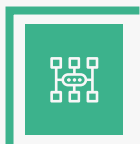


A Gateway is born (healthdatagateway.org)

The screenshot shows a web browser window with the URL <https://dev.healthdatagateway.org>. The page header includes "Innovation Gateway" and "Discover & Explore 414 datasets" with a "Login" link. The main content area features the "Health Data Research Innovation Gateway" logo, a large heading "Explore & Discover our 414 health datasets from across the UK.", and a sub-heading "Search and explore information about these datasets. If you have an OpenAthens account, you can log in to enquire about access to these datasets for research and innovation." Below this is a "BETA" notice: "This is a new service - your [feedback](#) will help us improve it." A search bar contains the text "bowel cancer". At the bottom, a section titled "What is the Innovation Gateway?" is visible.

The MVP continues to evolve

Since soft launch 16 January



Source available in GitHub
(@HDRUK/gateway-mvp)



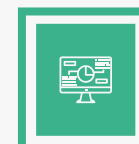
Institutional Login



Usage Dashboard



Status monitoring



Enhancements to
landing page



Enhancements to Dataset page

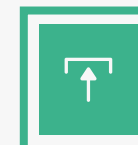
- Link to Metadata Catalogue



Improved mobile / tablet
rendering



Improvements to Dataset
Enquiry Form



Updates to privacy notice,
cookie handling and notice
and various bug fixes

and in the backlog



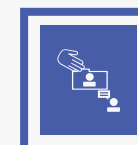
Social Login
(LinkedIn & Google)



Improved usability



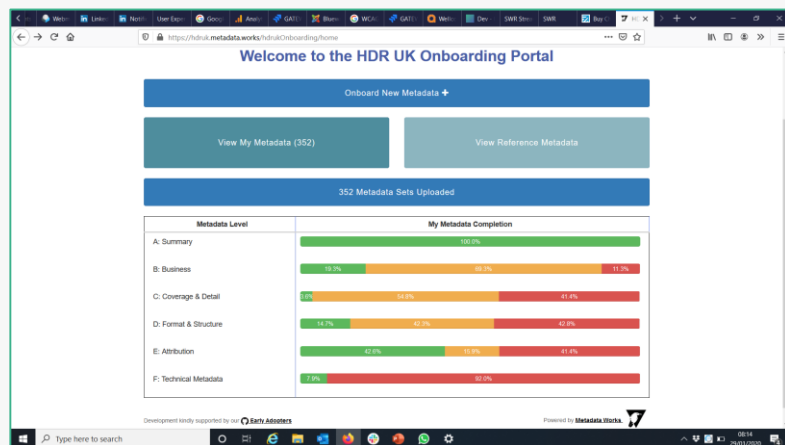
Search history



Multiple dataset
requests

We would also like you feedback to help us continue to improve the MVP email: support@healthdatagateway.org

Improving The Metadata



The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) primary care sentinel network and database

[Make enquiry](#)

DATE RELEASED: 11/12/2019 | PUBLISHER: RCGP and University of Surrey (to be superseded by University of Oxford) | LICENSE: In Progress | REQUEST TIME: 2-3 months

STANDARD: RCGP RSC hold data in a granular, generalized data warehouse model (see section F Technical Metadata for details) and can provide data in formats which conform to standard data model formats as requested, subject to cost.

The RCGP RSC network database is one of the freshest data sources of primary care data. Data are refreshed twice weekly. Through our network of we can (1) Identify and recruit patients for trials/studies, (2) Collect specimen, (3) conduct questionnaires.

Data Access

DATA CONTROLLER: The Royal College of General Practitioners (RCGP) is a policy level Data Controller for RCGP RSC. Currently RCGP has a contract with University of Surrey to be the Data Controller holding these data, this is migrating to University of Oxford during 2020.

DATA PROCESSOR: Not Applicable

ACCESS RIGHTS: Please complete our data request form at: https://www.rcgp.org.uk/-/media/Files/CIRC/Research-and-Surveillance-Centre/RCGP_RSC_Data_Request_Form_2018_ash7a-en Our data can be used for SOUIRE purposes: Surveillance Quality Improvement Research Education There is no restriction on who can use our data, just the purpose. RCGP Study Approval Committee and possibly Ethical approval will also be required. General information about our network is at: <http://www.rcgp.org.uk/rsc>, <https://clinintf.eu/index.php/rcgp-rsc/>, <https://clinintf.eu/index.php/wlo/>

Coverage

JURISDICTION	GEOGRAPHIC COVERAGE	DATASET START DATE	DATASET END DATE	PERIODICITY
GB	England	1960-01-01	Not specified	Twice weekly

Demographics

STATISTICAL POPULATION	AGE BAND
4000000	Any Age

Related Resources

PHYSICAL SAMPLE AVAILABILITY: Access is available to influenza virology samples. Public Health England (PHE) genotype all the Influenza viruses they isolate. These genotypes are available from PHE. Over the next 30 months they will be linked to RCGP RSC data through a collaborative grant from the Wellcome Trust. PHE also holds physical samples for 5 years of virology samples. Access may be available on request.

Related Datasets

GROUP: Not Applicable

LINKED DATASETS: Data is linked using an NHS Digital approved process. Linkage to: Hospital Episode Statistics (HES) Office for National Statistics (ONS) Cancer registry (CANREG) We can also link to virology samples and genetic data, subject to ethical approval.

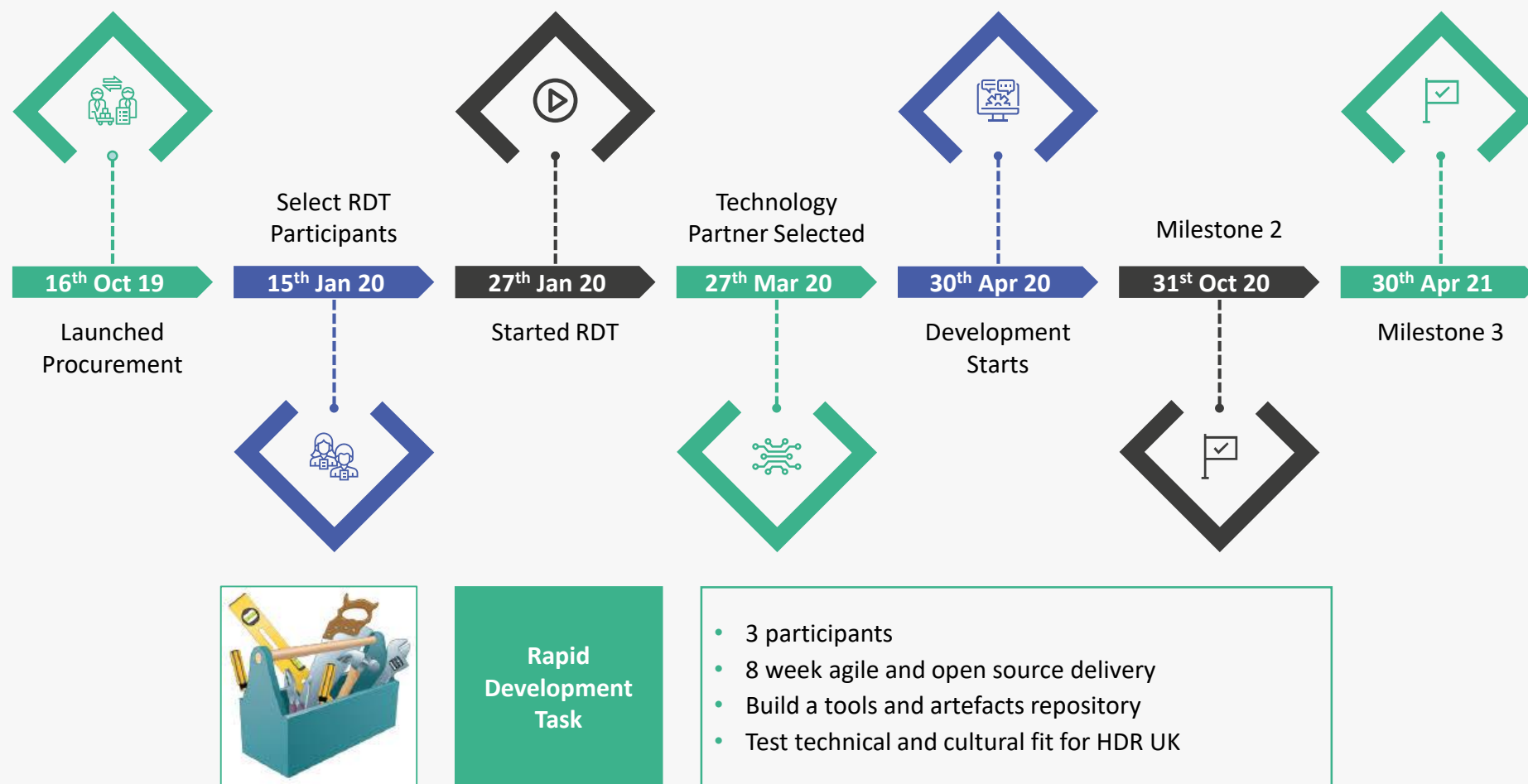
DERIVED DATASETS: The entire database is available to researchers, through the data access process (Pre-defined extracts are available at minimal cost (based on the principle that extraction processes previously developed will be re-used, thus incurring minimal costs for future researchers). These pre-defined extracts cover: Charlson comorbidity score • Frailty Index • Extracts of our monitored conditions. • Risk scores including: CHA2DS2-VASc score

Attributions

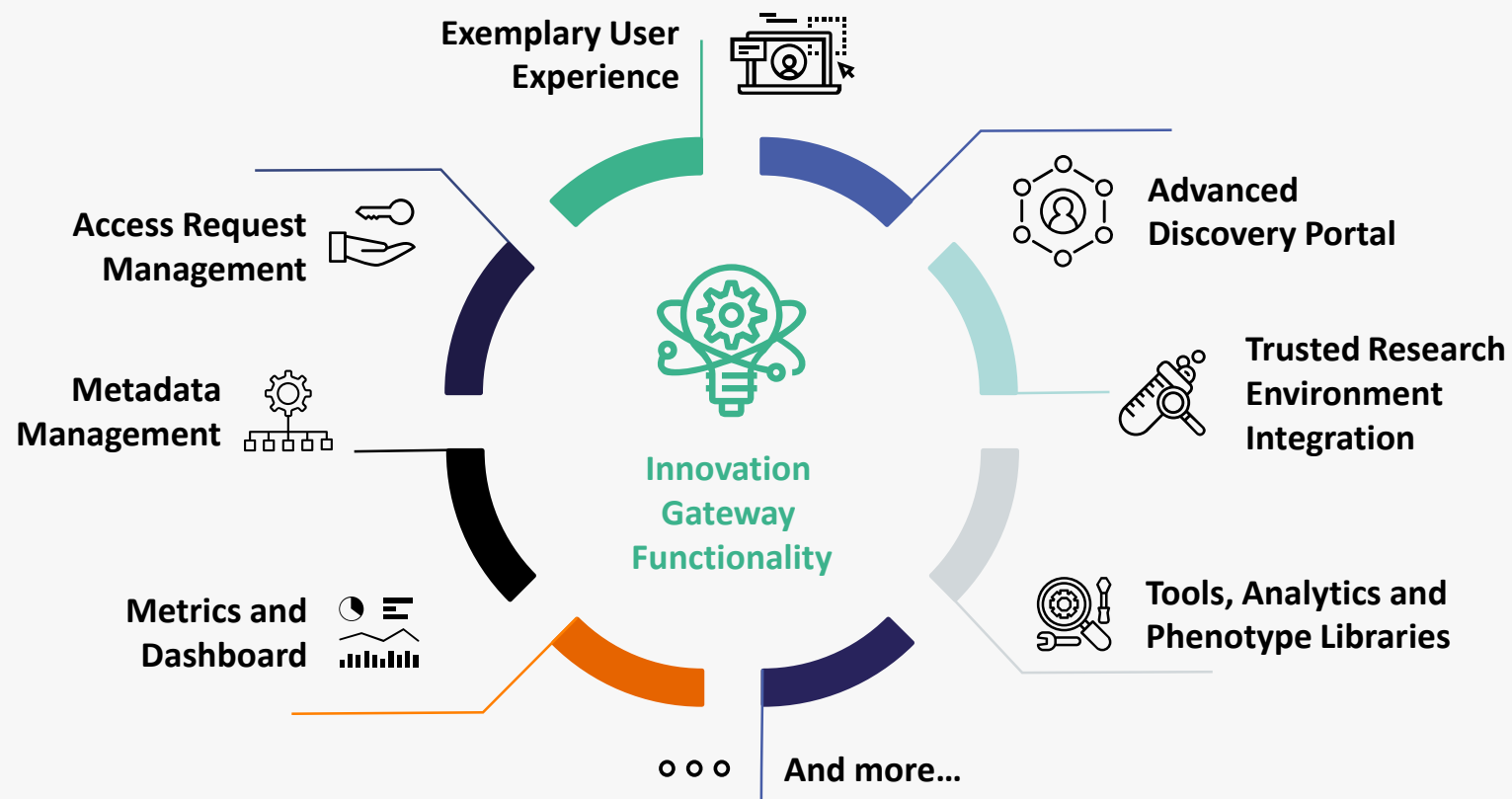
Next steps

- Improve metadata maturity across member organisations through continued support
- Iterate metadata specification and complete outstanding attributes
- Improve tooling and process automation
- Support members to create sustainable metadata pipeline
- Encourage members to publish an organisation metadata strategy

Innovation Gateway, building a Technology Partnership



Towards an End to End User Journey



“

Finally, please try the Gateway at
www.healthdatagateway.org

”

Thank you!

“

and provide feedback to
support@healthdatagateway.org

”

HDRUK

Health Data Research UK



THANK YOU



www.healthdatagateway.org

support@healthdatagateway.org





UK Health Data Research Alliance – Uniting the UK's health data

Jane Ingham, CEO, HQIP
Yvonne Silove, Associate Director, HQIP

David Seymour, Partnerships Director, Health Data Research UK

@HDR_UK | #hdralliance



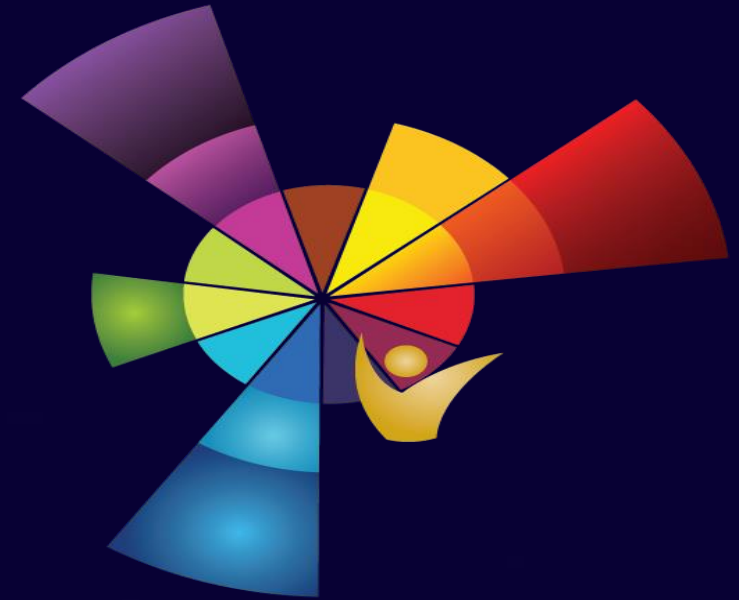
HQIP as a member of the Alliance - the journey so far

Jane Ingham

CEO

Yvonne Silove

Associate Director



Who is Healthcare Quality Improvement Partnership?



Our vision: enabling those who commission, deliver and receive healthcare to measure and improve services

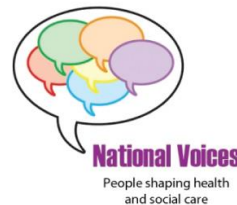


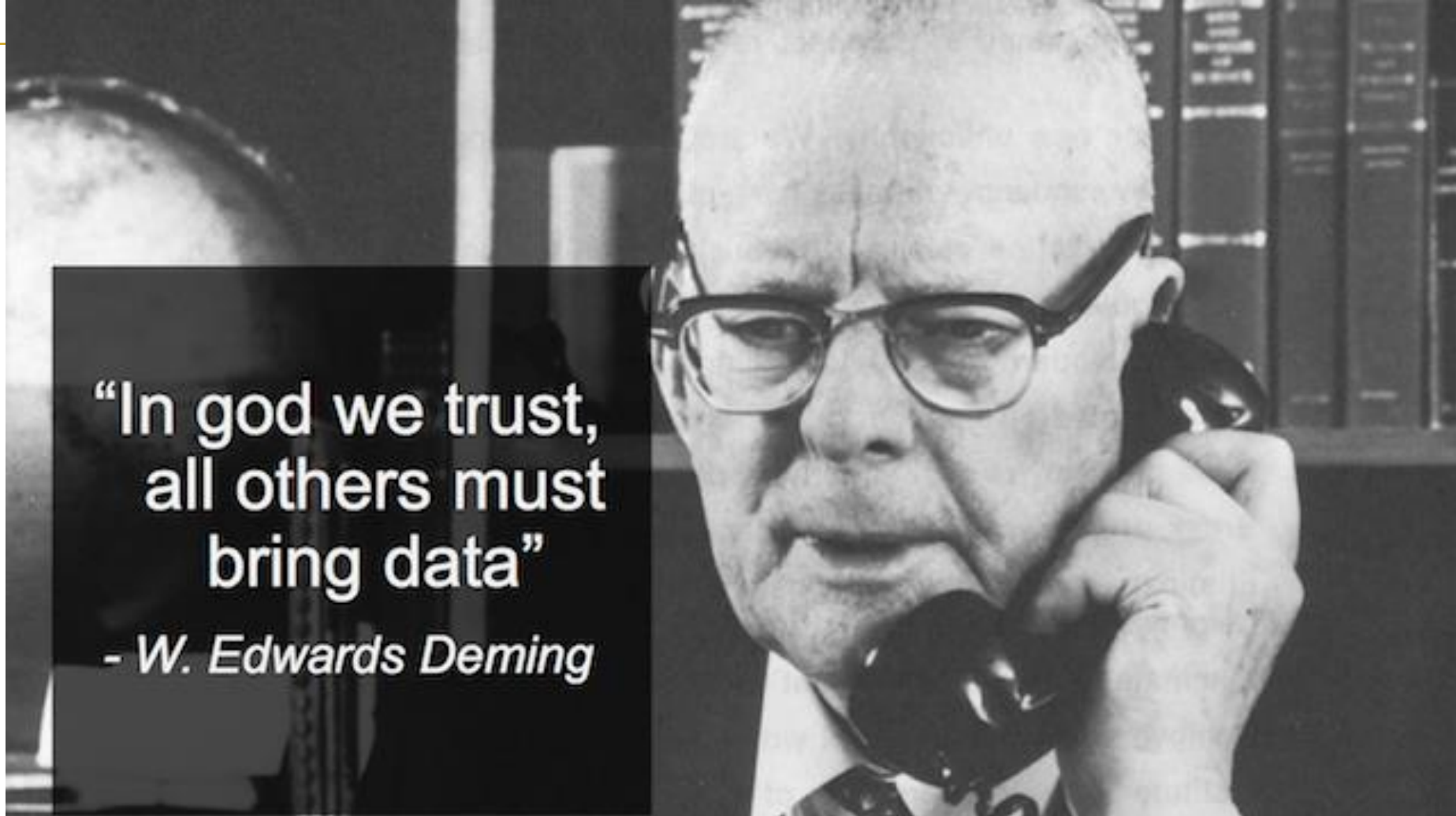
Our values: independent, working in partnership with patients and health professionals to improve practice



Our history: established in 2008, governed by the AoMRC, National Voices and RCN

ACADEMY OF
MEDICAL ROYAL
COLLEGES





“In god we trust,
all others must
bring data”

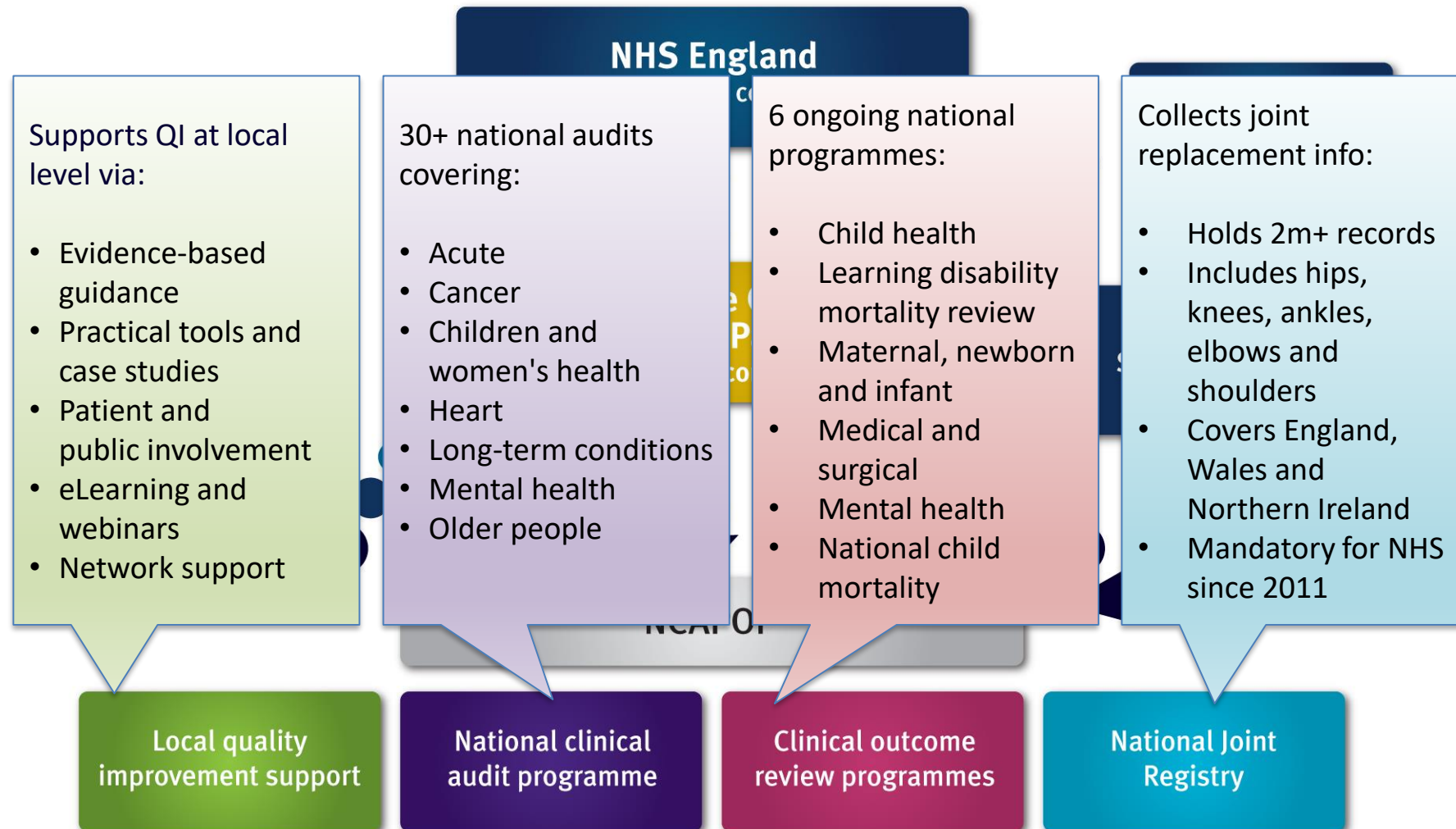
- *W. Edwards Deming*



HQIP

Healthcare Quality
Improvement Partnership

A description of HQIP work programme





“You cannot fatten
a cow by weighing it”
- Palestinian Proverb

Use of data for improvement

Manchester Royal Infirmary
 Central Manchester University Hospitals NHS Foundation Trust
National Audit of Dementia



Metric	CQC Key Question	2017 Report ¹	National Aggregate (England)	National Standard	Comparison to other sites	
20 carers	Percentage of carers rating overall care received by the person cared for in hospital as Excellent or Very Good	Caring	80.00%	68.90%	N/A	
110 staff	Percentage of staff responding "always" or "most of the time" to the question "Is your ward/ service able to respond to the needs of people with dementia as they arise?"	Responsive	78.80%	77.70%	N/A	
55 casenotes	Mental state assessment carried out upon or during admission for recent changes or fluctuation in behaviour that may indicate the presence of delirium	Effective	20.00%	44.90%	N/A	
39 casenotes	Multi-disciplinary team involvement in discussion of discharge	Effective	71.80%	81.90%	N/A	



www.rcpsych.ac.uk/dementiareport2017

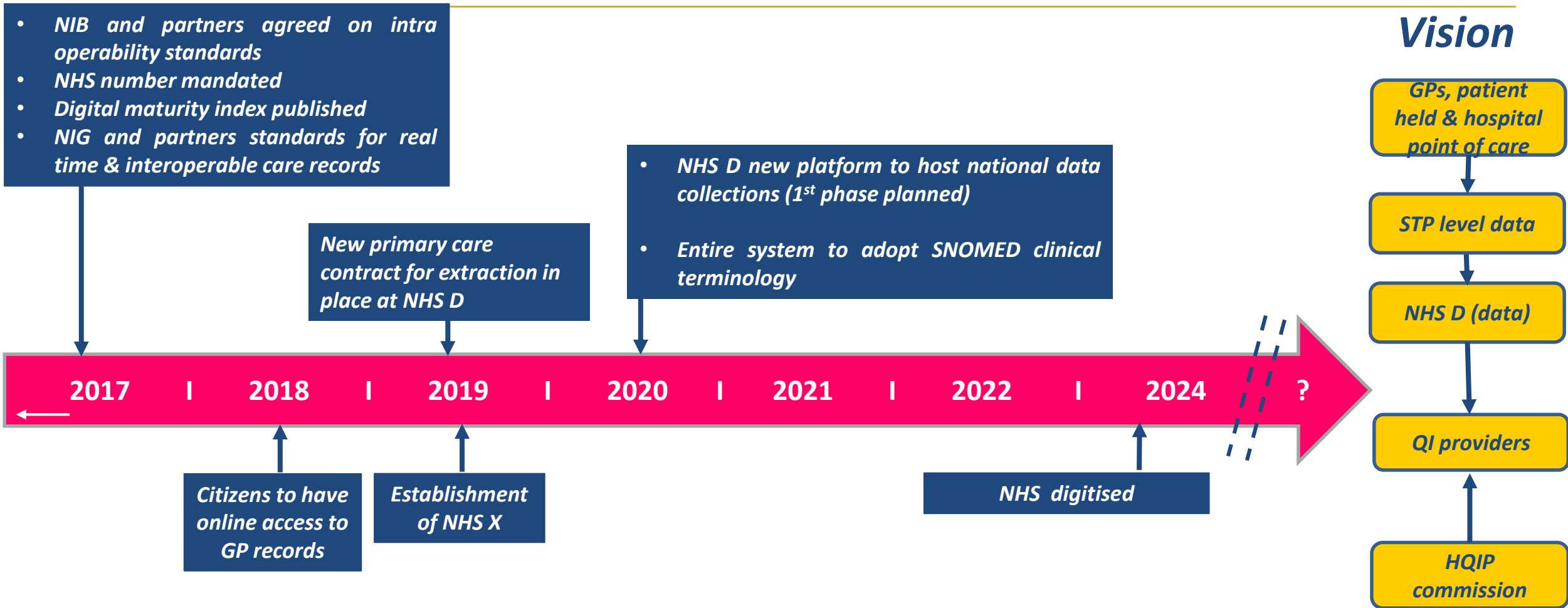
Produced by HQIP
 in partnership with the



KEY ONLY

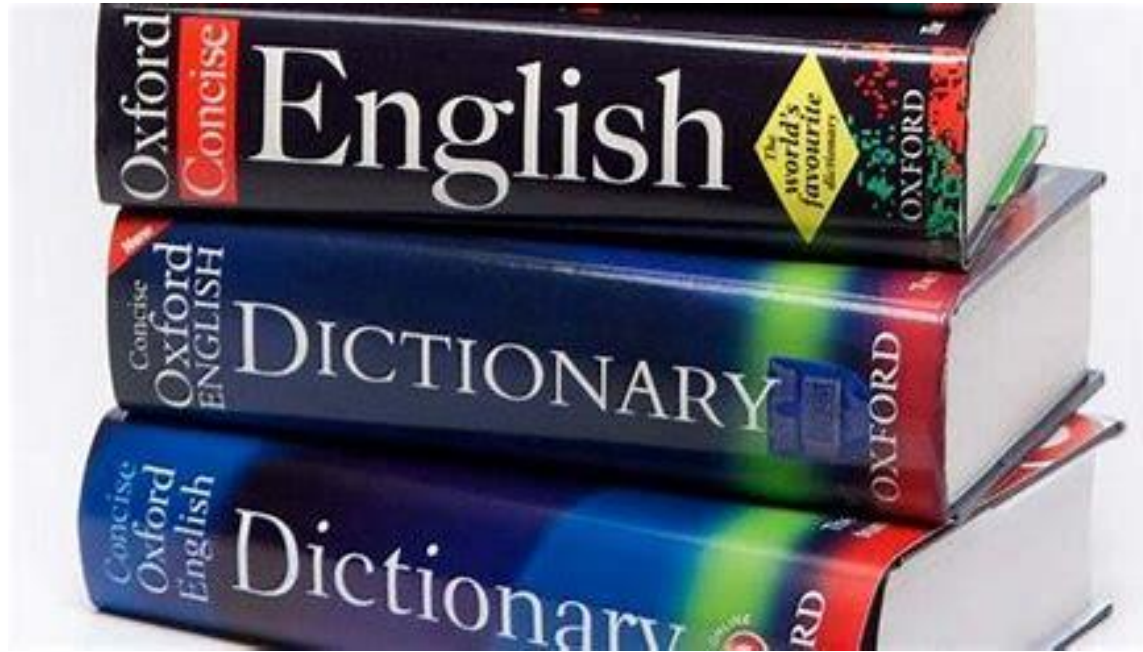


Information for QI in a digital NHS – steps to the *Vision*



Questions:

- Role and update of HES
- Patient objections - management of impact
- Influencing /positioning with EHR providers



Reflections on our journey so far...



Understanding the ambition, scope, design, ethos



Metadata Onboarding



Metadata Onboarding



Understanding Practice in Clinical Audit and Registries tool: UPCARE-tool

A protocol to describe the key features of clinical audits and registries

The Data Officers' Group



The Data Officers' Group



Clinical audit: New team registration



Complete this form to register a new team on the SSNAP webtool.

[Download Form](#)

SSNAP domains and key indicators

1. Scanning	
1.1	Proportion of patients scanned within 1 hour of onset
1.2	Proportion of patients scanned within 2 hours of onset
1.3	Proportion of patients scanned within 3 hours of onset
1.4	Proportion of patients scanned within 4 hours of onset
1.5	Proportion of patients scanned within 5 hours of onset
2. Stroke unit	
2.1	Proportion of patients directly admitted to a stroke unit within 8 hours of onset
2.2	Proportion of patients directly admitted to a stroke unit within 12 hours of onset
2.3	Proportion of patients directly admitted to a stroke unit within 16 hours of onset
2.4	Proportion of patients directly admitted to a stroke unit within 20 hours of onset
2.5	Proportion of patients directly admitted to a stroke unit within 24 hours of onset
3. Thrombolysis	
3.1	Proportion of eligible patients receiving intravenous thrombolysis
3.2	Proportion of eligible patients receiving intravenous thrombolysis within 1 hour of onset
3.3	Proportion of eligible patients receiving intravenous thrombolysis within 2 hours of onset
3.4	Proportion of eligible patients receiving intravenous thrombolysis within 3 hours of onset
3.5	Proportion of eligible patients receiving intravenous thrombolysis within 4 hours of onset
3.6	Proportion of eligible patients receiving intravenous thrombolysis within 5 hours of onset
3.7	Proportion of eligible patients receiving intravenous thrombolysis within 6 hours of onset
3.8	Proportion of eligible patients receiving intravenous thrombolysis within 7 hours of onset
3.9	Proportion of eligible patients receiving intravenous thrombolysis within 8 hours of onset
3.10	Proportion of eligible patients receiving intravenous thrombolysis within 9 hours of onset
3.11	Proportion of eligible patients receiving intravenous thrombolysis within 10 hours of onset
3.12	Proportion of eligible patients receiving intravenous thrombolysis within 11 hours of onset
3.13	Proportion of eligible patients receiving intravenous thrombolysis within 12 hours of onset
3.14	Proportion of eligible patients receiving intravenous thrombolysis within 13 hours of onset
3.15	Proportion of eligible patients receiving intravenous thrombolysis within 14 hours of onset
3.16	Proportion of eligible patients receiving intravenous thrombolysis within 15 hours of onset
3.17	Proportion of eligible patients receiving intravenous thrombolysis within 16 hours of onset
3.18	Proportion of eligible patients receiving intravenous thrombolysis within 17 hours of onset
3.19	Proportion of eligible patients receiving intravenous thrombolysis within 18 hours of onset
3.20	Proportion of eligible patients receiving intravenous thrombolysis within 19 hours of onset
3.21	Proportion of eligible patients receiving intravenous thrombolysis within 20 hours of onset
3.22	Proportion of eligible patients receiving intravenous thrombolysis within 21 hours of onset
3.23	Proportion of eligible patients receiving intravenous thrombolysis within 22 hours of onset
3.24	Proportion of eligible patients receiving intravenous thrombolysis within 23 hours of onset
3.25	Proportion of eligible patients receiving intravenous thrombolysis within 24 hours of onset

An easy to follow breakdown of SSNAP's 10 domains and 44 key indicators

[SSNAP domains and key indicators \(PDF\)](#)

SSNAP methodology



Overview on the methods of data collection, submission, and analysis on SSNAP

[Download SSNAP Methodology](#)

Current stroke datasets

The latest versions of the core and comprehensive stroke datasets

[SSNAP Core Dataset 4.0.0 \(Word\)](#)

[SSNAP Core Dataset 4.0.0 \(PDF\)](#)

[SSNAP Comprehensive Dataset](#)

How to register for the webtool



A step by step guide on how to register for the SSNAP webtool

[How to register for the SSNAP webtool](#)

Stroke dataset helpnotes



Helpnotes for users entering SSNAP data

[Helpnotes for core dataset 4.0.0](#)

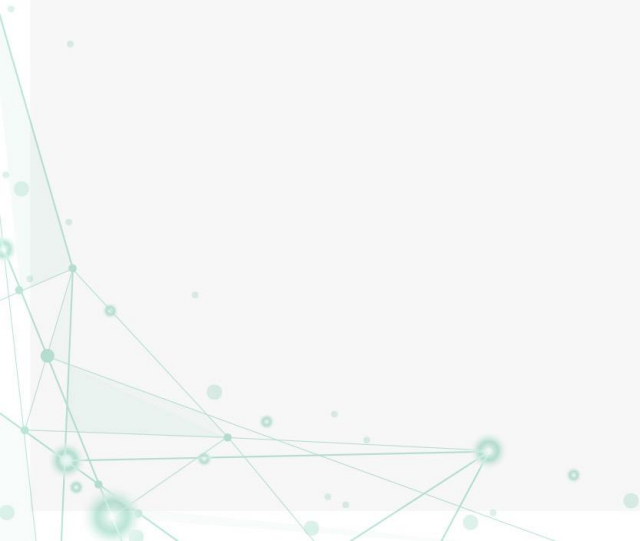


View to the future



Thank you

- Jane.Ingham@hqip.org.uk
- Yvonne.Silove@hqip.org.uk





UK Health Data Research Hubs – improving the UK's health data

Elizabeth Sapey, Director, PIONEER – The Health Data Research Hub for Acute Care

Ben Gordon, Digital Innovation Hub Programme Director, Health Data Research UK

@HDR_UK | #hdralliance






Health Data Research Infrastructure



Image Credit: NFL (Youtube)

Across the Hubs, 39 datasets are now discoverable on the Gateway

 Hub	 Datasets on Gateway	 Notes
Gut Reaction	7	Range of demographic and clinical data on IBD, relating to IBD registry
Breathe	7 (plus partial metadata from 10 more)	Includes cohorts, medical records and genome-wide association studies. All in SAIL: additional phenotype definitions, metadata and individual records available
Discover-Now	9	Multiple descriptions or components of single, linked dataset which includes community, primary care, prescribing and mental health data from North West London
Data-Can	3	Longitudinal real-world studies (Macmillan): Linked clinical and genomic (GeL); Clinical studies in paediatric cancer (Leeds and London)
Insight	2	Entire ophthalmic datasets from UHB and Moorfields — to form largest ophthalmic imaging dataset in the world
NHS DigiTrials	7	National datasets from NHS Digital to support feasibility service
Pioneer	4	Society of Acute Medicine database; acute inpatient episodes from UHB; out of hours primary care; longitudinal genomics data

All Hubs have developed tailored PPIE plans that have been reviewed by the HDR UK Public Advisory Board and each other

Each plan is tailored to the specific Hub
Common themes include:



Lay members or patient representatives involved in **all levels of governance**, from operational groups to steering boards



The use of **varied messaging and media** to communicate – including leaflets, focus groups, videos and workshops



Continued evaluation of the activities and impact



Exploration of **seldom heard groups**



Linking with **key partner organisations** to expand the reach of the Hub



Birmingham Health Partners

Collaboration. Innovation. Application.

On your marks, get set, go!
Hub development and delivery

Dr Elizabeth Sapey
Director of PIONEER



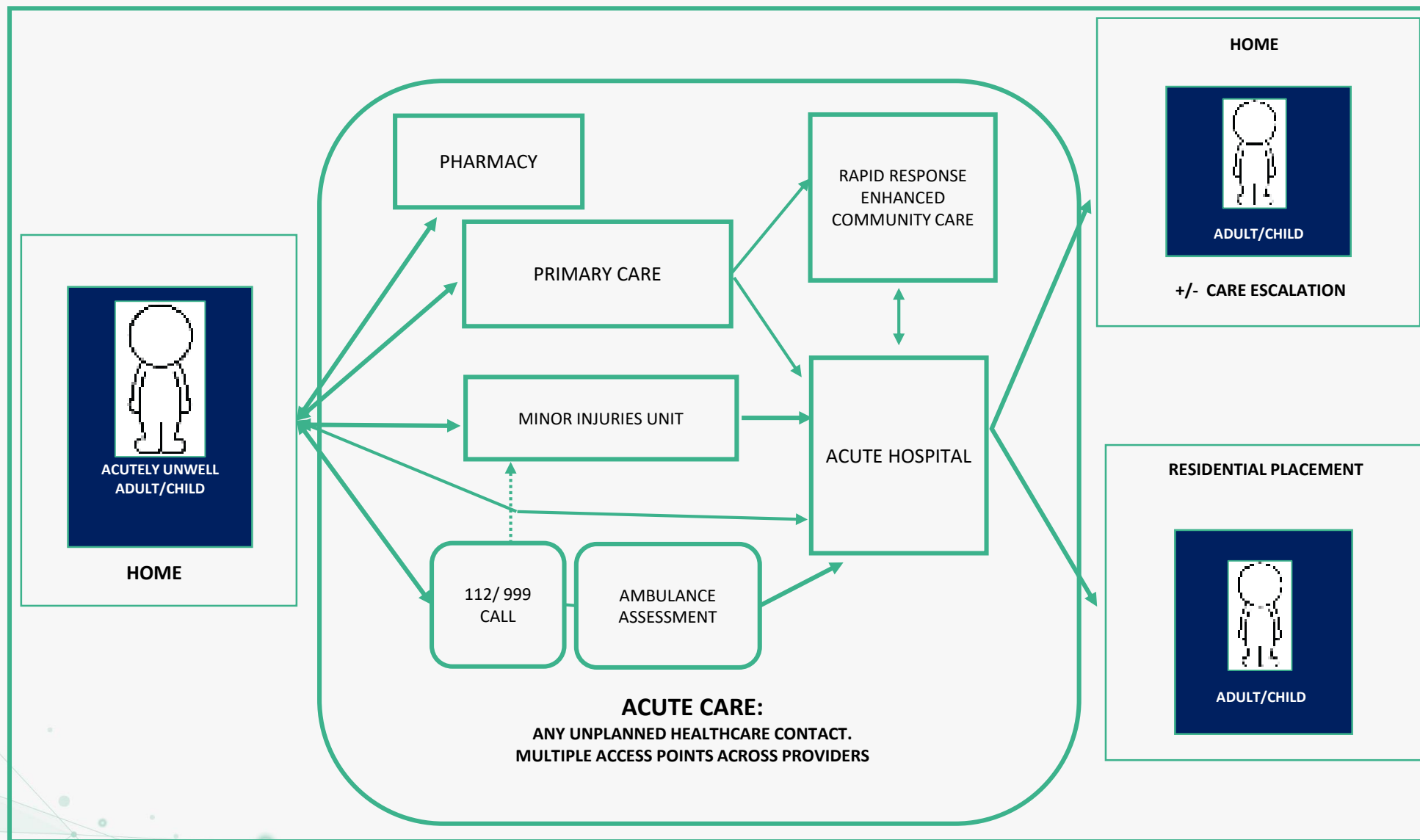
Delivered by

UNIVERSITY OF
BIRMINGHAM


University Hospitals
Birmingham
UH&B&CTU


Birmingham Women's
and Children's
UH&B&CTU

SILOED HEALTH CARE



Challenges

Patient and HCPs

- Lack of clarity of patient pathways
- Results not shared
- Lack of joined up thinking across providers
- Diagnostic delay
- Health inequalities
- Poor health outcomes

Researchers

- Unclear what data are there
- Unknown quality
- Poor accessibility
- Minimal linked data
- Slow to access
- Further curation needed

Our Aims across Hubs

- **Transparency in what data is available now**
- **New datasets generated**
- **Curated within hubs to improve quality**
- **Discoverable and accessible**
- **National governance policies implemented with equity**
- **Strongly directed by patient and public partners**
- **National framework to ensure benefit to NHS and UK population**

Our aim

- **Individually link data across acute health care providers**
 - **Primary care, Out of hours, Ambulance, Secondary Care, Community Care**
- **Start with an acute care contact, then curate data from onset to outcome**
- **Episode/ Disease agnostic and multi-morbid inclusive**
- **Ageless (including children and adolescents)**
- **De-identify for research and innovation**
- **Describe and model/ test points of change**
- **Patients at the heart of executive decision making**

Size and Scalability

- Integrated Regional Depth Applied to the National Environment



UHB	150,000 pa
BWCH	55,000 pa
Out of Hours GP	130,000 pa
SWBH	95,000 pa
BCHT	200,000 pa
WMAS (112/999)	1,440,000 pa

Largest Critical Care Unit in Europe
21,000 deliveries per year (> anywhere in EU)
Royal Centre for Defence Medicine



153 Acute Care Hospitals
(representing 64% of total acute care hospitals in UK)
and OHCA

Translation
and
scalability of
regional
innovation
into national
impact

Lessons learned. 1

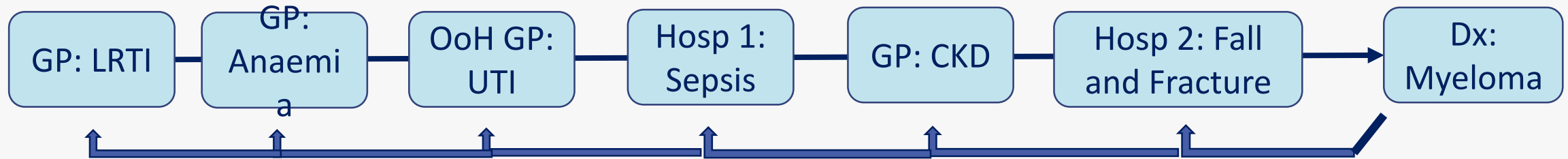
- **Public and patients are generous and altruistic in how their data is used**
 - **They are interested in this process and want this to work**
 - **But we must never take that for granted**
- **There is too much to do to and time is pressing**
- **Partnership working is vital**
 - **Public/ Patient/ Clinical/ Computer science/bio-informatics/ IG/ R&D/STP/ Academic/ Industry/ Policy**

Lessons learned. 2

- **NHS bodies are keen to share data**
 - **Recognise differences in digital maturity**
 - **Demonstrate patient benefit with the data shared**
 - **Reducing health inequalities**
- **Transparent governance framework and accountability**
- **Transparency in what we are doing**
 - **Licensing access for specific uses**
 - **Always for patient benefit**
- **Benefits are real and important**

Individual pathways explored in detail

Diagnostic delay of years



- New devices, software prompts, Healthcare AI to detect earlier patterns of disease
- Earlier diagnostic opportunities
- New populations for novel interventions

Real world data to stratify/personalise



- Real world diagnosis/ therapeutic pathways and real world patients
- Health care processes / Strategy and policy

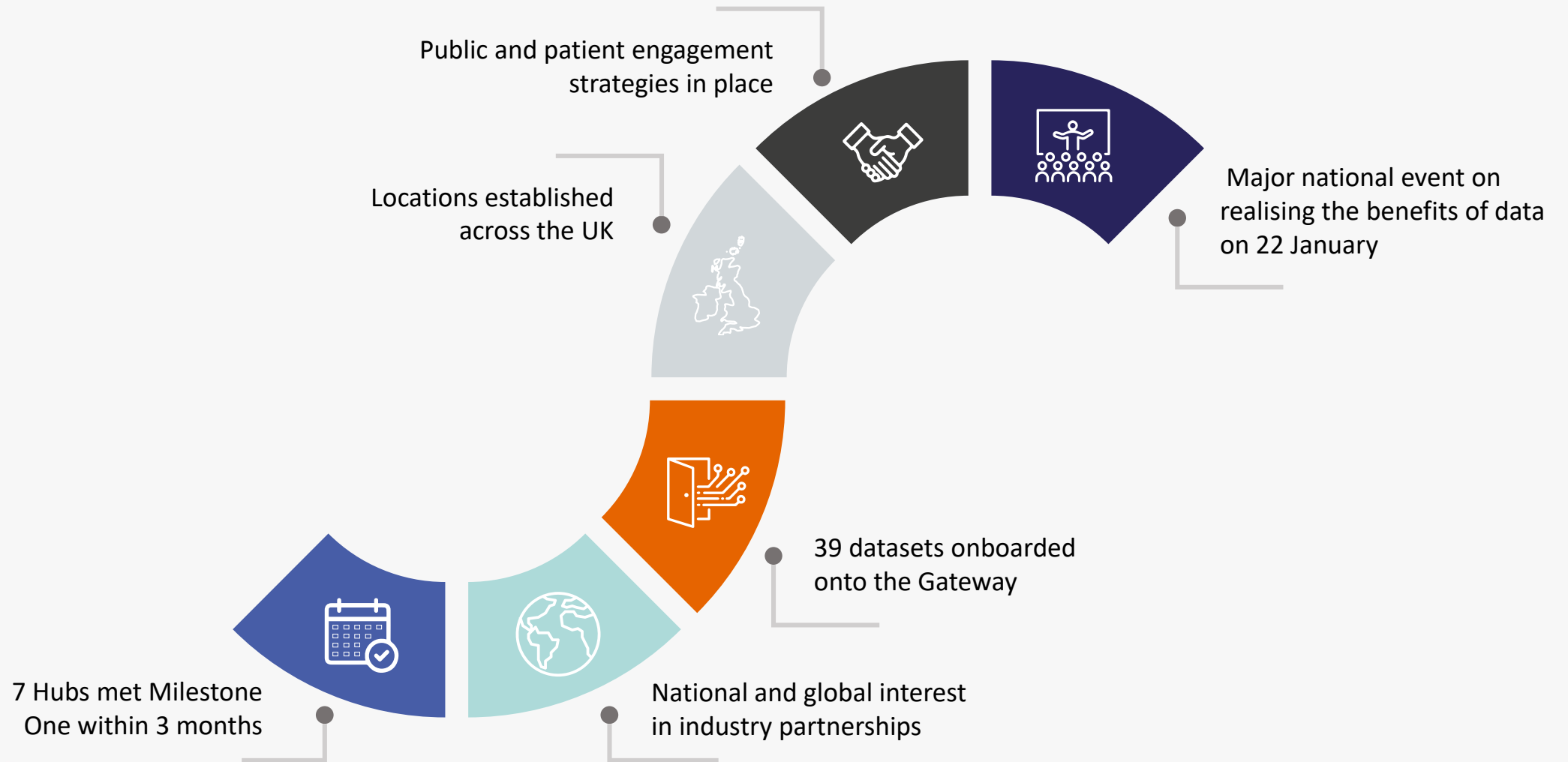
Using AI to optimize care



Next Steps

- **More and better**
- **Keep reinforcing the huge potential for benefit**
- **Be transparent in our processes**
- **Learn from each other**
- **Own our mistakes and learn from these**
- **Be guided by our patient and public voices**
- **Make our patients extremely proud of what their data has and will achieve for the UK and beyond**

Health Data Research Hubs



What comes next for hubs?

- **Milestone 2: Service Delivery – 31 March 2021**
- “Evidence that the **quality of the datasets has been improved** (curated) and that the curated data is discoverable through the Gateway. Provide publishable enhanced service **case studies** (e.g. from industry, academia and NHS) that demonstrate impact (and expected impact) and value to researchers and innovators from a range of sectors, and to patients and populations. Provide evidence that the Hub is **continuing to engage and involve patients and the public** in a meaningful manner.”

Area	Now	Going forwards
PPIE	Groups established PPIE plans in place	Learning Continuous improvement
Data	Discoverable Quality baseline	Increasing scale, coverage, depth, quality and linkage
Use cases	Work underway	Demonstrating impact Supporting NHS, academia and industry



Panel Q&A

Angela Coulter, Chair, Public Advisory Board, Health Data Research UK
Ben Gordon, Digital Innovation Hub Programme Director, Health Data Research UK

Elizabeth Sapey, Director, PIONEER – The Health Data Research Hub for Acute Care

Caroline Cake, COO & Deputy Director, Health Data Research UK

Jane Ingham, CEO, HQIP

Yvonne Silove, Associate Director, HQIP

Chair: David Seymour, Partnerships Director, Health Data Research UK

@HDR_UK | #hdralliance





Cross-cutting workstreams & breakout workshops

Refreshment break: 11:30-11:50

Cross-cutting workstream presentations: 11:50-12:15 (*livestreamed*)

Cross-cutting workstream breakout workshops: 12:15-13:00

Lunch: 13:10-14:00

@HDR_UK | #hdralliance



Workshop:
Earning, building and sustaining public trust

Amanda White, Communications & Engagement Direction, Health Data Research UK

11/02/2020





Aim: to agree practical steps and methodologies for how we earn trust and enable patients and the public to have confidence in our work

Expected outputs of workshop

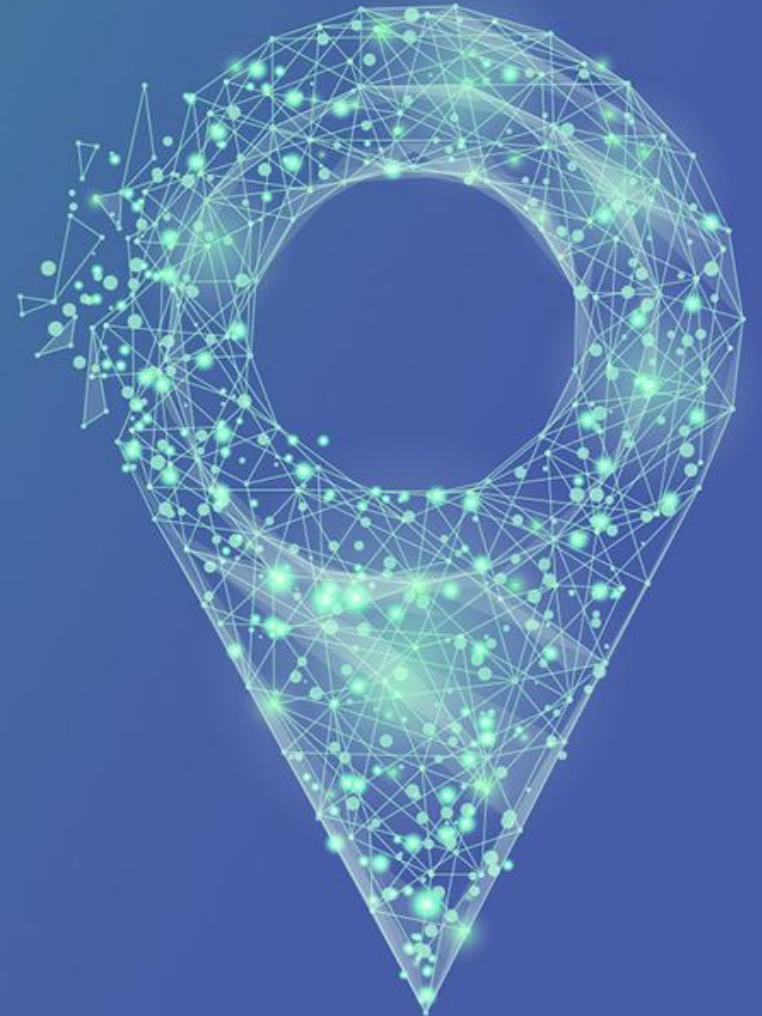
- Examples of ways that organisations are already earning public and patient trust
- Suggestions for reaching seldom heard groups
- Examples of existing resources, tools and methodologies



Workshop: Commercial sustainability

Joanne Hackett, Chief Commercial Officer, Genomics England

11/02/2020



Workshop: Ethics & information governance

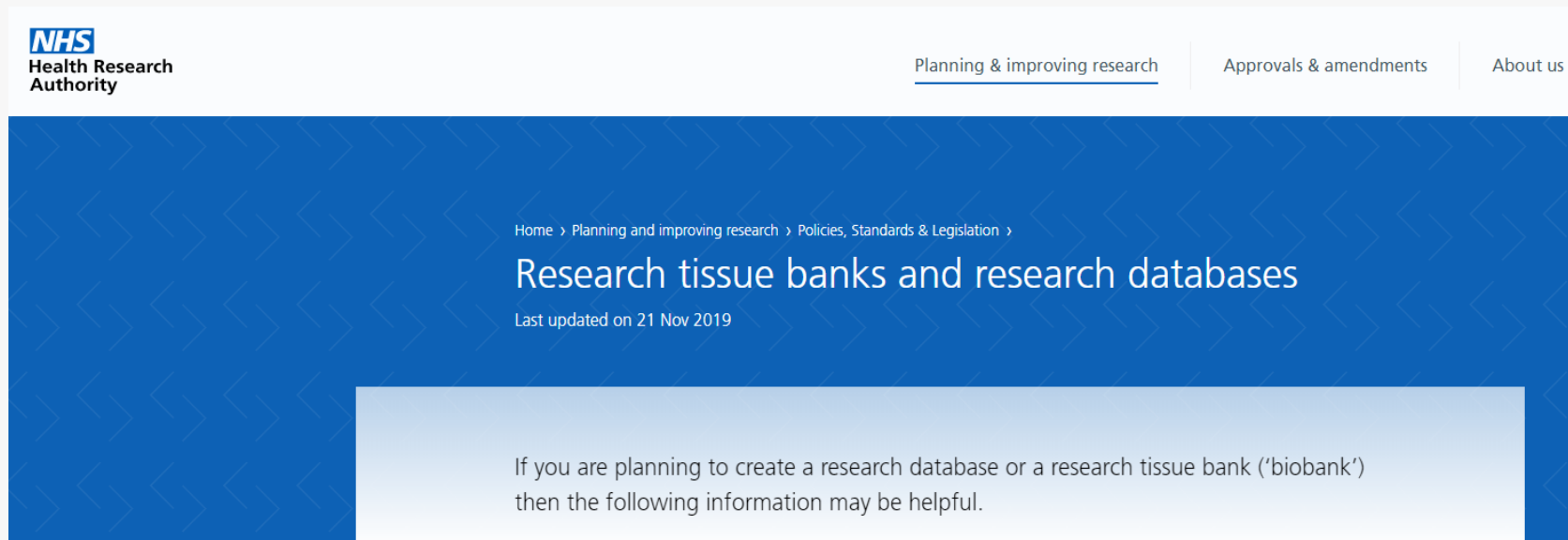
Clark Crawford, Ethics & IC Lead, INSIGHT – The Health Data Research Hub for Eye Health

11/02/2020



Ethical Frameworks

- Health Research Authority



The screenshot shows the NHS Health Research Authority website. The top left features the NHS Health Research Authority logo. The top right has navigation links: 'Planning & improving research' (underlined), 'Approvals & amendments', and 'About us'. The main content area has a blue background with a white breadcrumb trail: 'Home > Planning and improving research > Policies, Standards & Legislation >'. The title 'Research tissue banks and research databases' is prominently displayed in white, with 'Last updated on 21 Nov 2019' below it. A light blue box contains the text: 'If you are planning to create a research database or a research tissue bank ('biobank') then the following information may be helpful.'

- Entirely Voluntary
- **409 Databases are approved by an NHS REC**
- Akin to Tissue Directory and Coordination Centre : <https://biobankinguk.org/>

Legal Frameworks

The Common Law Duty of Confidentiality

Topics: [Good management](#), [good records](#), [Legal and professional obligations](#)

Common law is not written out in one document like an Act of Parliament. It is a form of law based on previous court cases decided by judges.

- Lessons Learned:
 - Identifying Legal Entities (Trusts, Cloud Service, etc)
 - Identifying Existing Data Controllers (e.g. Trusts)
 - Identifying Data Controller(s) for the Hub Activity
 - Defining activities under the control of the “Hub Data Controller(s)”
 - Identifying Legal Basis and exemptions
 - Assurances around Security
 - Transparency to data subjects

The screenshot shows the 'Data Protection Act 2018' page on the UK Public General Acts website. The page title is 'Data Protection Act 2018' and the breadcrumb trail is 'UK Public General Acts > 2018 c. 12 > Table of contents'. There are three tabs: 'Table of Contents', 'Content', and 'Explanatory Notes'. Below the tabs, there is a 'What Version' section with two radio buttons: 'Latest available (Revised)' and 'Original (As enacted)'. The 'Original (As enacted)' option is selected. To the right of the 'What Version' section, there is a 'Status' box that says 'Status: This is the original version'. Below the 'Status' box, there is a 'Collapse all -' button.

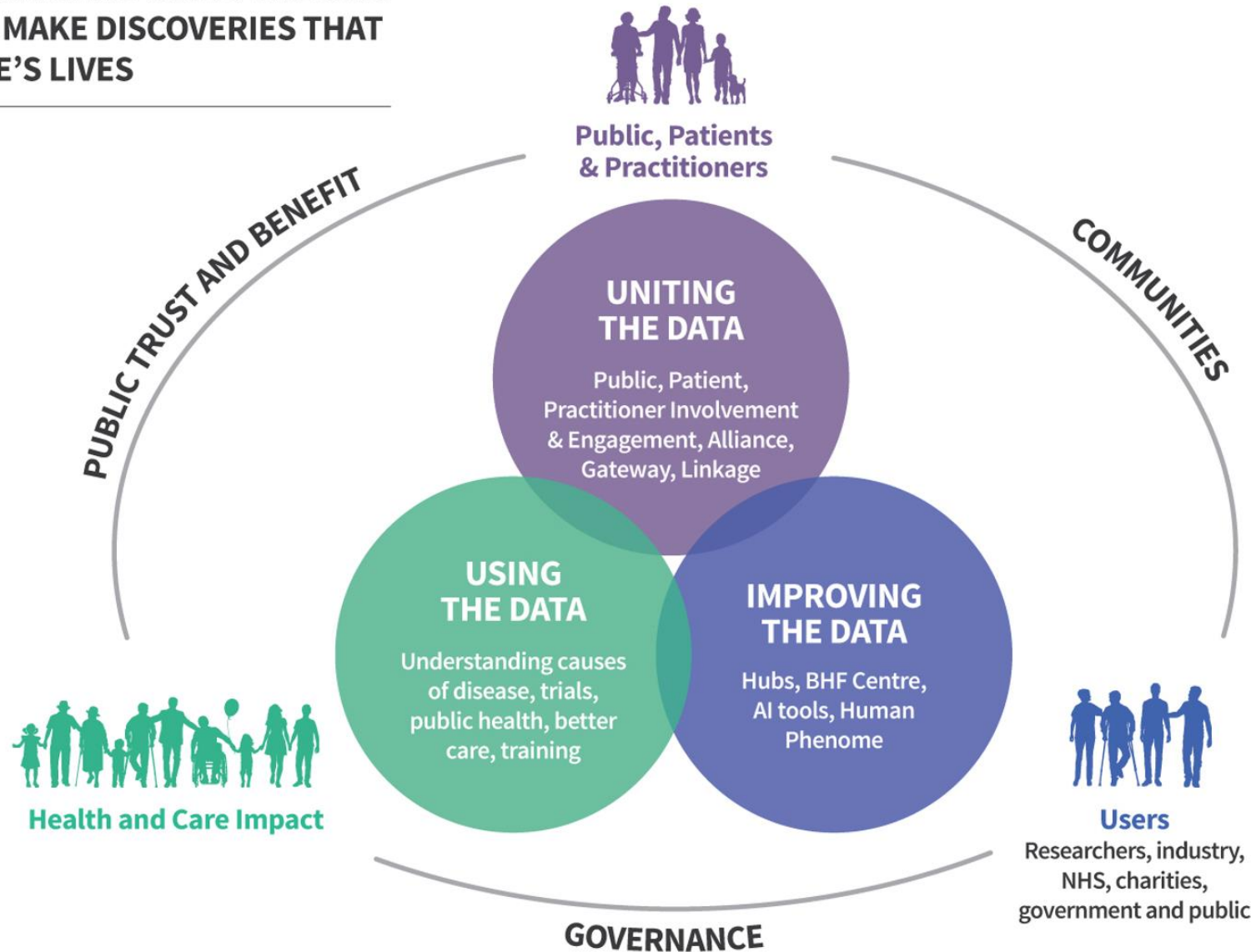
Workshop:
**Data quality & standards – improving the
data**

Neil Sebire, Chief Data Officer, Health Data Research UK
Monica Jones, Chief Data Officer, DATA-CAN – The Health Data Research
Hub for Cancer

11/02/2020



UNITING, IMPROVING AND USING THE UK'S HEALTH DATA TO MAKE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES



Health Data Standards and Interoperability issues

Define data elements

WHAT to collect

Define format of data

HOW to collect, store, represent, model, syntax, structure/Data model
(FHIR, OMOP, openEHR, proprietary etc)

Define terminologies / classifications / ontologies used

WHICH terms and definitions and their relationships
(SNOMED CT, LOINC, DM&D, ICD10 (diagnoses), OPCS4 (procedures), HPO)

Define semantics / phenomics

MEANING clinically of the terms and data values plus knowledge representation

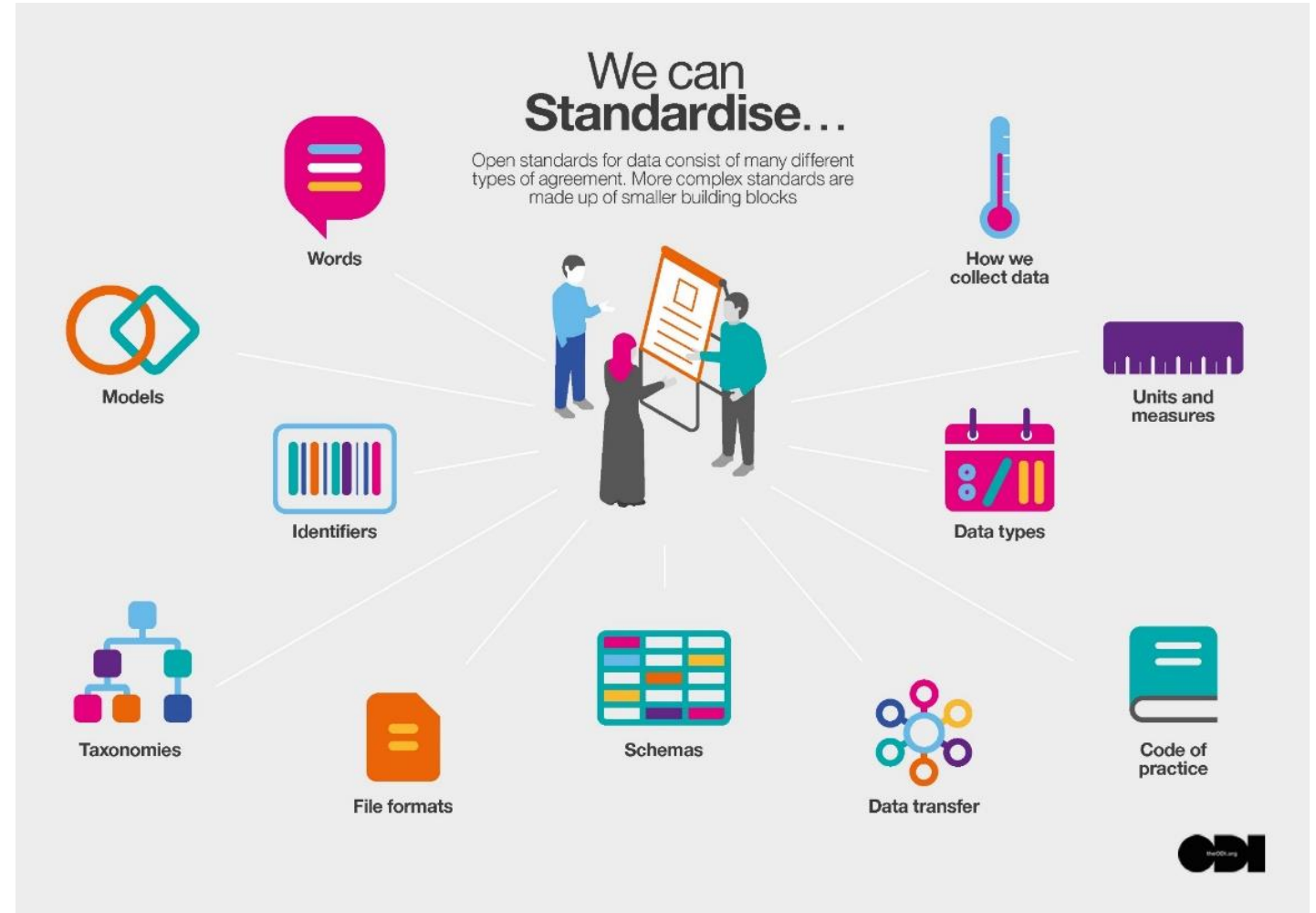
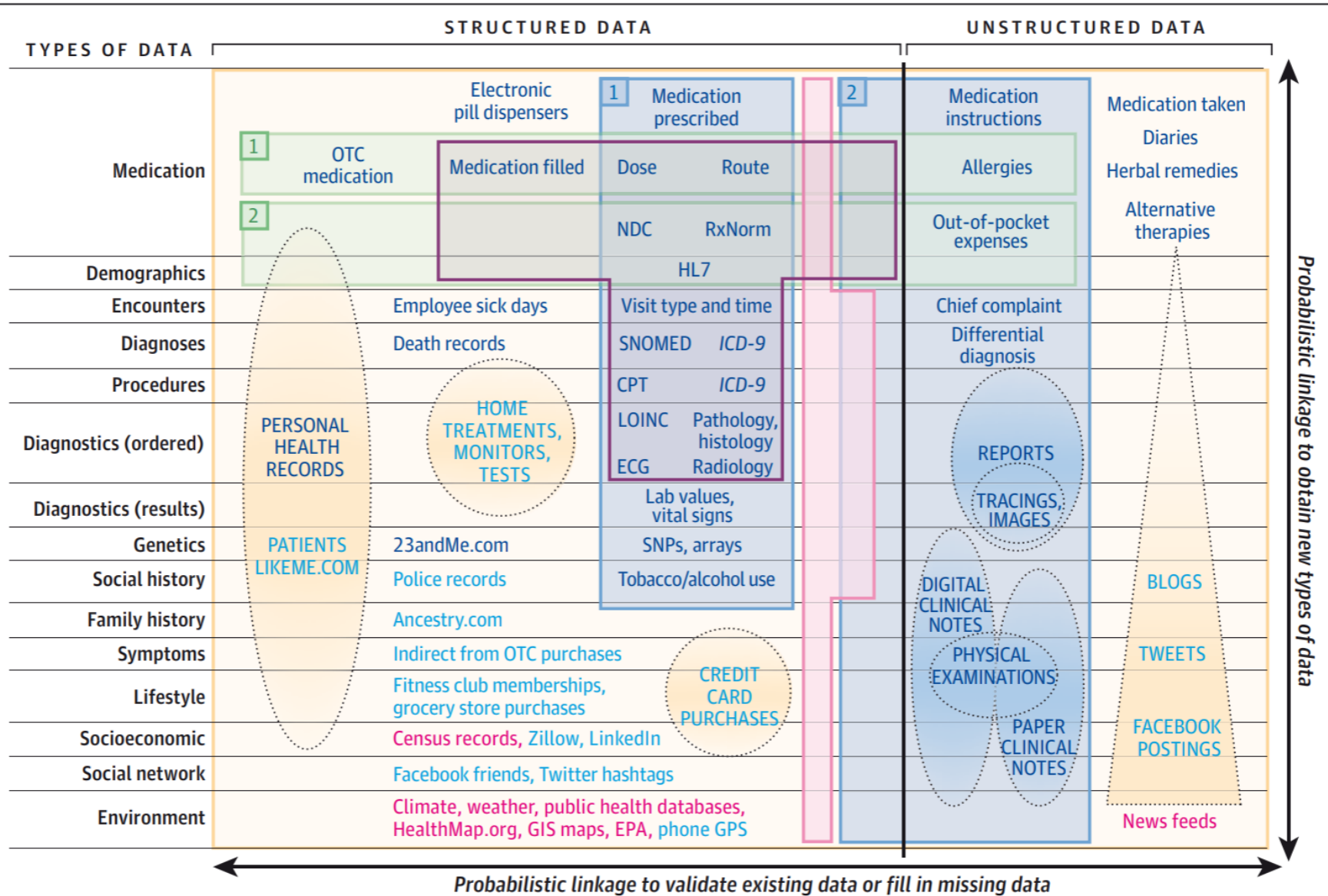







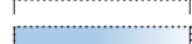




Figure. The Tapestry of Potentially High-Value Information Sources That May be Linked to an Individual for Use in Health Care



Probabilistic linkage to obtain new types of data

Probabilistic linkage to validate existing data or fill in missing data

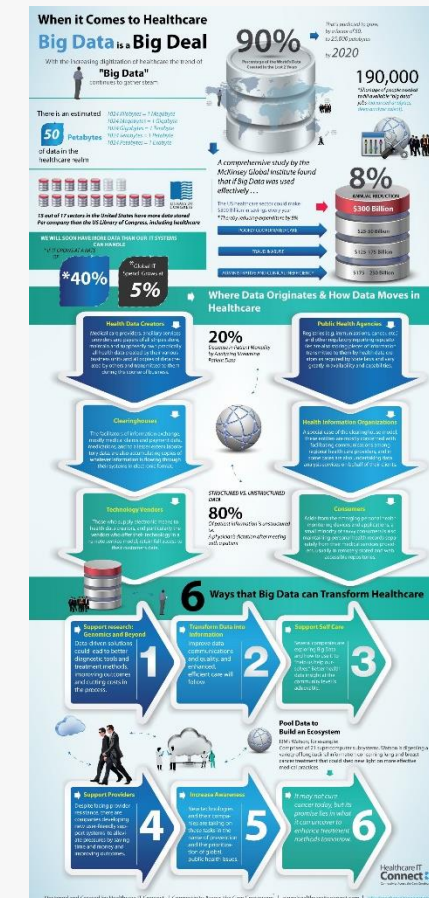
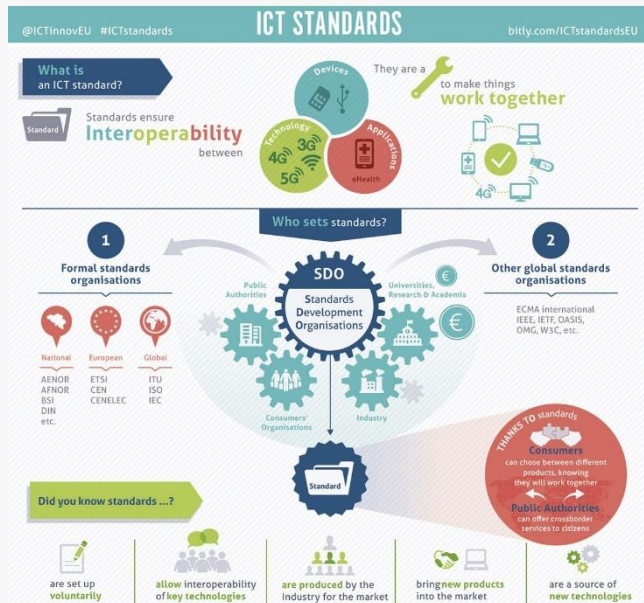
Examples of biomedical data		Ability to link data to an individual	Data quantity
 Pharmacy data	 Health care center (electronic health record) data	 Easier to link to individuals	 More
 Claims data	 Registry or clinical trial data	 Harder to link to individuals	 Less
 Data outside of health care system		 Only aggregate data exists	

Dataset Utility: Data standards, data quality, other factors

Which standards should HDRUK support

How should we measure data quality / utility?

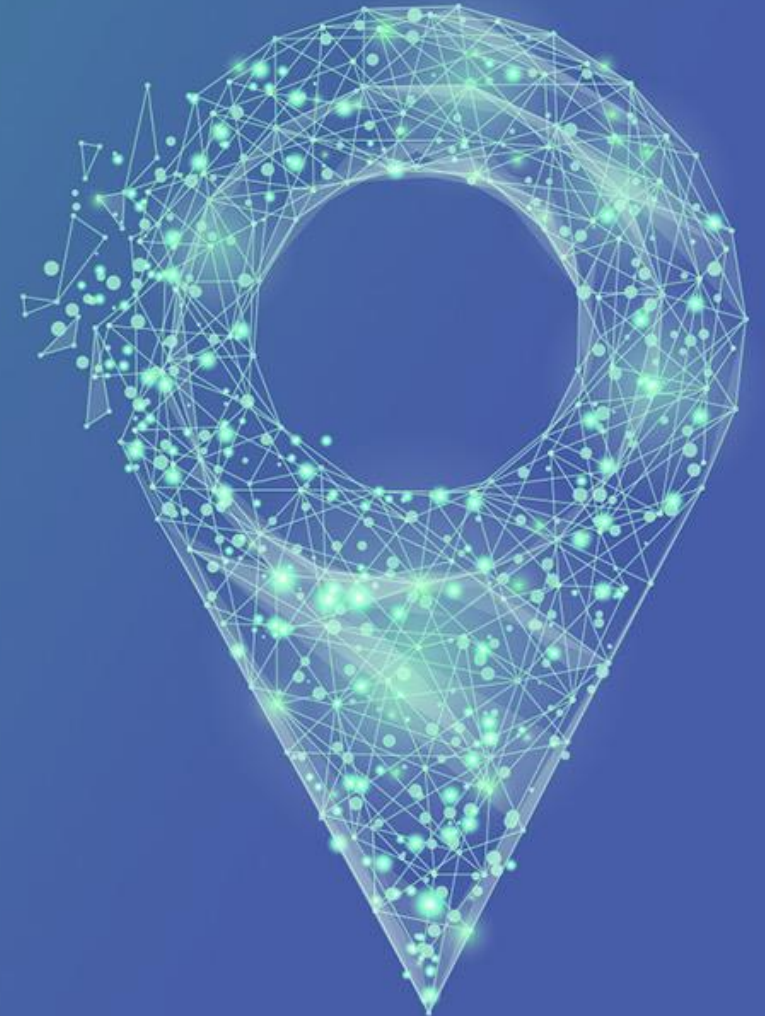
What level of 'curation' is reasonable / valuable



Workshop:
Trusted research environments

Tim Hubbard, Associate Director, HDR UK London

11/02/2020



Trusted Research Environments

- Proposal that UK HDR Alliance commits to “5 safes”.
- Central principle:
 - no distribution of individual level data;
 - all processing and analysis within Trusted Research Environment (TRE) (safe setting)
- Multiple examples of TREs operating successfully in this way
- Public is highly sensitized to issues around use of health data. Clear commitment to operating in this way provides a unique chance to “reset” public confidence.

Data reuse *via distribution*

- Specific request for access received, reviewed, approved
- Specific anonymised dataset prepared within safe haven
- User downloads dataset and carries out analysis on own computer system

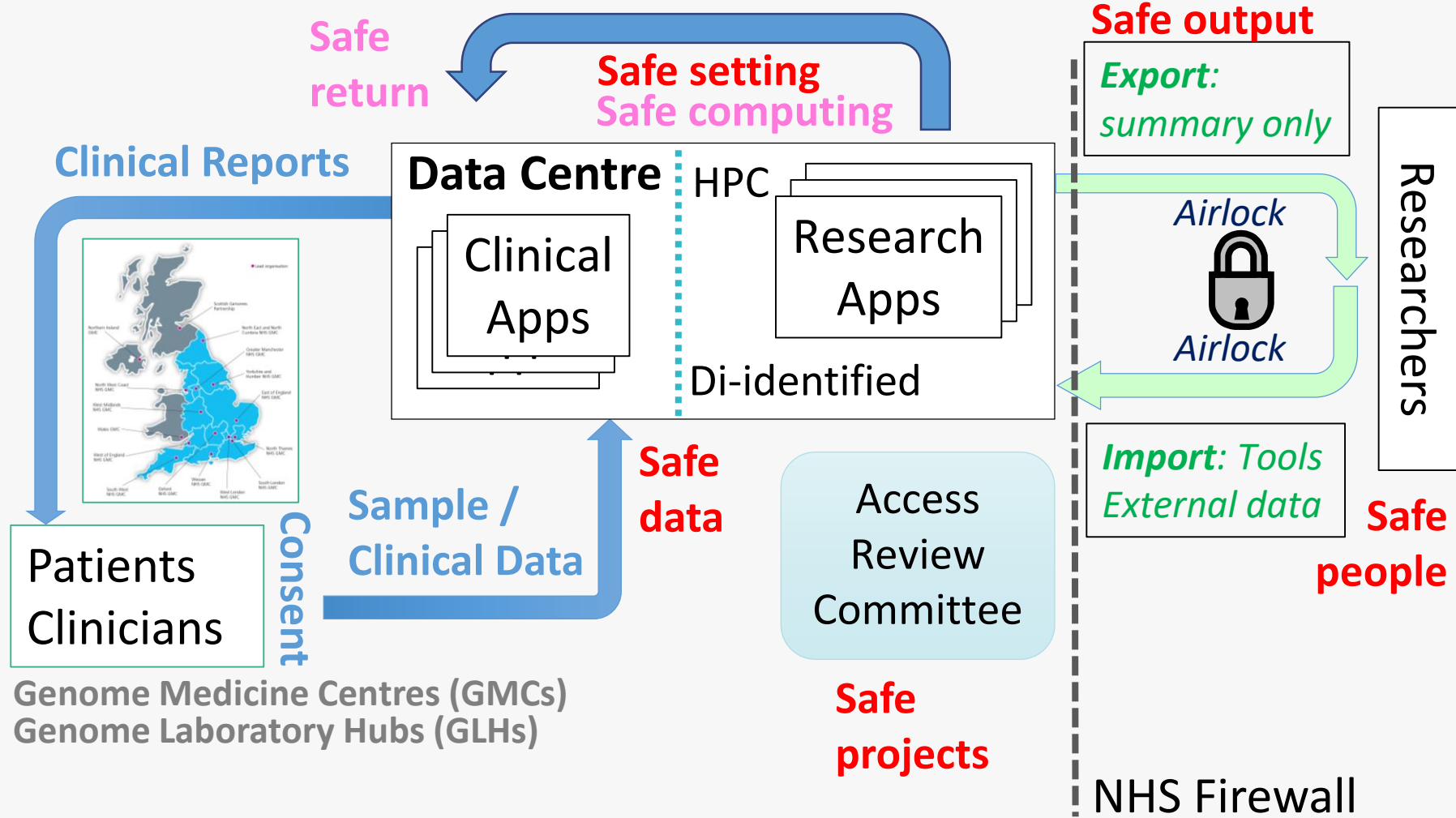
- Issues
 - Data custodian loses of control of data
 - Cannot completely guarantee anonymity
 - Genome data can only be de-identified
 - Lack of public trust
 - Holding health data can be a liability under GDPR
 - Overall expense of transferring, storing, maintaining multiple copies of large datasets

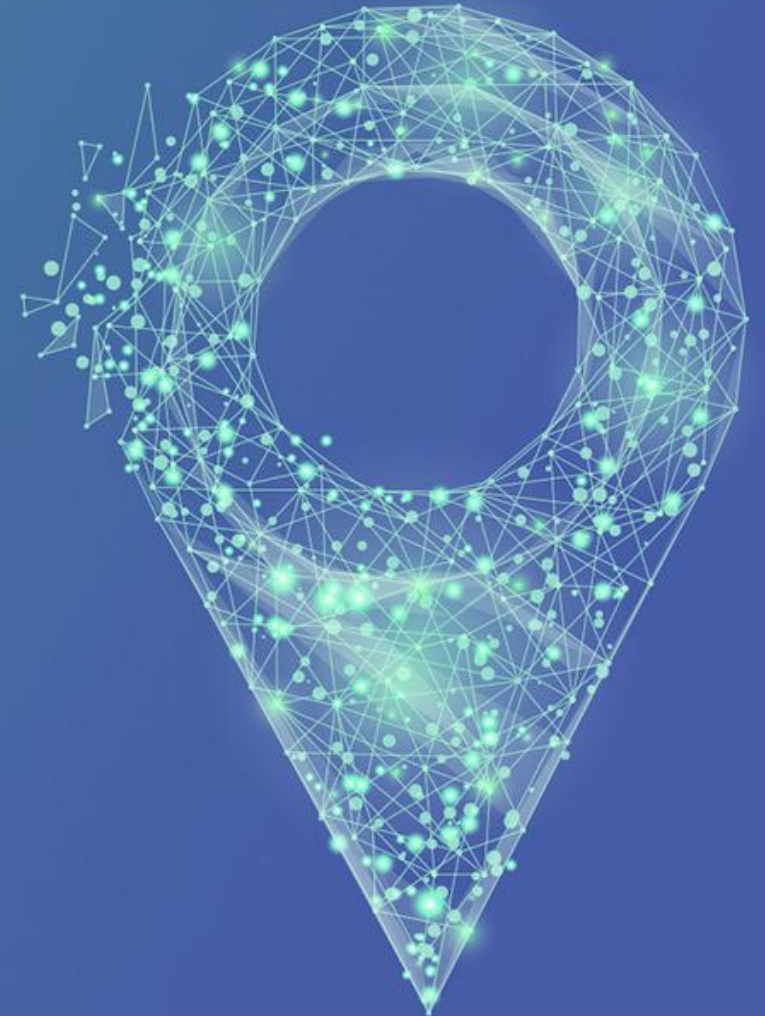
Data reuse *via access*

- Generic anonymised dataset prepared within a Trusted Research Environment (TRE)
- Generic request for access received, reviewed, approved
- User logs into TRE and carries out analysis
- User requests export of results (summary data)

- Benefits
 - Data custodian maintains control of data; retains public trust
 - Data privacy no longer depends on anonymization process
 - Less costly holding and securing a single copy of the data
 - Virtualisation / cloud computing make being algorithms to date and provision of scalable compute environment practical
 - Privacy controls shifts from individual to summary level

Data design for 100,000 genomes project Trusted Research Environment (TRE+)





World café sessions

Lunch: 13:10-14:00

World café pitches: 14:00-14:25 (*livestreamed*)

World café discussions x 4 (15 mins each): 14:25-15:25

Refreshment break: 15:25-15:40

@HDR_UK | #hdralliance

World café: format & purpose

- **Purpose:** the world café roundtables are an opportunity to network, showcase work across the infrastructure and celebrate achievements
- **Format:**
 - Each world café host will deliver a 1 minute pitch. You will then decide which four to attend.
 - Join 4 x 15 minute world café discussions over one hour
 - These are informal – enjoy!



World café

1. Data infrastructures: How will we know when we've got it right? Paul Jackson & Elizabeth Waind, ADR UK	11. How can NHS DigiTrial deliver the maximum impact for customers and Alliance members Heather Pinches & Charlie Hooper, NHS Digital/IBM
2. Methods and evidence standards for data Adrian Jonas, NICE	12. Real-World Evidence: Current barriers and future wishlists Chris Russell, Discover-NOW – The Health Data Research Hub for Real World Evidence
3. HDRA terminology: Are we speaking the same language? Macey Murray, UCL	13. Knowledge based metadata to drive automation and interoperability Luke Smith, Databiology
4. What does a fair partnership for accessing NHS patient data look like?	14. How can Alliance members work with DATA-CAN? Charlie Davie, DATA-CAN – The Health Data Research Hub for Cancer
5. How can we become more efficient providers of health data for researchers? A regulator's challenge Amanda Evans & Stevan Cirkovic, HFEA	15. How do we assess, capture and report the non-financial value of Hubs? Monica Fletcher, BREATHE – The Health Data Research Hub for Respiratory Health
6. How do we evaluate the impact of using healthcare data to support safer prescribing? Arlene Gallagher, CPRD	16. Can we, should we, use the public cloud for analysis of health data? Neil Walker, NIHR Bioresource
7. Data + Human Tissue = Impact Katy Glazer & Amir Gander, UKCRC Tissue Directory & Coordination Centre	17. CANCELLED
8. Health Data and Industry – opportunities for health data to drive research and innovation in the SME sector Joanne Hartley, MDC	18. Use of FHIR for data science and app delivery to the clinic David Sibbald & Rodrigo Barnes, Ardihia
9. Open science, GitHub, and open approaches to health data research Caroline Morton & Helen Curtis, EBM Datalab, University of Oxford	19. HDR UK metadata standard: future progress Ashutosh Tripathi, Parity & Adam Milward, MetadataWorks
10. Standardising national data access for health tech product development e.g. machine learning components (commercial or non-commercial) Deborah Wake, MyWay Digital Health	20. What are the skills and credentials required to be a user of data? George Moulton, HDR UK



Closing keynote & wrap up

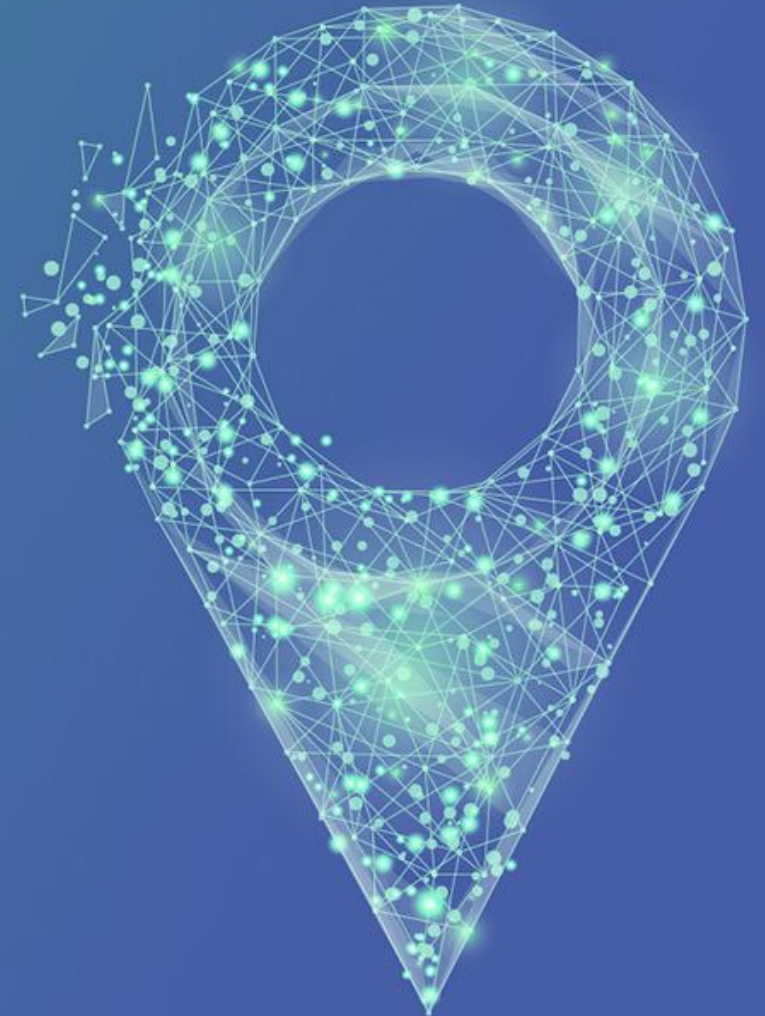
Refreshment break: 15:25-15:40

Closing keynote: 15:40-16:00

Wrap-up: 16:00-16:15

Networking reception: 16:15-17:00

@HDR_UK | #hdralliance

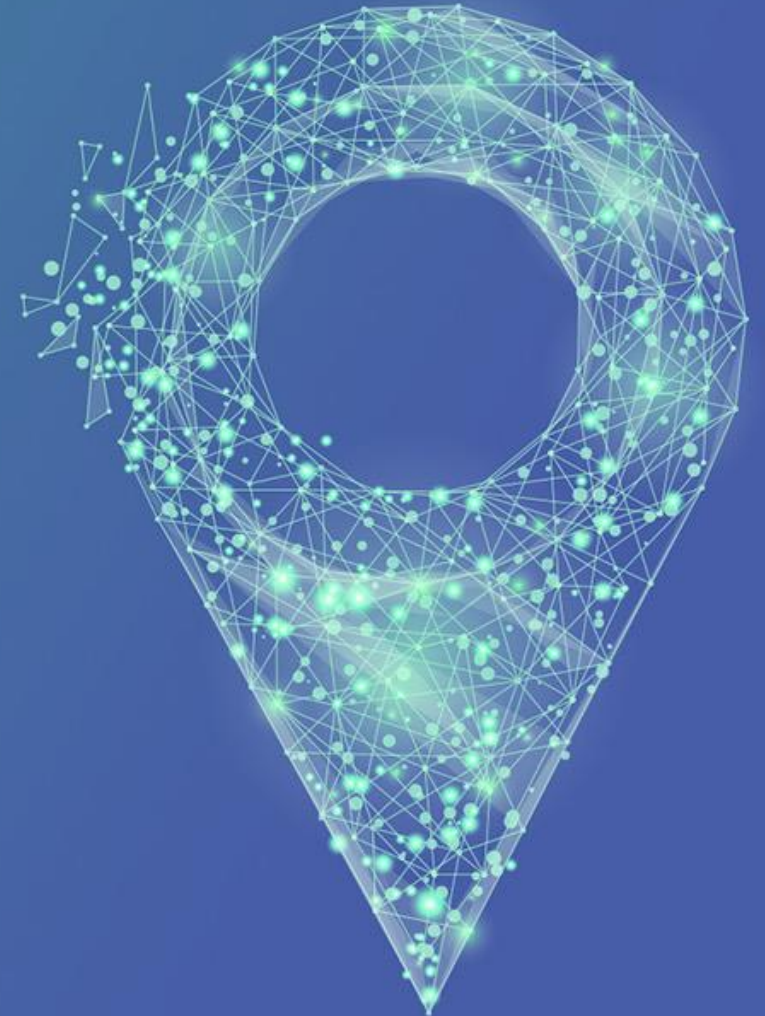




Closing keynote: Trustworthy access to data

Jeni Tennison, CEO, Open Data Institute

@HDR_UK | #hdralliance



Trustworthy access to data

Jeni Tennison

@JeniT

theODI.org



The ODI



**Dr Jeni
Tennison**
CEO



**Sir Nigel
Shadbolt**
Chairman



**Sir Tim
Berners-Lee**
President

Founded in 2012, the Open Data Institute (ODI) is an international, independent and not-for-profit organisation based in London, UK.



Vision

We want a world where data works for everyone.



Mission



We work with companies and governments to build an open, trustworthy data ecosystem.

Trustworthy access to data

- ▶ **The data value chain and threats to its impact**
- ▶ The importance of trust and how to retain it
- ▶ Emerging institutional models for trustworthy access



Getting value from data

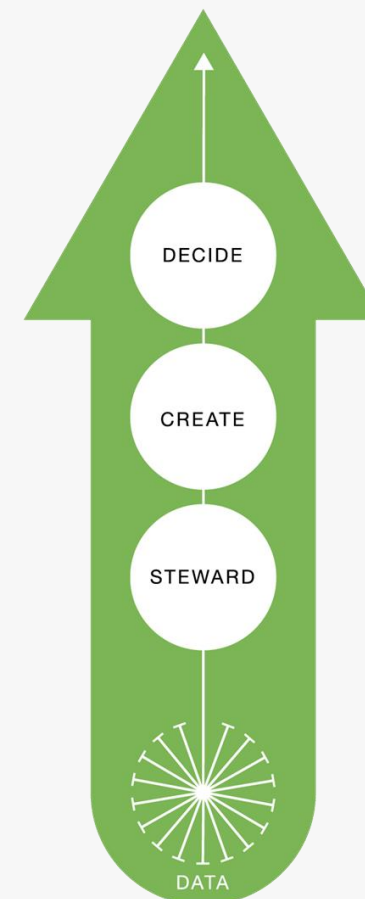
Many different people and organisations making decisions

Many creating tools & insights that help those decision makers

Requires stewards to share data

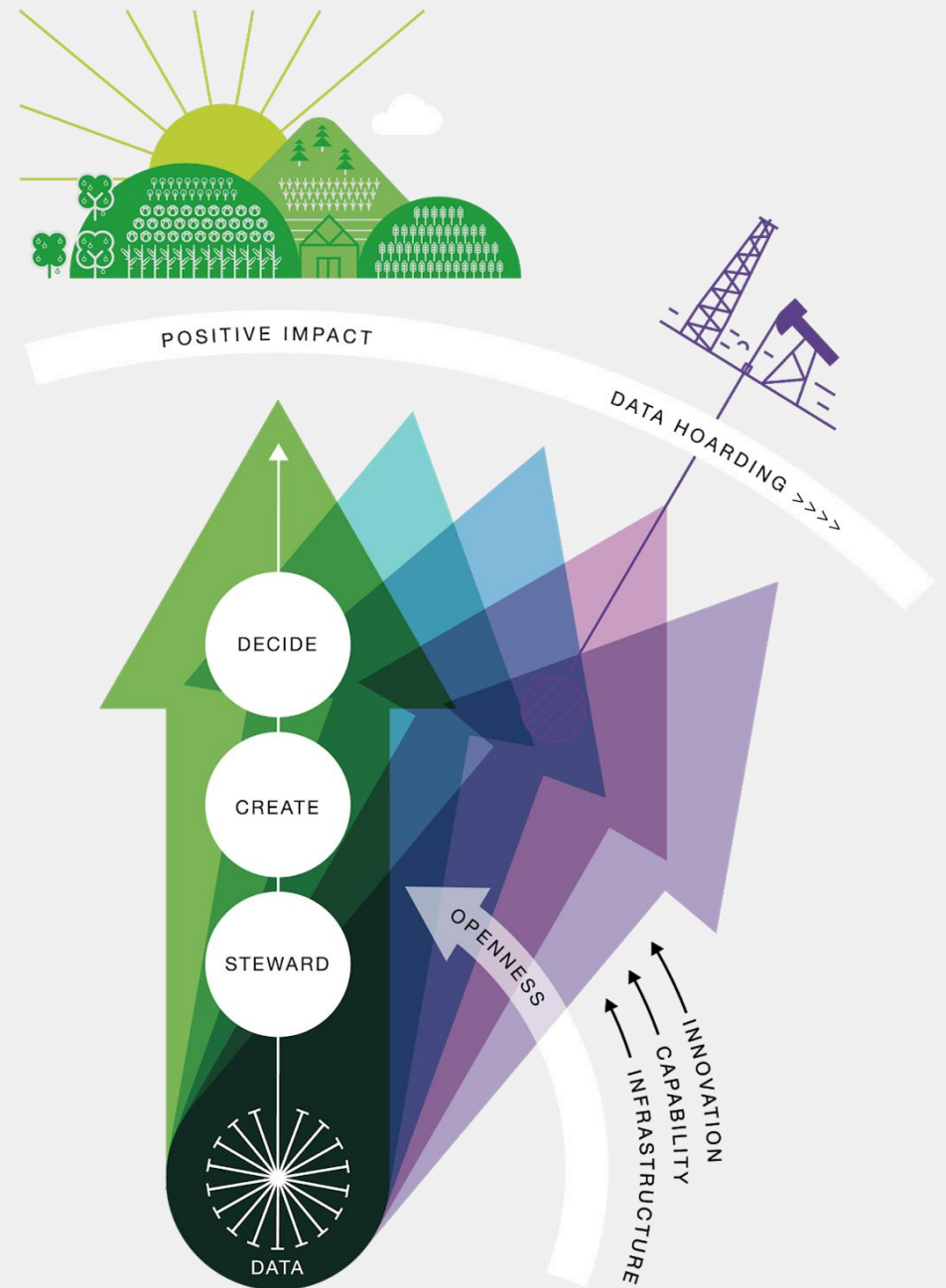


POSITIVE IMPACT



Data hoarding threat

- “Data is oil” attitude
- Vertical, in-organisation data use
- Narrow decision maker needs
- Restricted intermediary market
- Data extraction



Data fearing threat

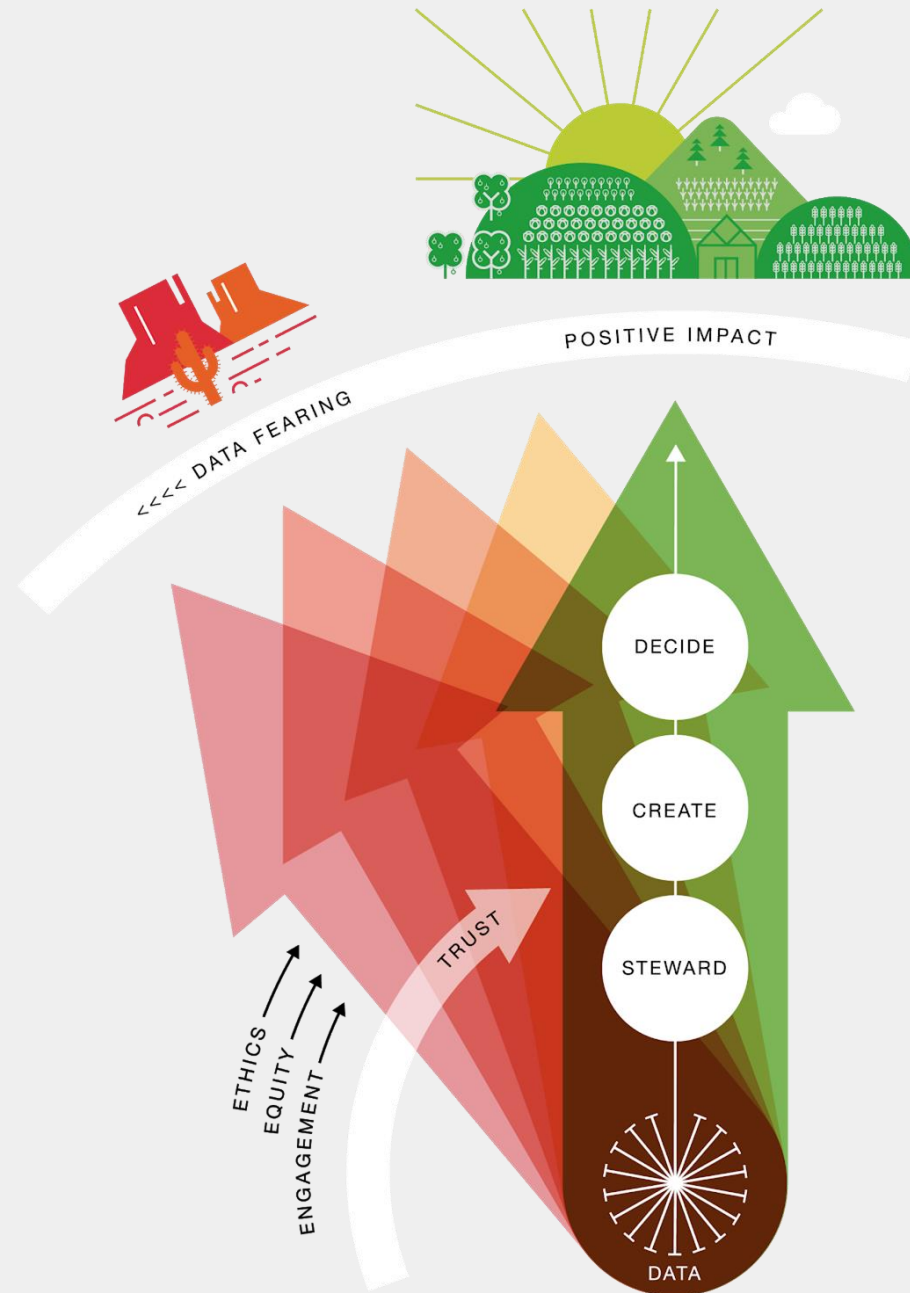
Not collecting data

Not using data

Withdrawal of consent

Misleading data

Poor conclusions



Trustworthy access to data

- ▶ The data value chain and threats to its impact
- ▶ **The importance of trust and how to retain it**
- ▶ Emerging institutional models for trustworthy access



Fear and distrust is infectious

care.data

Distrust about personal health records being sold to pharma and other companies.

Royal Free/Google DeepMind

ICO ruling that Royal Free breached Data Protection Act; distrust about Google's access to patient records.

General distrust of big tech

Most surveys show only about 4-5% of people trust big tech to handle data ethically.

General fear of NHS privatisation

Political debate about NHS provision and role of the private sector.




Trusted

How can organisations win the trust of others?



Trustworth



What does doing the right thing look like?

Points of failure in trust

Misunderstanding

When the people you want to trust you don't interpret what you tell them in the way you intended

Misrepresentation

When you don't do what you say you're going to do (whether that was intentional or not)

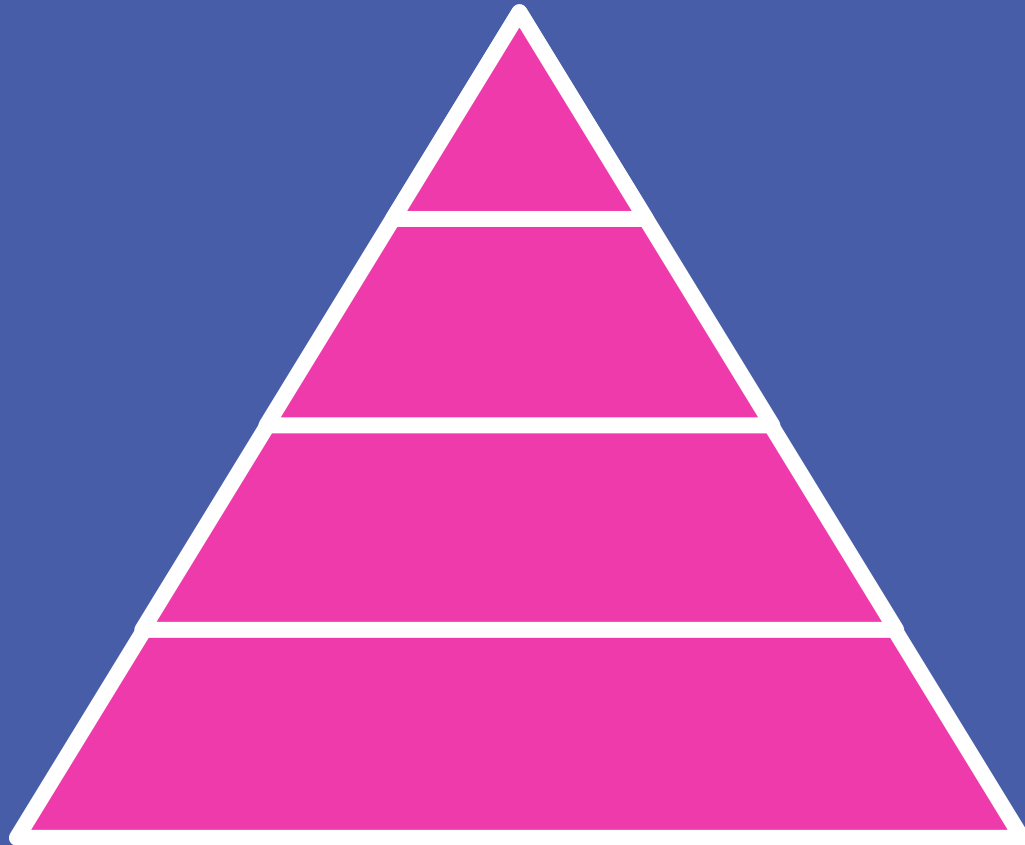
Lack of updates

When you don't tell people how and when circumstances change, that impact your actions

Lack of information

When you and others don't have enough information to make sound decisions

Trustworthy access to data: **Hierarchy of trust**



equity

engagement

ethics

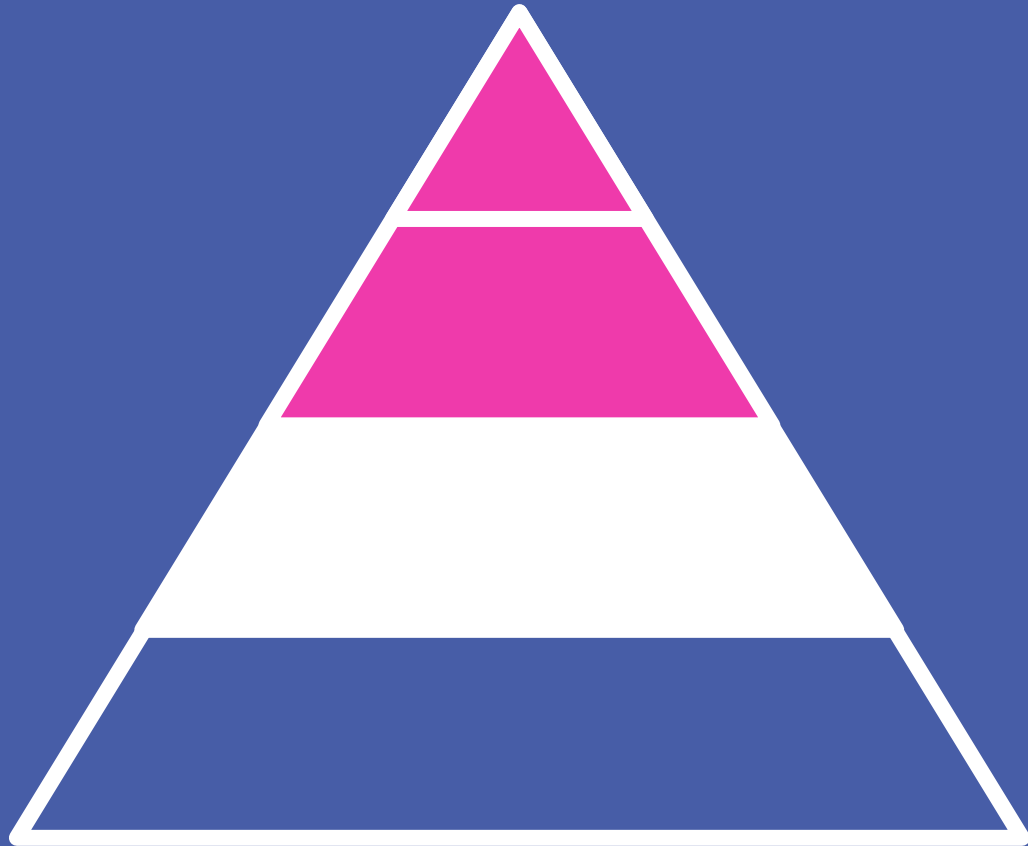
compliance

Trustworthy access to data: **Hierarchy of trust**



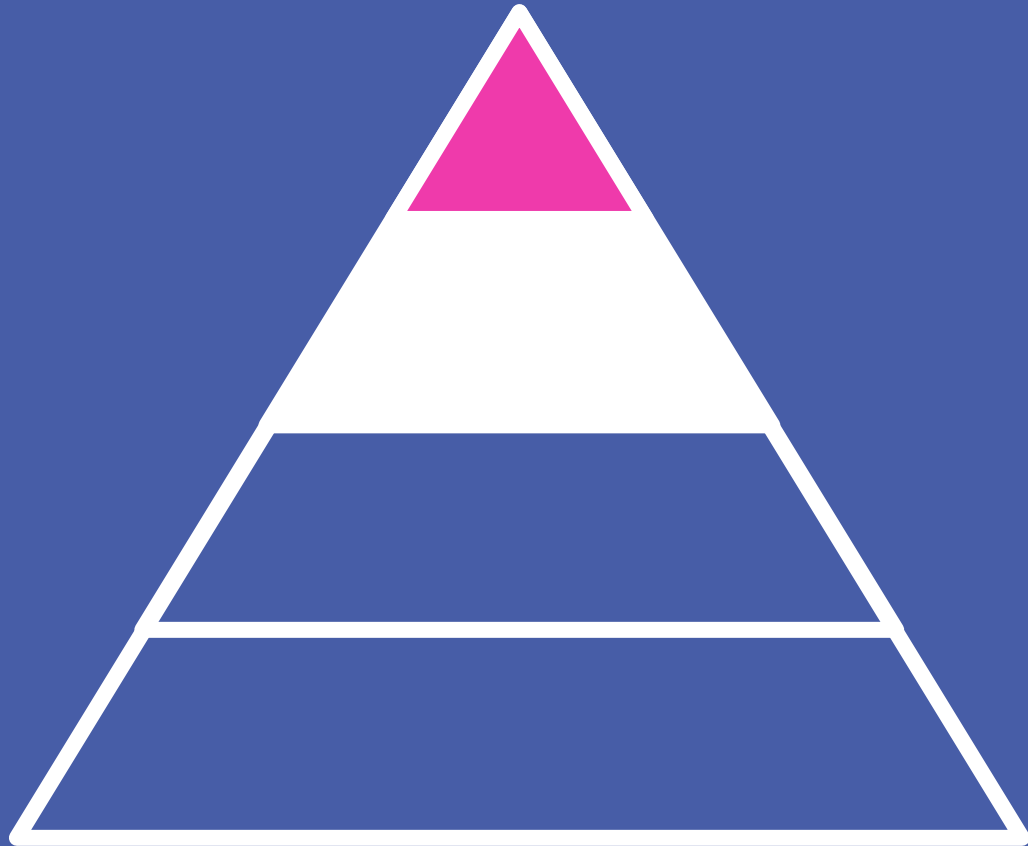
Compliance

Protect personal data
Comply with the law Comply
with codes of practice



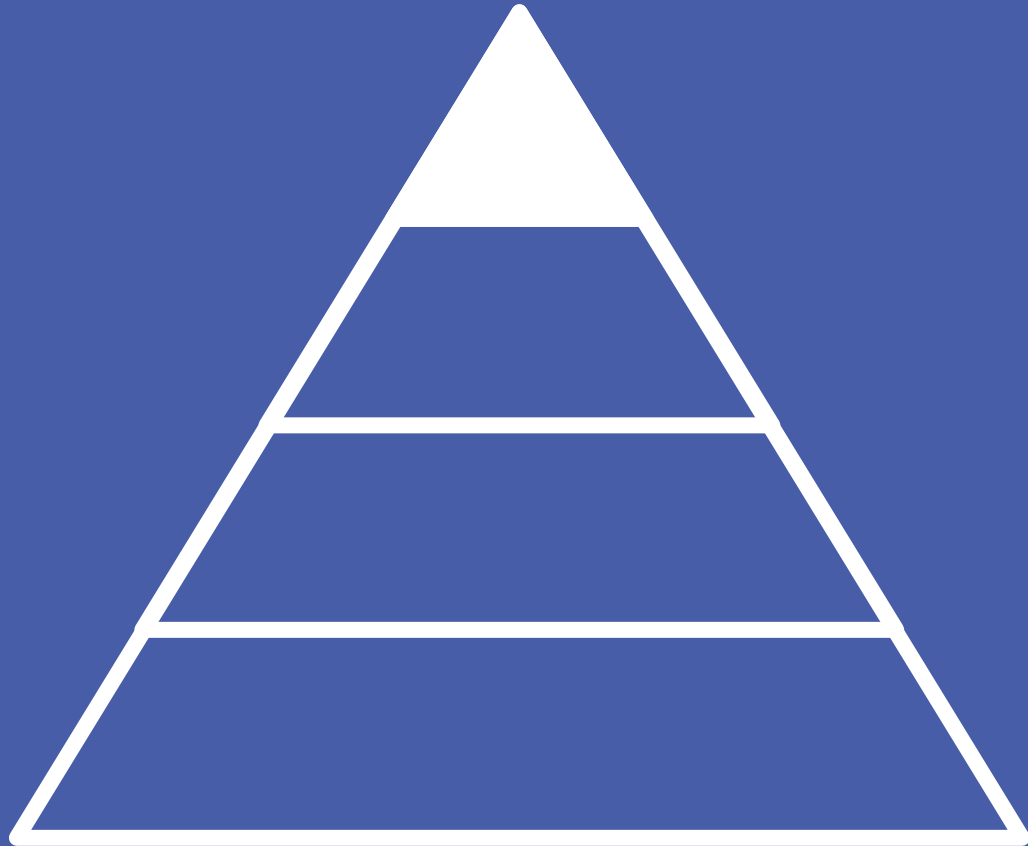
Ethics

Be open and transparent
Publish ethical principles
(eg HDRUK DIH Principles of
Participation)
Do compliance well
Embed ethical practices



Engagement

Establish advisory boards
Ask affected communities
what they think (eg PPIE)
Act on what they say
Give them veto power



Equity

Describe revenue models
Describe value return
Ensure equity in who benefits
Build into corporate structures

Trustworthy access to data

- ▶ The data value chain and threats to its impact
- ▶ The importance of trust and how to retain it
- ▶ **Emerging institutional models for trustworthy access**

DASHBOARDS

TECHNICAL EDONIA

THE TENDER ISLES

DATA PHILANTHROPY

SHARED BIOME

OD ISLAND

SYNTHETIC DATA

THE PRETENDER

THE INDEPENDENT ISLE

DATA TRUSTS

GOVERNORS ISLAND

PERSONAL DATA STORES

THE ISLE OF HUMAN

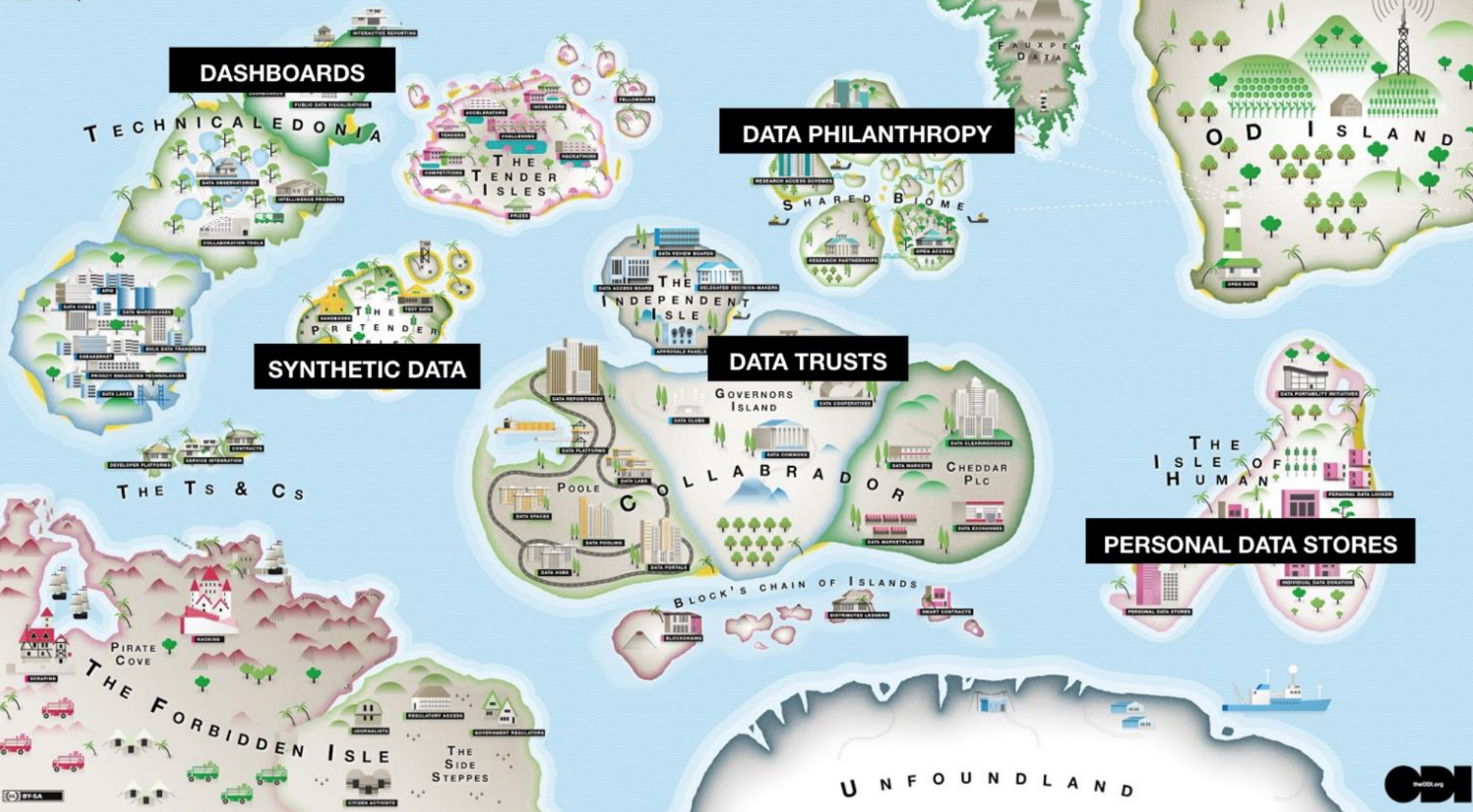
THE TS & Cs

COLLABRADOR

THE FORBIDDEN ISLE

BLOCK'S CHAIN OF ISLANDS

UNFOUNDLAND





Data institutions

Increase data access

Enabling organisations to work and innovate together across a sector or for purpose
Supporting benchmarking
Creating new opportunities
Ensuring the benefits of data are shared more widely

Balance conflicting views and incentives

Between those who have data, those who could use it, and those affected by the way it's used

Reduce the costs of data stewarding

By providing a common service to manage the operations of well governed data access

Data institutions

Data clubs

Inspired by members clubs, enable organisations to share data with other organisations who are part of the club

Data commons

Inspired by the management of common pool resources, help organisations or people collaborate to create and maintain shared data assets

Data cooperatives

Inspired by mutual organisations, enable members of the cooperative (organisations or people) to share data with others, with data governance decisions made by those members

Data trusts

Inspired by legal trusts, enable people or organisations to share data with others, with data governance decisions made by 'trustees' with fiduciary responsibilities

Trustworthy access to data



Trustworthy & trusted

What could the external environment provide to help confer trust (eg accreditation, certification, insurance)?



Sustainable

What structures and revenue models provide for longevity, reliability and coherence with purpose?

Trustworthy access to data

- ▶ The data value chain and threats to its impact
- ▶ The importance of trust and how to retain it
- ▶ Emerging institutional models for trustworthy access

Any questions?

If you would like to talk to us about collaborating, partnering, supporting our work, or anything else, we'd love you to get in touch.

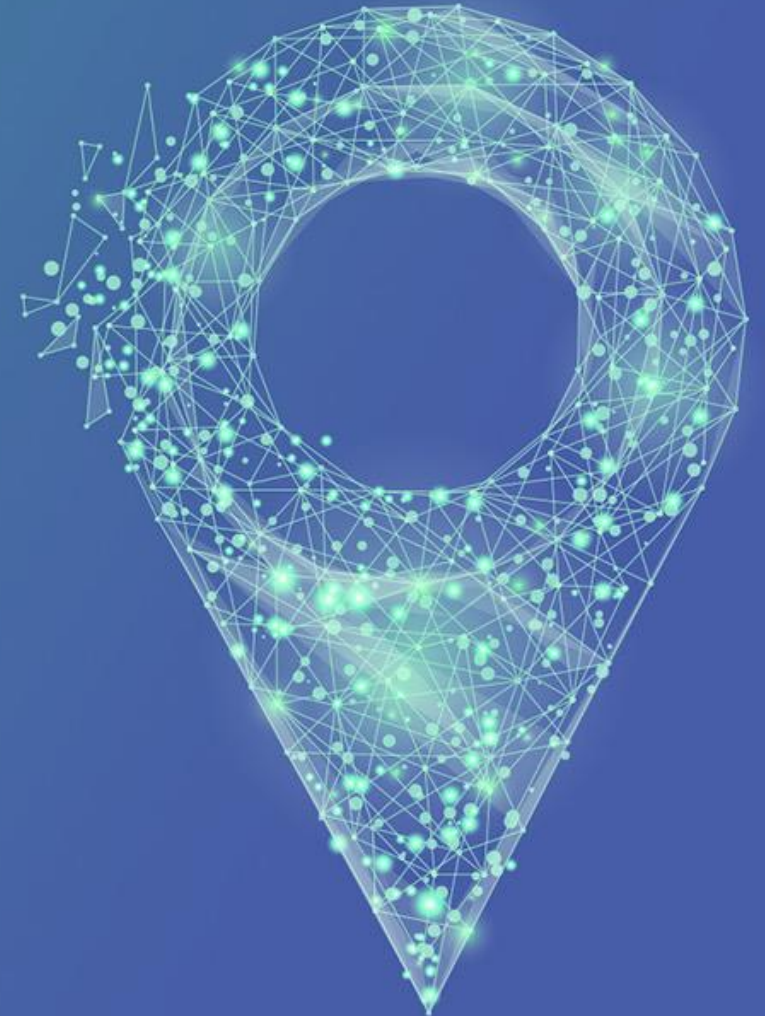
jeni@theodi.org

+44 (0)20 3598 9395

[@ODIHQ](https://twitter.com/ODIHQ)

RUUK
with Data Resea

 open
data
institute



Wrap up & close

David Seymour, Partnership Director, Health Data Research UK

@HDR_UK | #hdralliance

HDRUK
Health Data Research UK

UK Health Data
Research Alliance



Thank you!

11/02/2020 |