

Deep dive: exploring data platform and research initiatives

Prof Tim Hubbard

Associate Director
HDR UK London

Prof William Dixon

Chair in Digital Epidemiology,
Division of Musculoskeletal &
Dermatological Science
University of Manchester

Chair: Natalie Chishick

Director
Silver Buck

Dr Lauren Walker

Senior Clinical Lecturer in
Clinical Pharmacology &
Therapeutics and Internal
Medicine
University of Liverpool

How the Technology Ecosystem can help address the grand challenges in health data research

Health Data Research (HDR) UK

07/11/2023 | Professor Tim Hubbard (for Professor Emily Jefferson)



MISSION: to unite the UK's health data to enable discoveries that improve people's lives

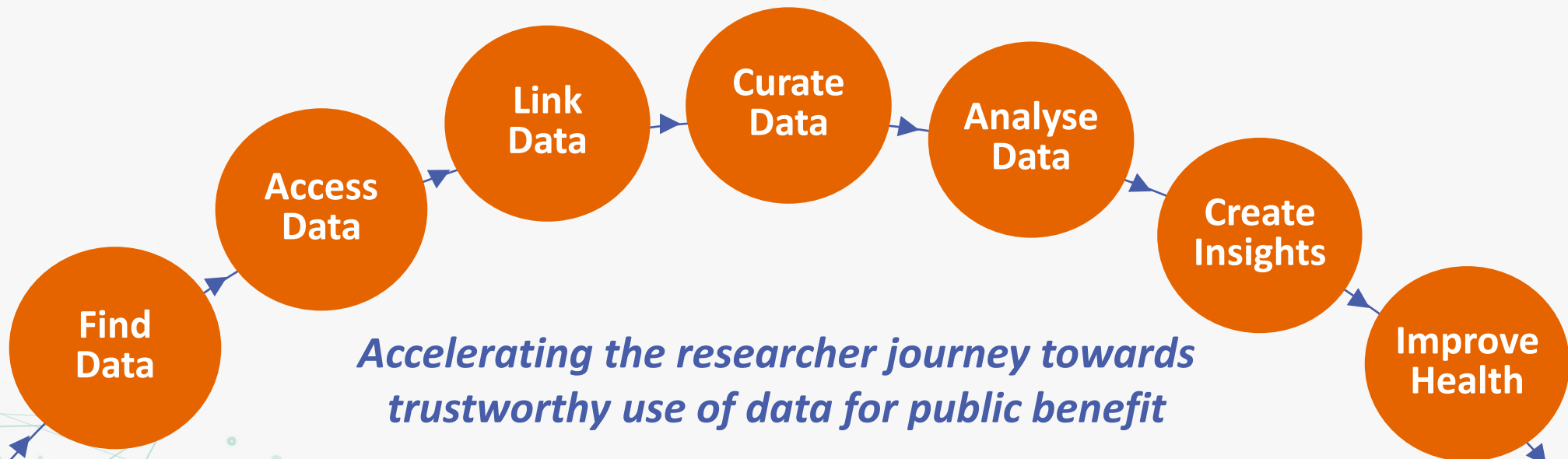
VISION: for large-scale data and advanced analytics to benefit every patient interaction, clinical trial and biomedical discovery, and to enhance public health

What are we about?

Improve health and boost UK science by making it easier for researchers to find, access and use diverse, high quality data

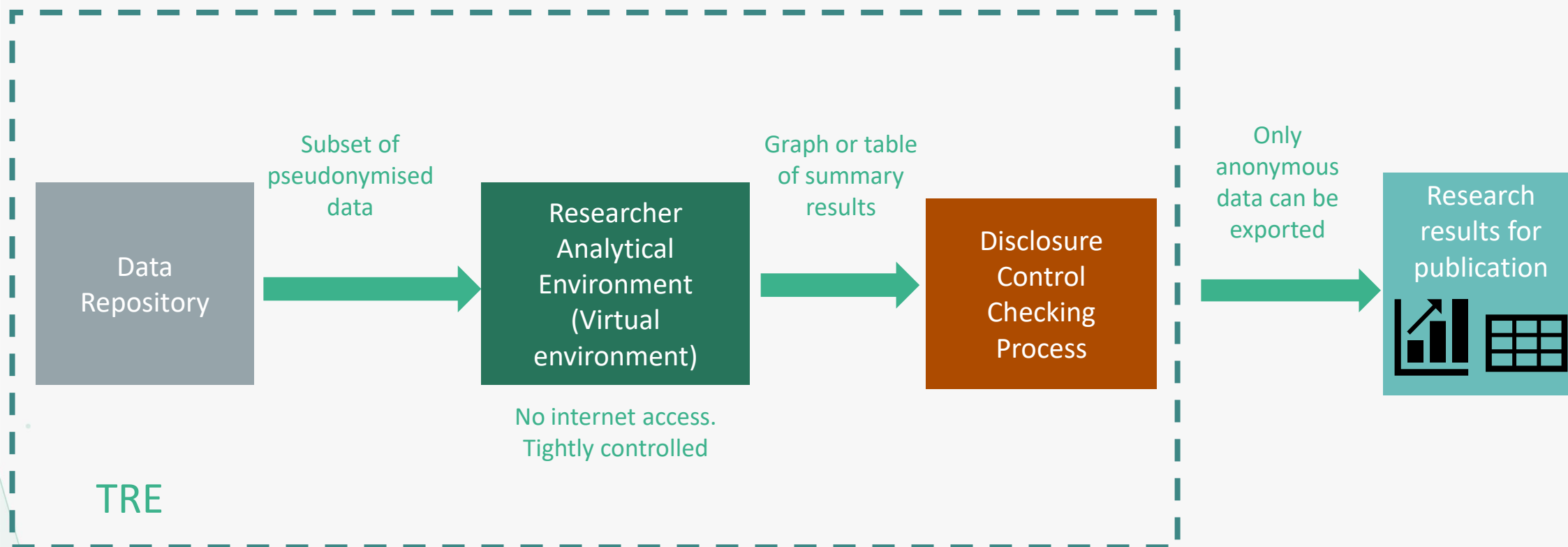
Provide leadership to fix difficult technical problems, by creating innovative solutions needed for researchers to use large-scale data safely and securely

Accelerate & streamline health data science by developing open collaborations that connect data, people and organisations across the UK and internationally



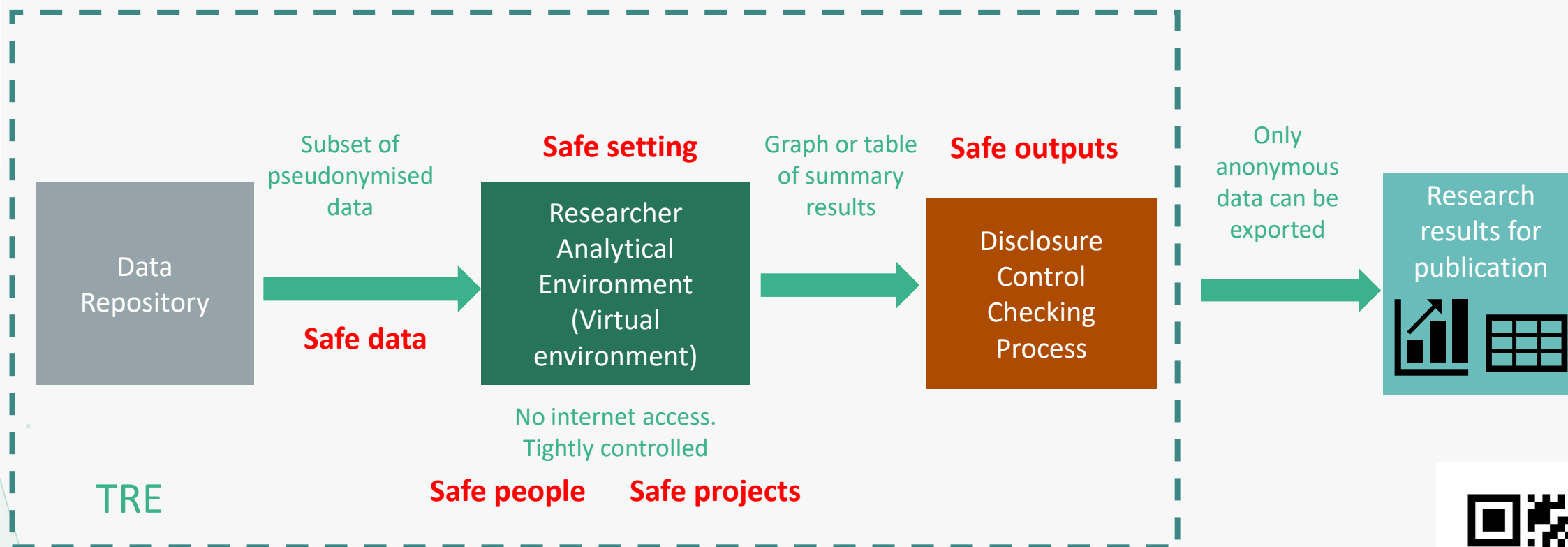
What is a Trusted Research Environment (TRE)?

(Also known as a Secure Data Environment (SDE) or Safe Haven)



What is a Trusted Research Environment (TRE)?

(Also known as a Secure Data Environment (SDE) or Safe Haven)

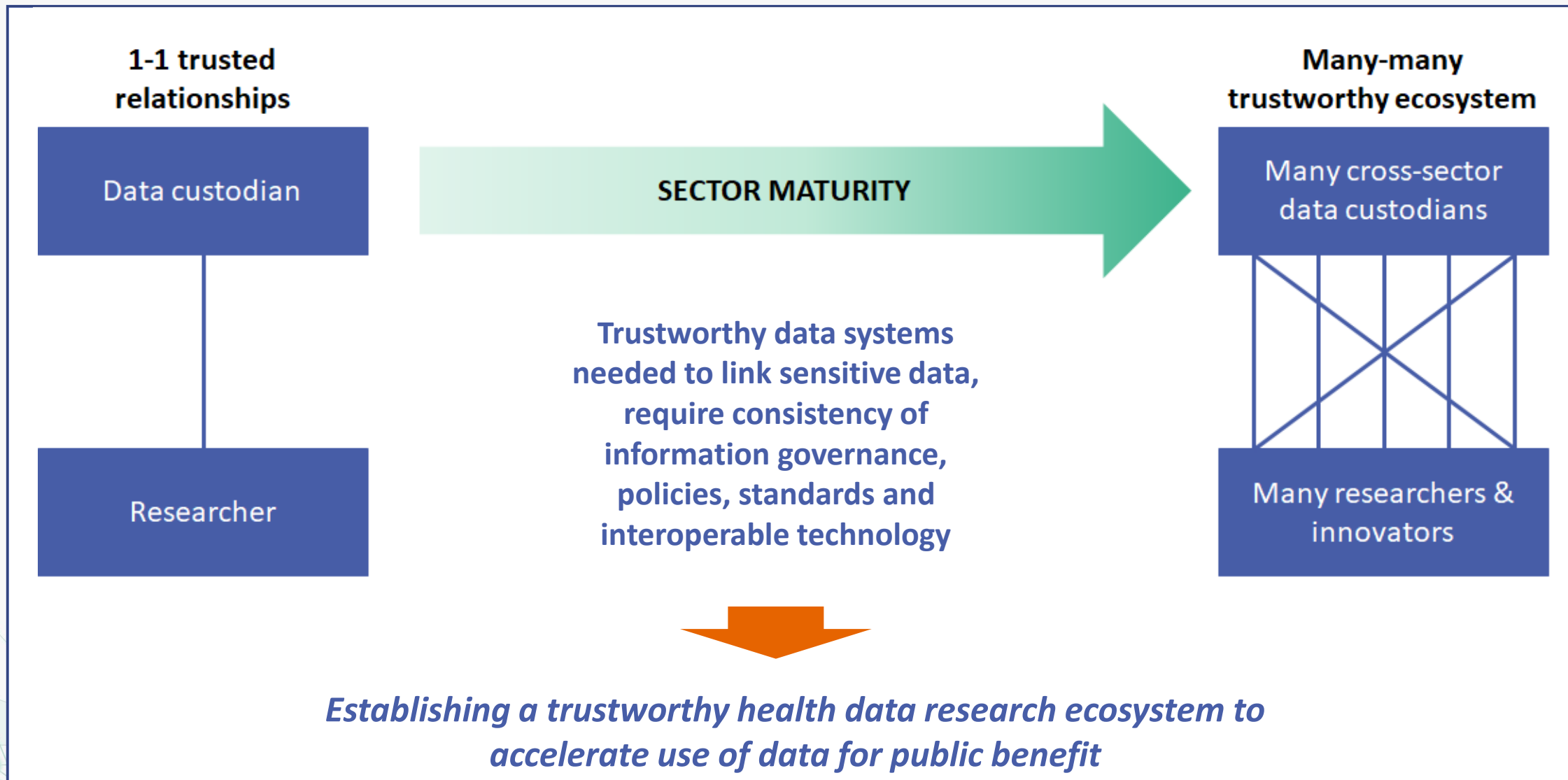


<https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>

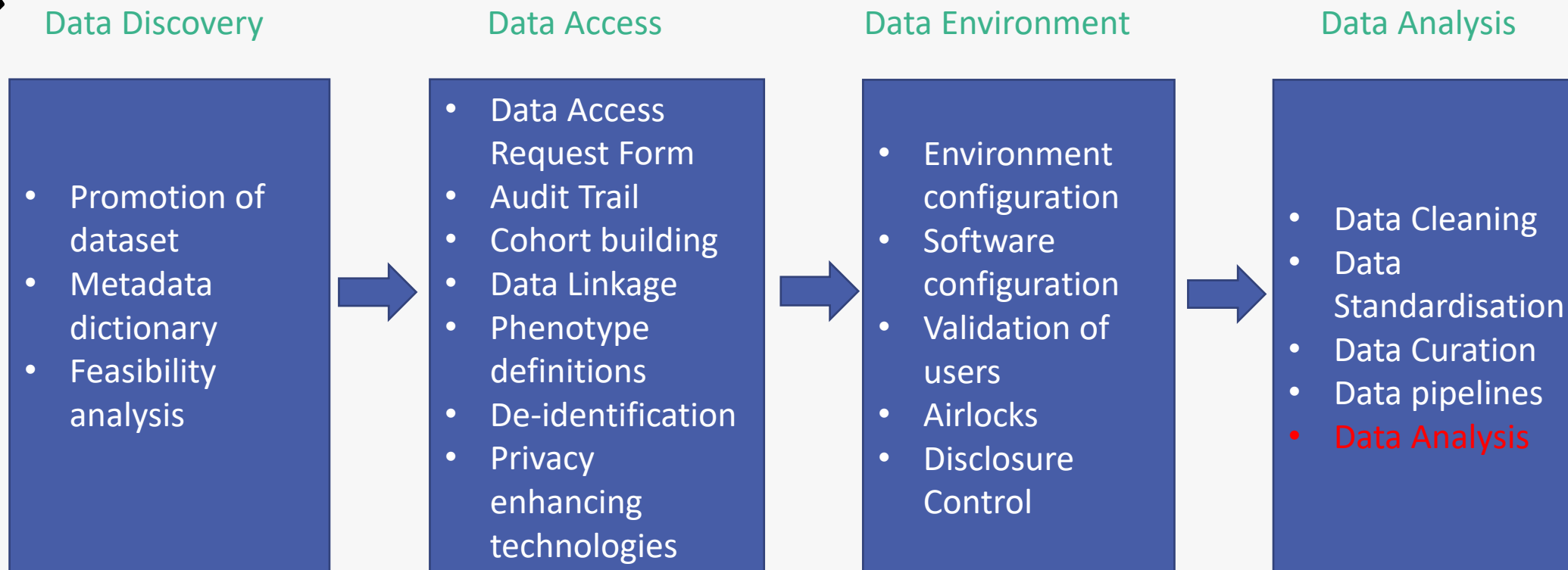
<https://zenodo.org/record/5767586#.ZADDR3bP2Ht>



The problem HDR are trying to solve

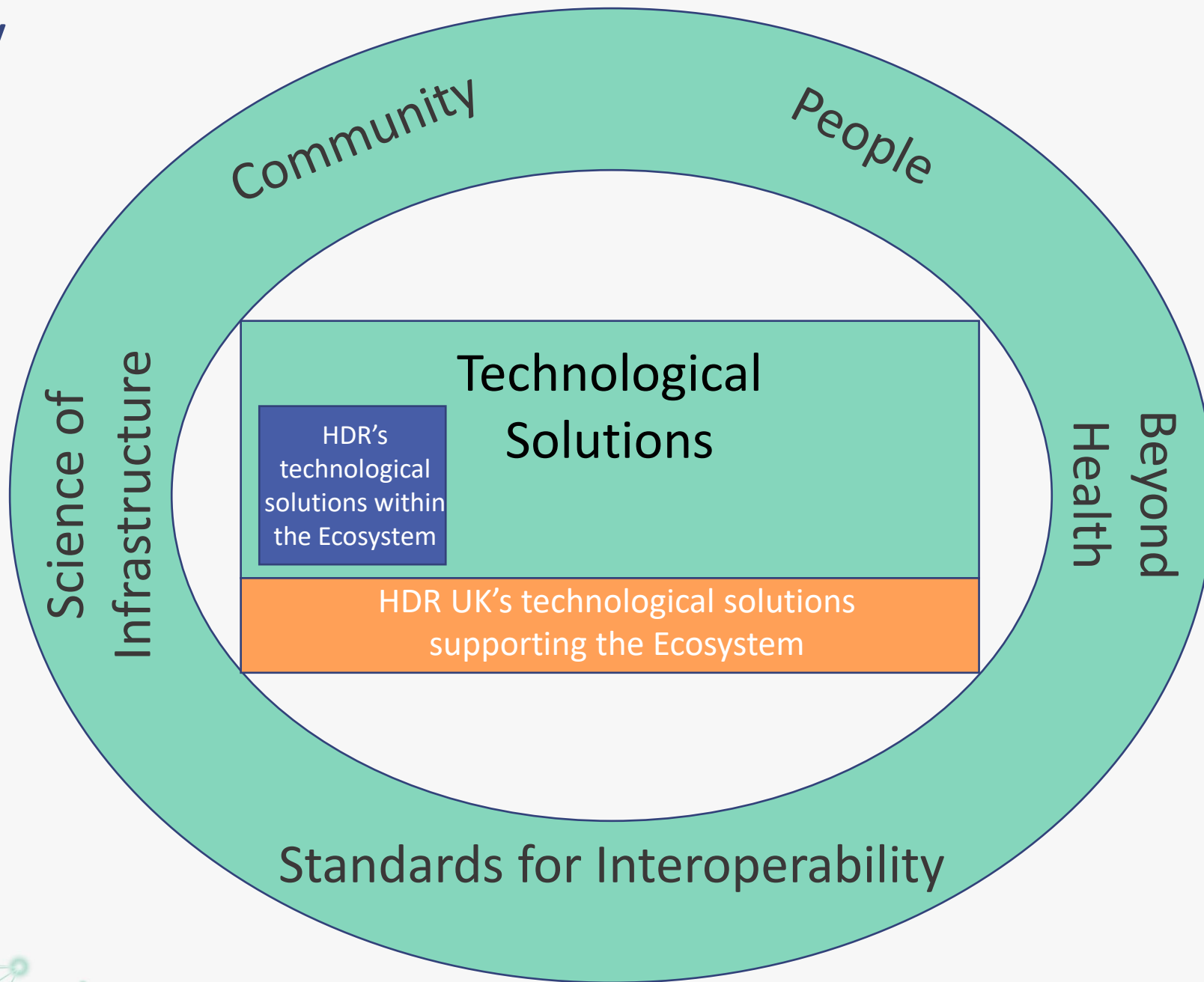


Technology across research journey



TREs, SDEs, Hubs, Research Cohorts, Research Programmes have all built tech across this space

Technology Ecosystem



Next 5-year HDR funded technology components

Gateway

Phenotype Library

Federated
Analytics

Driver
Programmes

Whole is greater than the
sum of the parts!

- Interoperable tech
- Re-use and sharing of tech/code
- Training and capacity building
- Knowledge sharing
- Community

Hubs

BHF DSC

Prognostic Atlas

Data Standards

~60 Technologists working across these components

Next 5-year HDR funded technology components

Gateway

Phenotype Library

Driver
Programmes

Federated
Analytics

Hubs

BHF DSC

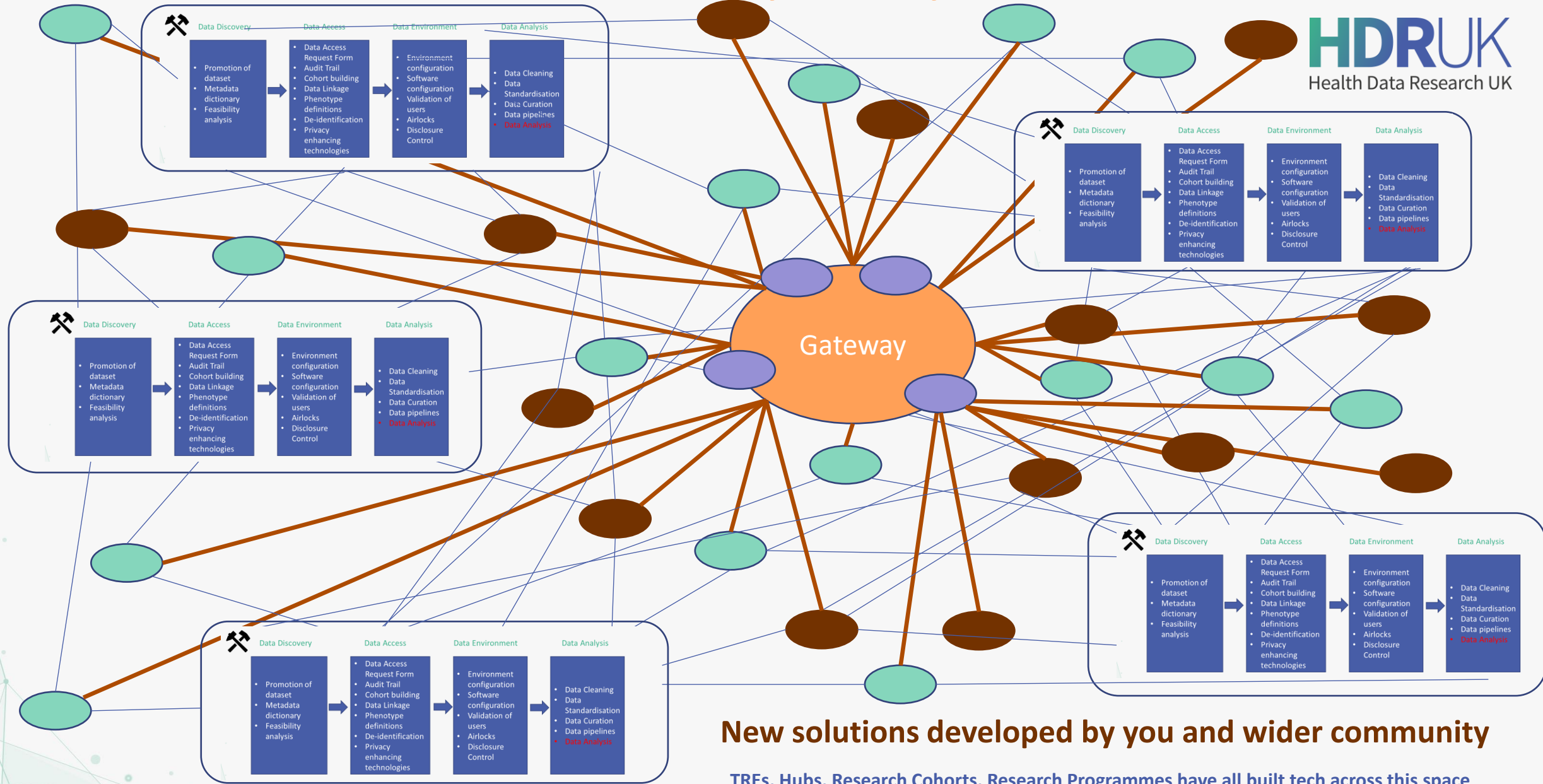
Prognostic Atlas

Data Standards



Some great solutions already developed – but not widely shared or systematically used

Gateway – Directory to solutions



New solutions developed by you and wider community

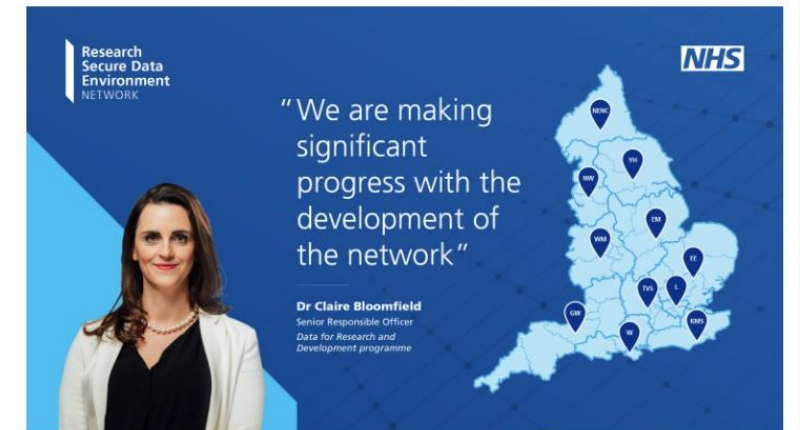
TREs, Hubs, Research Cohorts, Research Programmes have all built tech across this space

Gateway – Developing solutions / interoperability with existing solutions

Developing the Innovation Gateway as the 'single front door' for the NHS Research SDE Network

“The Gateway will evolve to provide users with a clear and consistent user journey for data discovery and cohort finding; reducing complexity and confusion for those wanting to securely access data to improve care and reduce burden on the NHS.”

<https://www.linkedin.com/pulse/transforming-data-enabled-research-landscape-england-bloomfield/>



Transforming the data-enabled research landscape in England

The Current Gateway (Mk1)



Data Discovery: Metadata



Data Access Requests
(DAR): Enquiries,
Application & Approvals



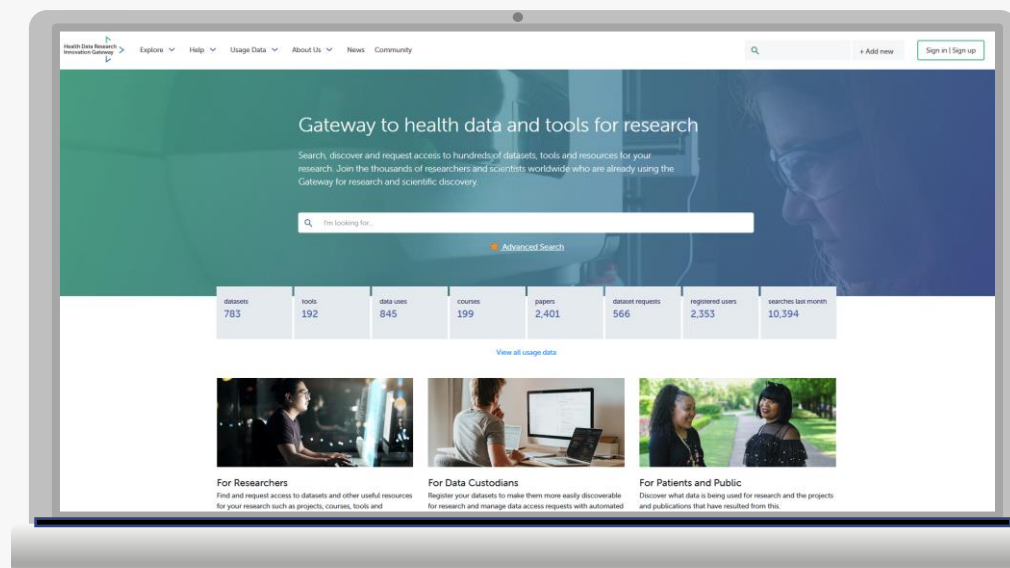
Data Use Register:
Transparency on use of
data:



Data Discovery: Cohort
Discovery Tool



Data Discovery:
Collections, Tools,
Publications



<https://www.healthdatagateway.org/>

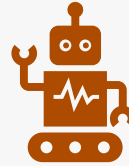
Gateway (Mk2) – enhanced design

Reducing manual curation of datasets, publications and tools/software

Community rather than HDR's Gateway: help us build it to meet your network's needs



Co-created



Automated



Interoperable

Lots of other great tools (collaboration rather than competition)

To automate & interoperate with other tools

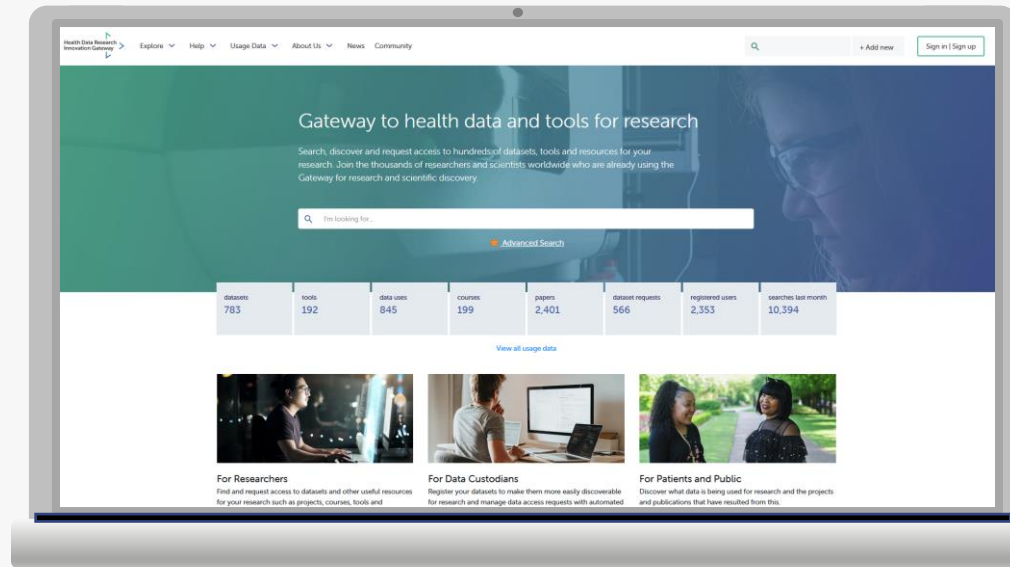


Open Source & APIs

Understanding what is possible

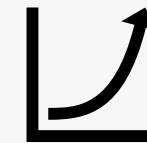


Visualisations



Plug-in technologies

Assemble, re-use and enhance



Scalable and modular

As requirements & data increases

Gateway (Mk2) – enhanced features

Use Gateway tech in your website and brand your entries on the Gateway

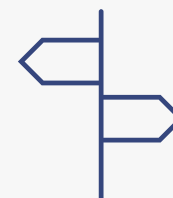
Using NLP and Elastic Search



Advanced Search



Brandable



Sign-posting

Point people at other solutions

Analysis scripts, detailed metadata, federated analytics

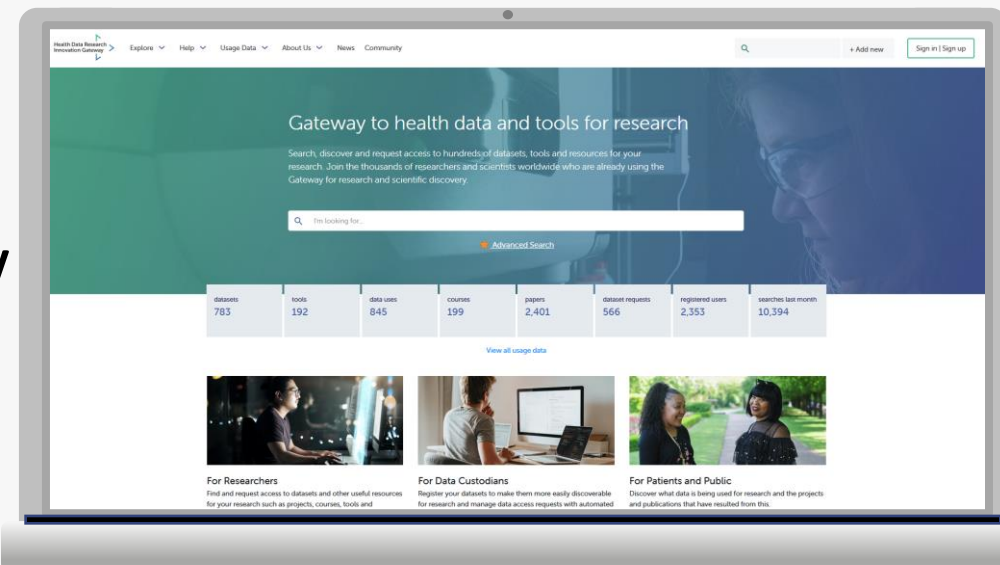


End to End Researcher Journey

DCAT, DDI, Schema.org



Multiple metadata standards



Multiomic, imaging, open, non-health

Additional Data Types

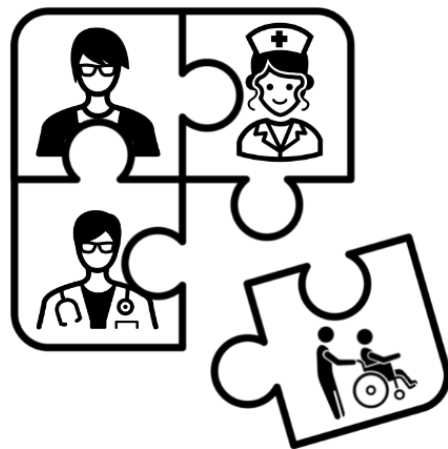


International exposure

Share with Gateway - we share with other networks



Thanks for listening!



Assembling the Data Jigsaw

Powering robust population research in MSK disease

Will Dixon, Professor of Digital Epidemiology & Honorary Consultant Rheumatologist

Digital Health AI+Data, 30Oct23

**CENTRE FOR
EPIDEMIOLOGY
VERSUS
ARTHRITIS**

MANCHESTER
1824
The University of Manchester

NHS
Northern Care Alliance
NHS Foundation Trust

Funded by:

**Nuffield
Foundation**

**VERSUS
ARTHRITIS**

Electronic health record (EHR) research

- Health data supports clinical care, with additional opportunities to support research and more



Electronic health record (EHR) research

- Health data supports clinical care, with additional opportunities to support research and more
- Limitations, however, come from the availability and quality of data
- Musculoskeletal research exemplifies many health data problems:

1. Absent



2. Inaccurate



3. Unstructured

Date/Time of Appt: 28th July 2015 at 09:00
Clinic: RHEUMATOLOGY
Type of Appt: Follow Up

Rheumatological Diagnoses: Osteoarthritis
Fibromyalgia
Anxiety and depression
Previous vitamin D deficiency

Non-Rheumatological Diagnoses: Ischemic heart disease leading to ST elevation
MI 2006
Type II diabetes
Migraine

Medication: Naproxen
Bisoprolol
Simvastatin

Mr. Dixon attended the clinic today with ongoing symptoms of fatigue. He appeared concerned that his current sleep pattern may be contributing. His knee osteoarthritis was not too significant.

4. Siloed



Assembling the Data Jigsaw



- A **'local data integration pilot'** that

1. Pilots **novel data collection, extraction and integration** in rheumatology,

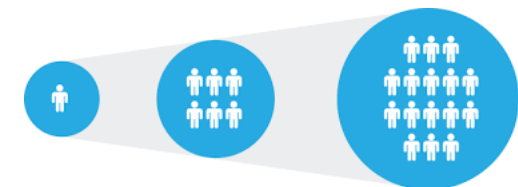
in order to...

2. Answer **important research questions**,



and designed to...

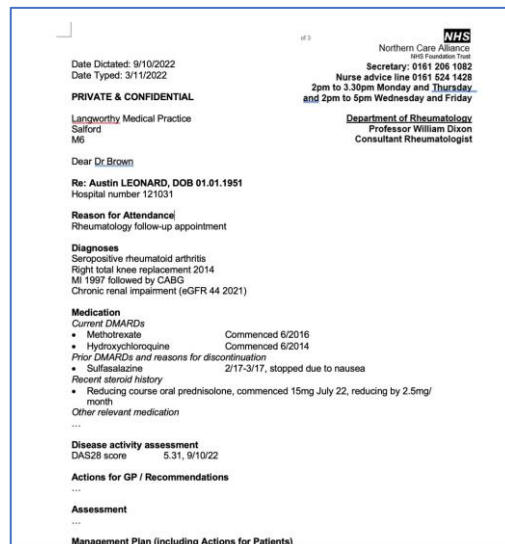
3. Maintain **public trust**, and for **sustainability and scalability** beyond the pilot



Data sources



Primary care EHR



Hospital letters



Inpatient EHR



Web-based surveys



Social care data

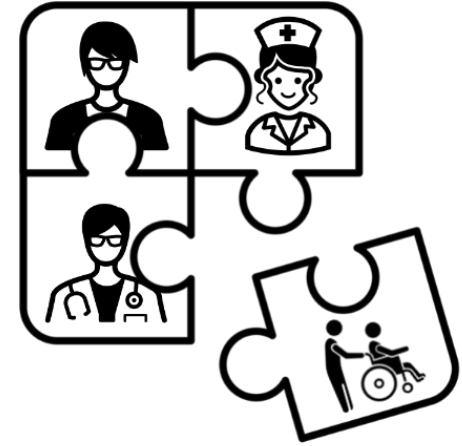


Interviews and focus groups

Outline

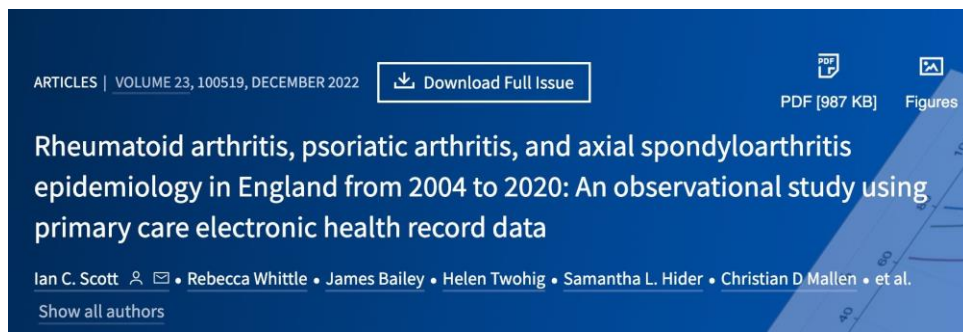
Focus on...

- Clinical question: Prevalence of disease
- Data sources: Primary care EHR (inaccurate codes)
Outpatient letters (unstructured data)
- Public trust: Public notification of research plans



Prevalence of disease

- First line of every funding bid
 - eg “Psoriatic arthritis affects one in 300 people”
- Needs to be up to date and accurate
- Measured by counting ‘number of people with disease’ out of the whole population
- Possible to do using primary care electronic health record databases



Prevalence of psoriatic arthritis in Salford

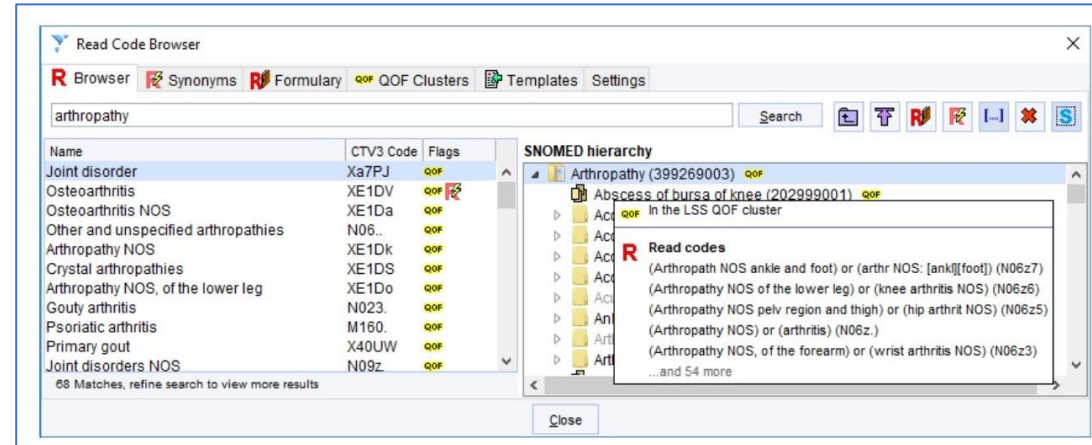
Data from Salford Integrated Record GP records, 2011-2019

- Number of patients with a code for PsA = 332
- Number of patients in primary care data = 188,286
- Prevalence = 0.18%

However...

- Not all patients with GP codes have the disease
- Not all patients with disease have a GP code
- Opportunity to understand mismatch using linked hospital data

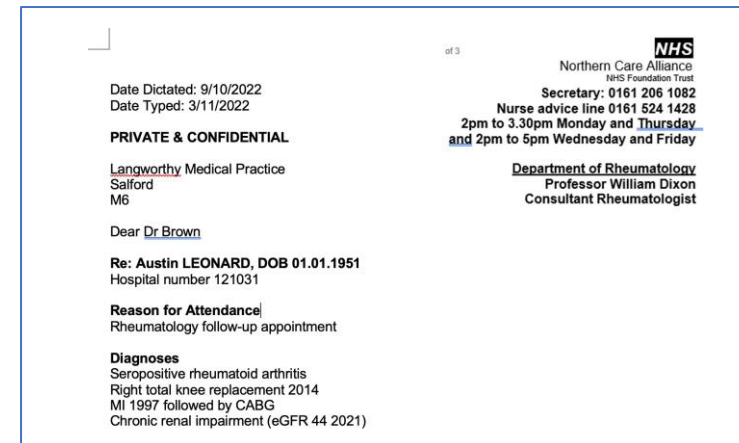
Primary vs secondary care data



Primary care

Read / Snomed codes

- Supports direct care, planning and research



Secondary care

Free text

- OK for direct care
- Not machine-readable
- Cannot count across populations

Natural language processing

St Elsewhere Hospital
Tel: 0161 001 1234

Patient name:
DOB:
Clinic Date:
Reason for attendance: Rheumatology follow-up

Diagnoses: Seropositive rheumatoid arthritis
Chronic renal impairment (eGFR 44)
MI 1997 followed by CABG
Anxiety and depression

Medications: Methotrexate
Sulfasalazine
Hydroxychloroquine

Mr Nenadic has had arthralgia for the last three years, diagnosed only six months ago as rheumatoid arthritis.

FREE-TEXT INPUT

Identify **diagnoses** not **procedures**

001	Seropositive rheumatoid arthritis
002	Chronic renal impairment (eGFR 44)
003	MI 1997 followed by CABG
004	Anxiety and depression

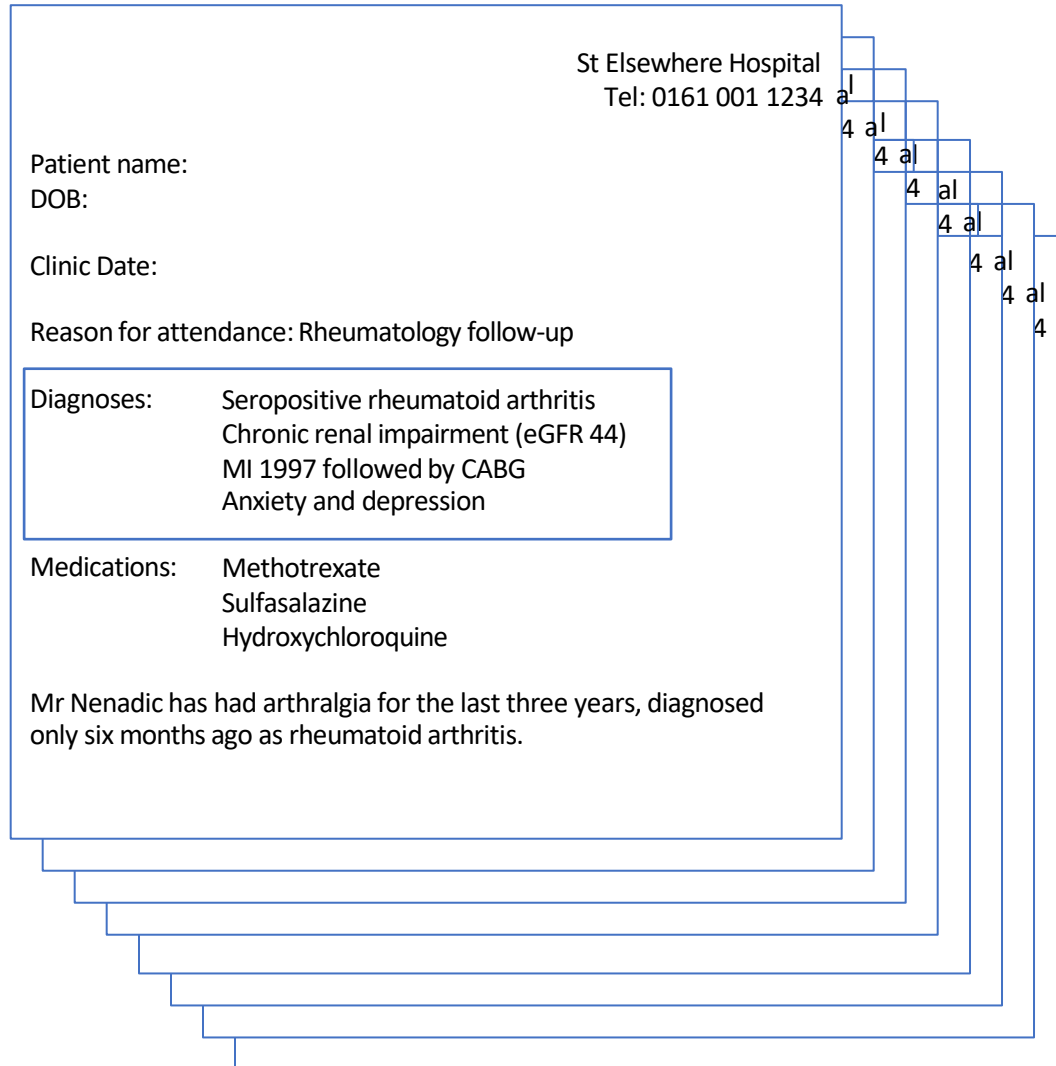
OUTPUT

Map **diagnoses** to Snomed terms

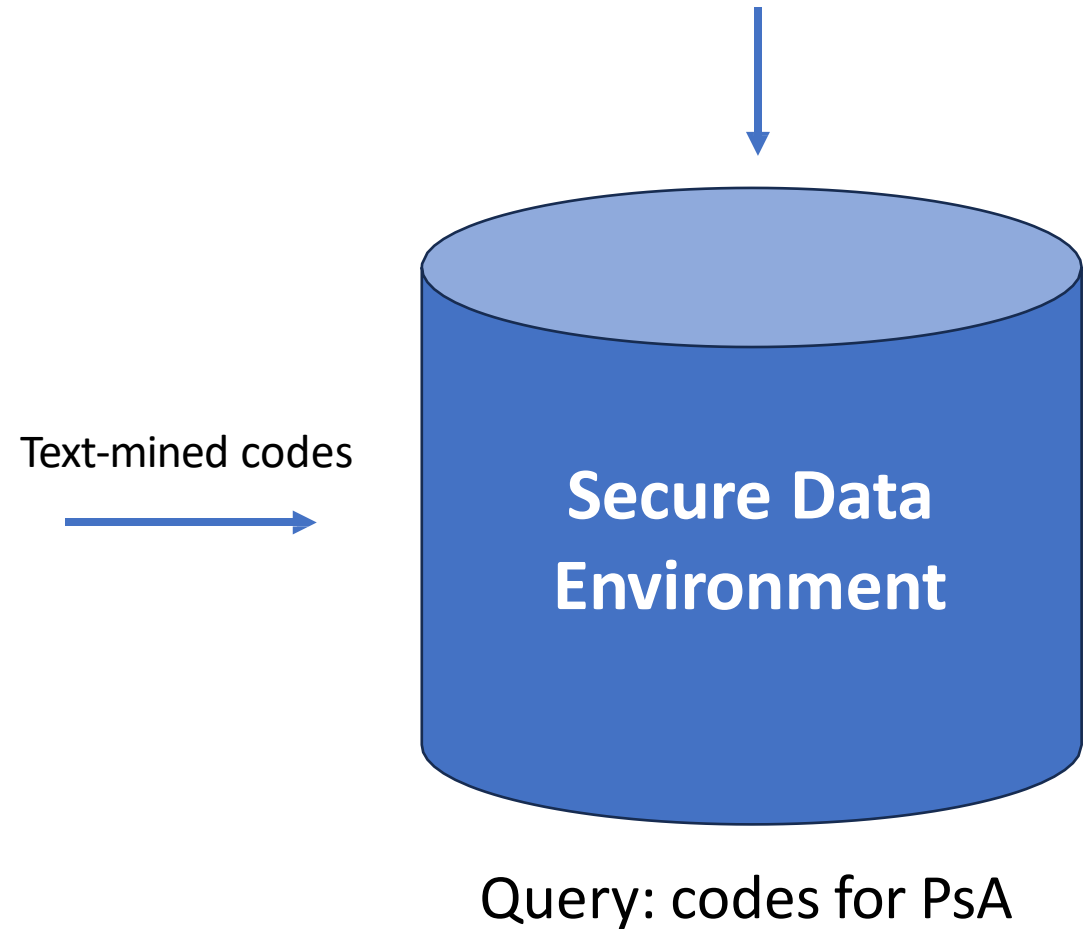
	Free-text term	Snomed ID
001	Seropositive rheumatoid arthritis	2967032
002	Chronic renal impairment	90688005
003	MI	22298006
004	Anxiety	197480006
	Depression	35489007

Unpublished data not shown

All rheumatology letters 2011-19



All primary care diagnoses 2011-19



What we want to know

Prevalence of disease

	Disease positive	Disease negative	
Test (codelist) positive	A	B	E
Test (codelist) negative	C	D	F
	G	H	N

What we had before

Prevalence of disease

	Disease positive	Disease negative	
Test (codelist) positive	A	B	E
Test (codelist) negative	C	D	F
	G	H	N

Unpublished data not shown

What linkage tells us

- Important opportunity to examine accuracy of national data in a subset where linkage exists
- Allows for 'correction factor' for national estimates
- Shows benefits of text-mining outpatient letters
- But... should improve the structured data foundation of outpatients

Assembling the Data Jigsaw

- A 'local data integration pilot' that

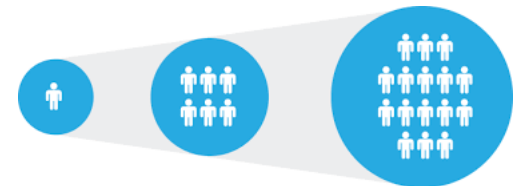
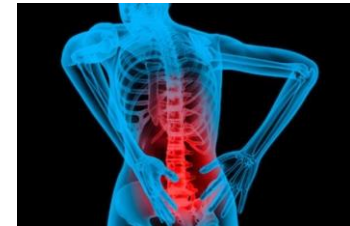
1. Pilots **novel data collection, extraction and integration** in rheumatology,

in order to...

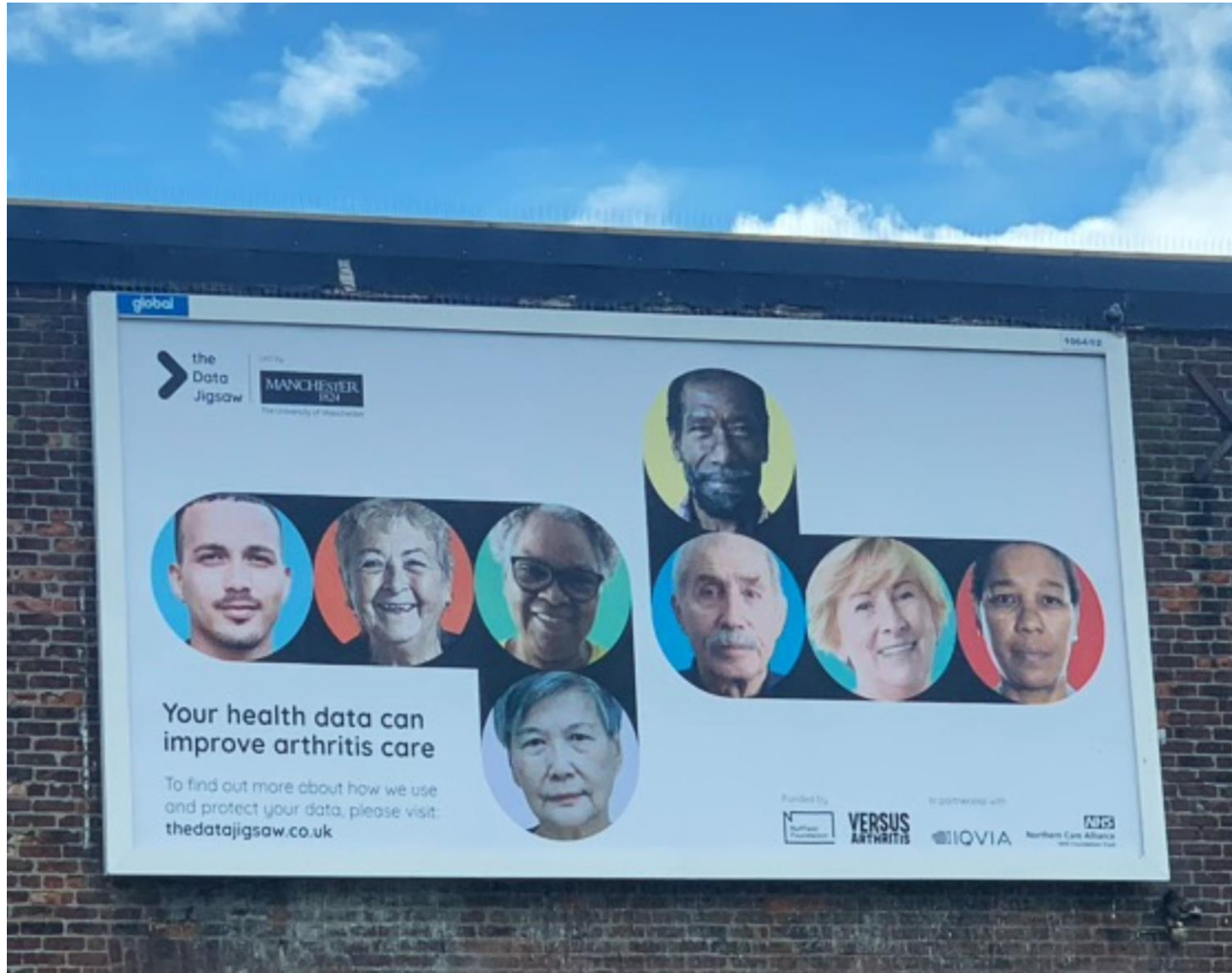
2. Answer **important research questions**,

and designed to...

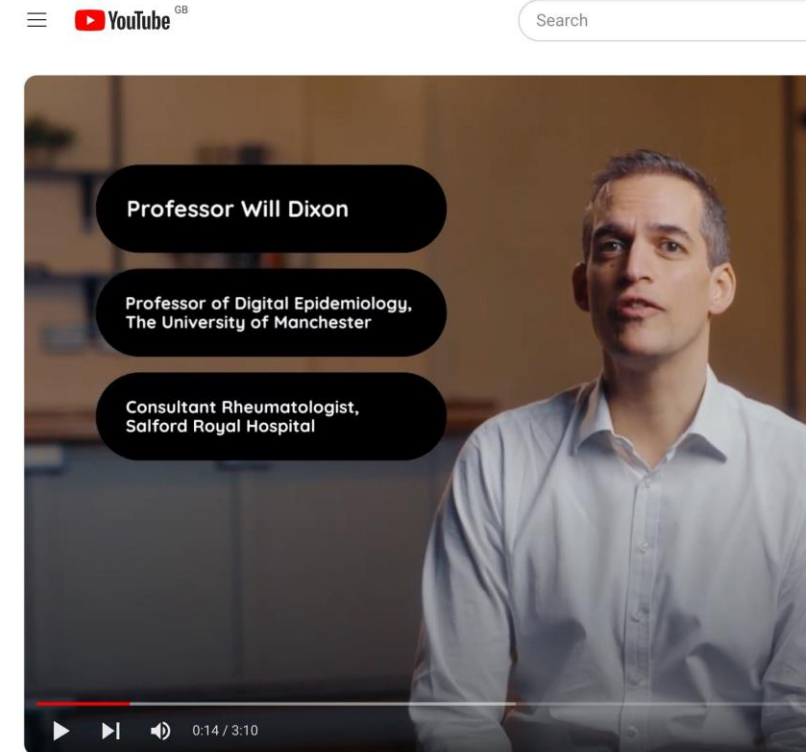
3. **Maintain public trust**, and for **sustainability and scalability** beyond the pilot



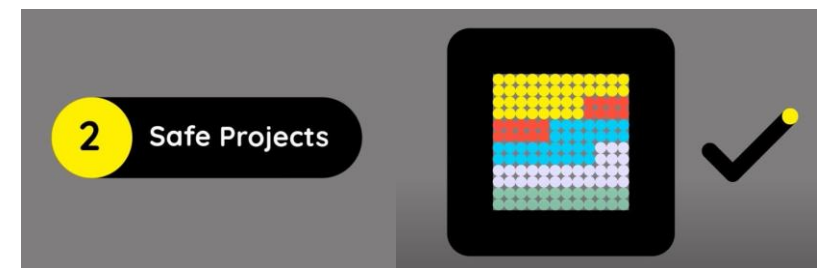
Public notification campaign



thedatajigsaw.co.uk



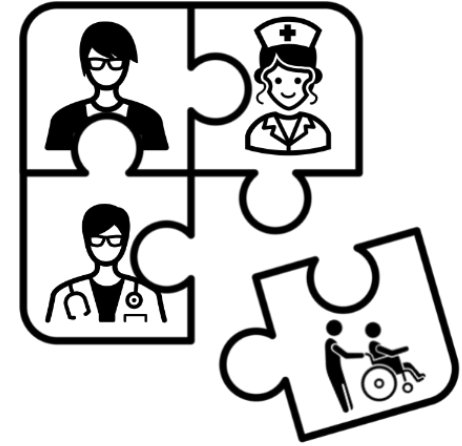
The Data Jigsaw: Secure Data Environments



Outline

Focus on...

- Clinical question: Prevalence of disease
- Data sources: Primary care EHR (inaccurate codes)
Outpatient letters (unstructured data)
- Public trust: Public notification of research plans



FUNDED BY

NIHR

National Institute for
Health and Care Research



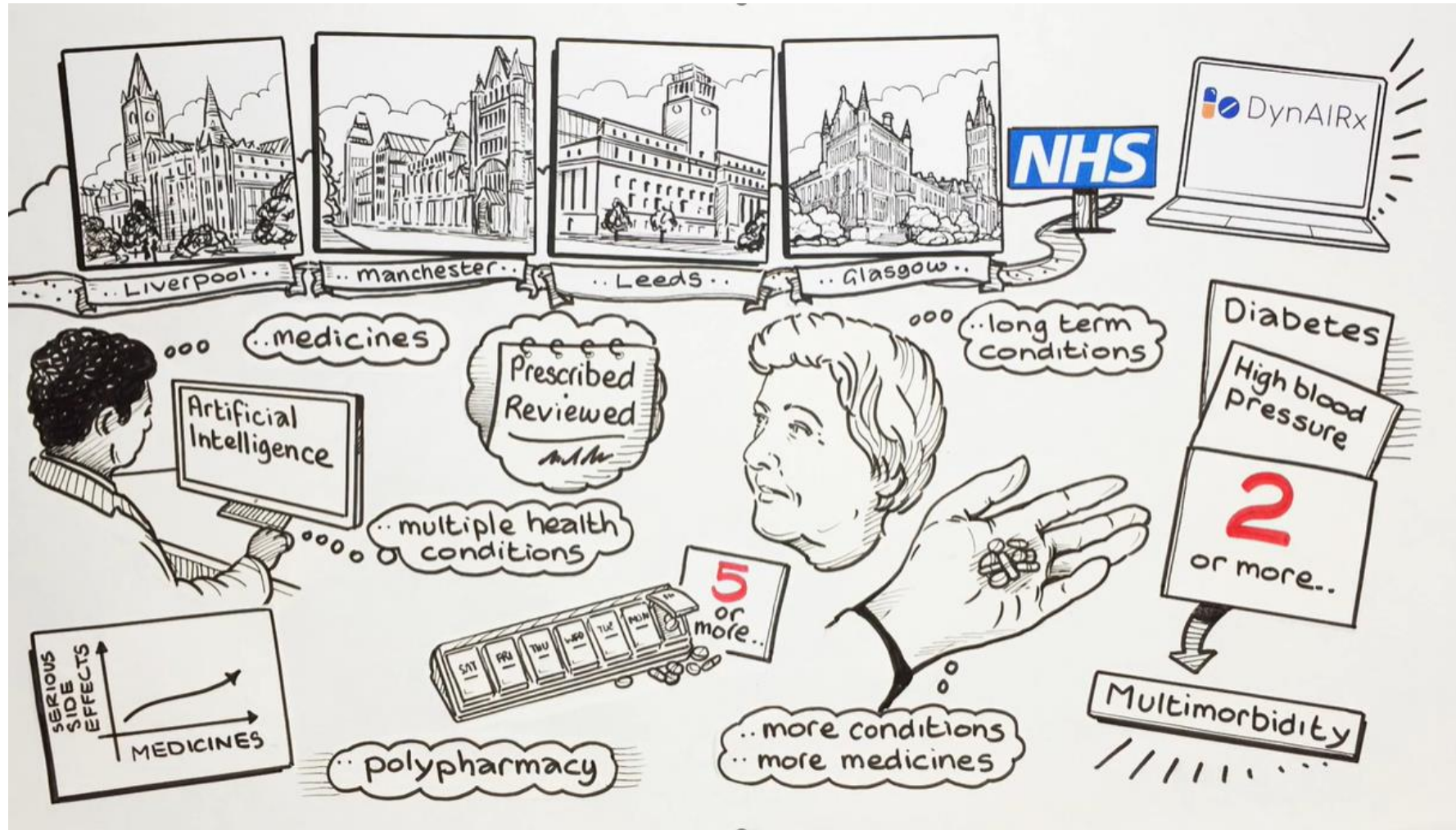
DynAIRx

Als for Dynamic Prescribing (Dyn-AI-Rx)

Transforming Medicines
Optimisation
in Multimorbidity Through Als

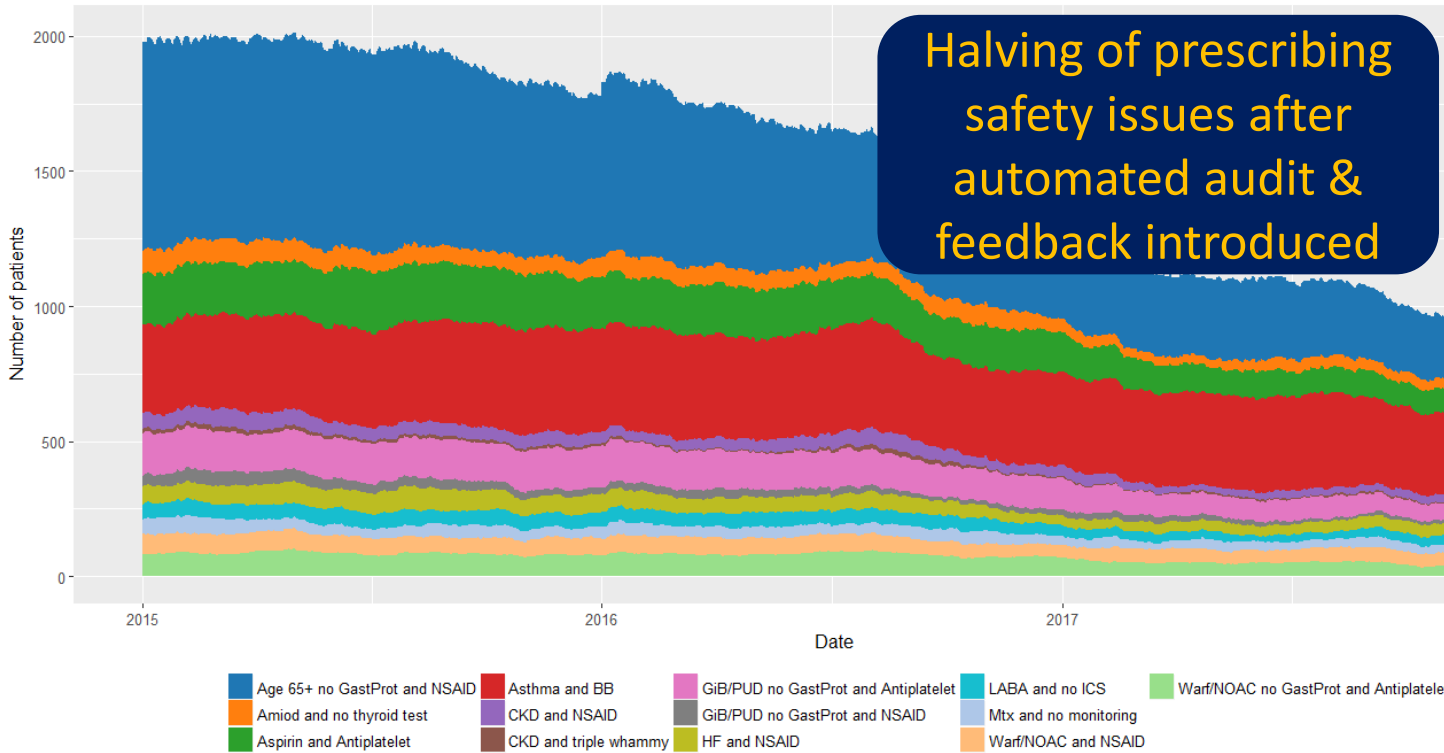
Dr Lauren Walker, co-PI
On behalf of the DynAIRx Investigators





Single condition focus of prescribing AIs to date

Number of patients at risk



<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003286>

Patient Safety Dashboard Users: Richard Williams

Single Practice / Glendale Medical Centre

Report date: 15 Nov (Latest) Comparison date: 16 Oct (30 days ago) Sort by: Affected patients

Practice summary Table Charts Export

Indicator	Affected patients	% of eligible patients affected	CCG Avg (%)	New cases	Trend	Show on top
Age ≥65 no GastProt and NSAID	19	2.04	0.32	3	1	<input type="checkbox"/>
Mtx and no monitoring	12	11.01	2.67	2	-3	<input type="checkbox"/>
GIB/PUD no GastProt and Antiplatelet	8	6.61	2.49	1	-1	<input type="checkbox"/>
Asthma and BB Click to view patients...	8	3.67	1.51	2	0	<input type="checkbox"/>
Aspirin and Antiplatelet	7	3.47	1.11	7	7	<input type="checkbox"/>
CKD and triple whammy	5	2.86	1.30	5	5	<input type="checkbox"/>
Warf/NOAC and NSAID	4	19.05	9.05	1	0	<input type="checkbox"/>
HF and NSAID	3	2.94	2.11	2	-2	<input type="checkbox"/>
LABA and no ICS	2	0.85	1.07	0	2	<input type="checkbox"/>
Amiod and no thyroid test	2	9.09	11.54	4	-3	<input type="checkbox"/>

SINGLE-CONDITION/HAZARD medication audit & feedback is working

MULTI-CONDITION/HAZARD medication reviews are required but not targeted well

Case



This is a fictional patient case included for illustrative purposes only

AVR, aortic valve replacement; **BD**, twice a day; **BP**, blood pressure; **eGFR**, estimated glomerular filtration rate; **HbA_{1c}**, haemoglobin A_{1c}; **HFpEF**, heart failure with preserved ejection fraction; **IV**, intravenous; **MR**, modified release; **OD**, once daily; **PO**, oral administration; **QDS**, four times a day; **TDS**, three times a day

- Tissue AVR
- Aortic regurgitation
- Atrial fibrillation
- HFpEF
- Falls
- Hypertension
- IV diuretics
- Refractory fluid overload
- ?end-of-life

Weight: 86 kg

BP: 132/60 mmHg

eGFR: 50 mL/min

Creatinine clearance: 82 mL/min

Creatinine: 120 µmol/L

Potassium: 5 mmol/L

HbA_{1c}: 50 mmol/mol

Multiple admissions for IV furosemide infusion

Retired from armed forces, previous marathon runner, lives with daughter

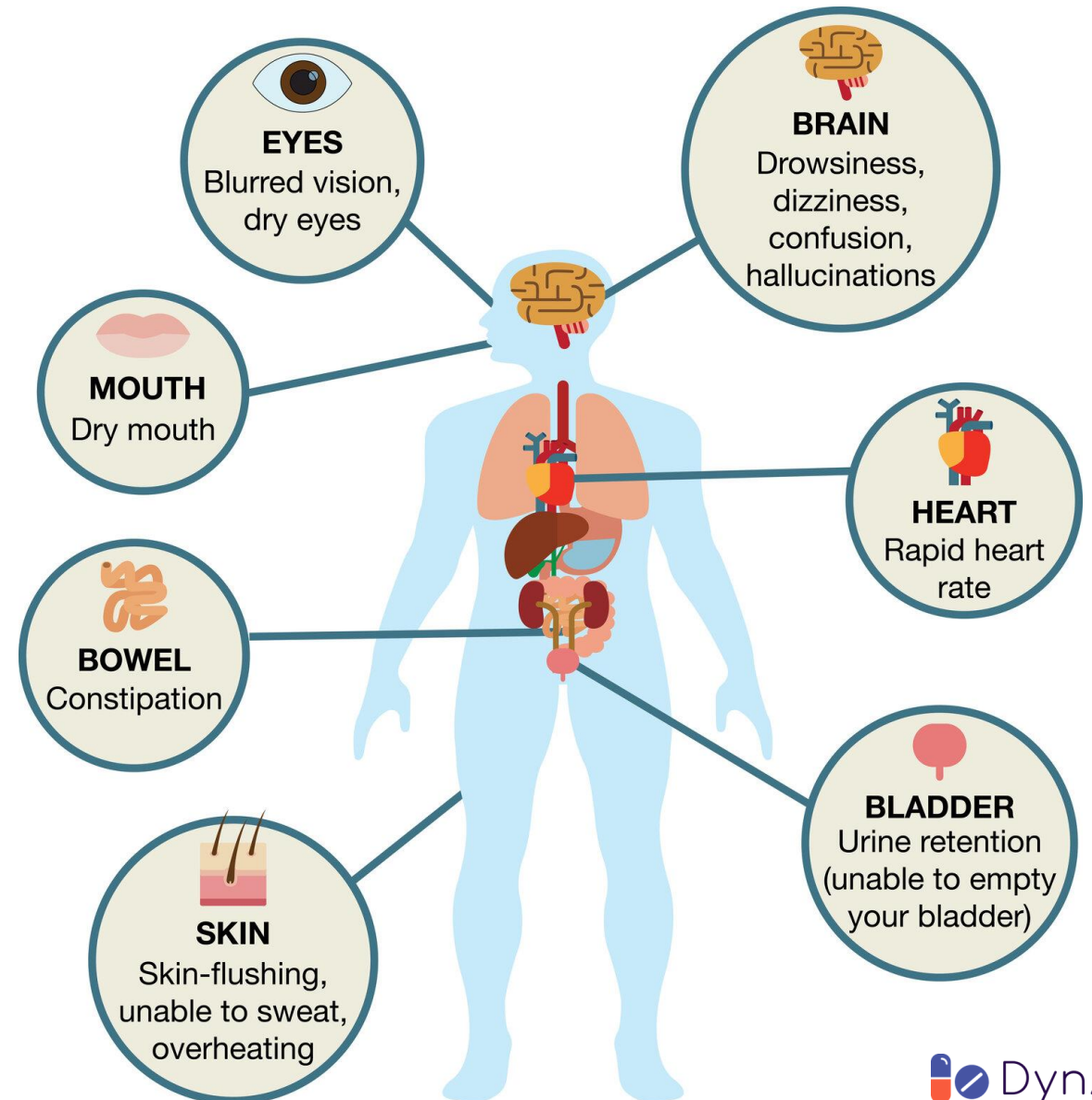
Medication:

- Bendroflumethiazide 2.5 mg OD
- Bumetanide 1 mg OD
- Dapagliflozin 10 mg OD
- Spironolactone 25 mg OD
- Movicol oral powder TDS
- Aspirin 75 mg OD
- Atorvastatin 80 mg OD
- Ferrous sulfate 200 mg tablets OD
- Lansoprazole 15 mg OD
- Bisoprolol 1.25 mg OD
- Mirtazapine 15 mg tablets nocte
- Longtec 10 mg MR BD
- Oxycodone 5 mg/5 mL PO, 5 mL 4-6 hours
- Apixaban 5 mg BD
- Tamsulosin 400 mcg OD
- Allopurinol 400 mg OD
- Colchicine 500 mcg QDS

Medications Treat and Cause Multiple Conditions

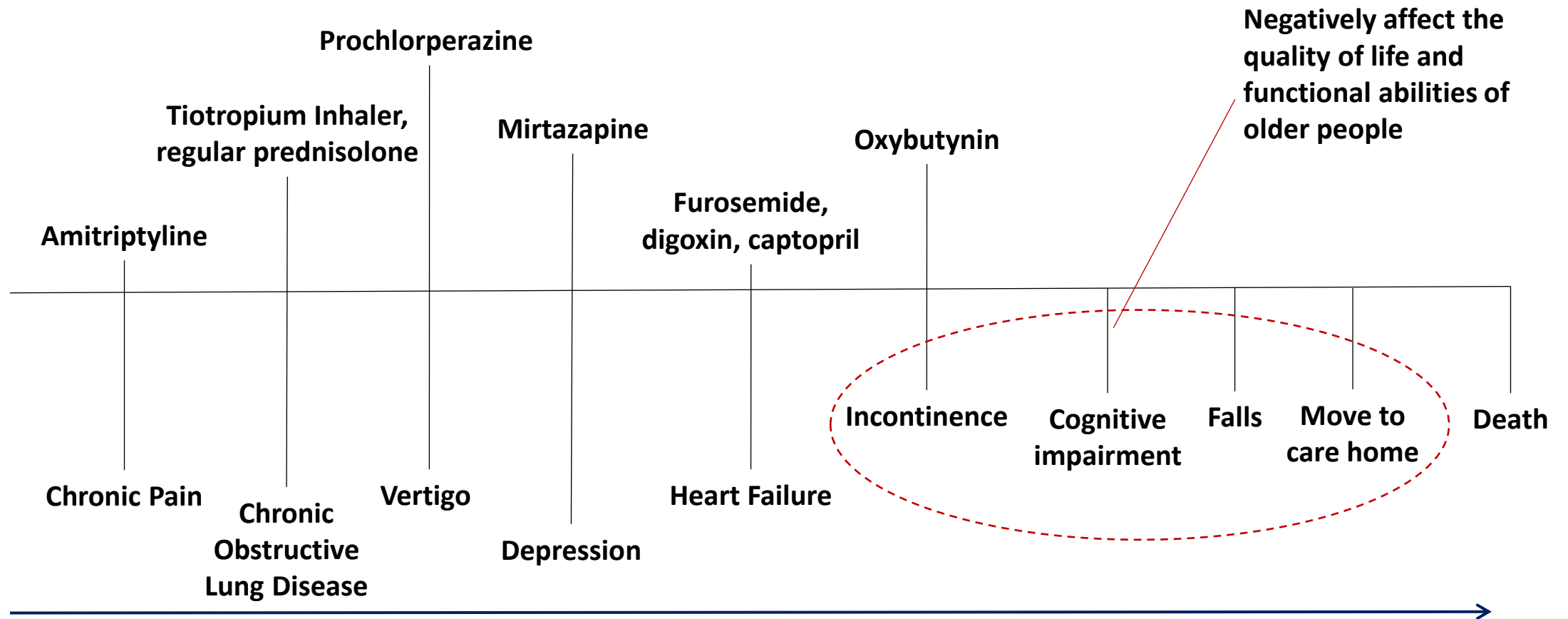
Example: Many common drugs prescribed for a wide range of reasons have **anticholinergic** side effects

- especially problematic for older people, but often overlooked by clinicians leading to harmful side effects such as delirium, falls and dementia



Hard Problem: Accumulation of Medicines Risks

Escalating anticholinergic burden over long periods of time not easily seen in records



age ↑; kidney function ↓; muscle mass ↓; unsteady; frail

WP1: WHAT ARE THE BARRIERS AND FACILITATORS TO THE UPTAKE AND UTILISATION OF AN AI-AUGMENTED PRESCRIBING SUPPORT SYSTEM FOR SMRS FROM THE PERSPECTIVE OF PRIMARY AND SECONDARY CARE CLINICIANS, PHARMACISTS, PATIENTS AND COMMISSIONS/MANAGERS INVOLVED IN SMR SERVICES?

Data collected	Number completed	Types of participants
Focus groups	8	<ul style="list-style-type: none"> • 2 x GPs (9) • 2 x Pharmacists (10) • 2 x Patient focus groups (MH and young MM, frailty outstanding) • 1 x Clinical pharmacologists (3) • 1 x Psychiatrists (3)
Semi-structured interviews	5	<ul style="list-style-type: none"> • 2 x Pharmacists • 1 x GP • 1 x Policy-maker • 1 x Psychiatrist
Data collected	Coding completed	



PRELIMINARY FINDINGS OF FOCUS GROUPS/INTERVIEWS

Five preliminary themes

1. Medication reviews in practice –TIME taken to *prepare* for SMR, pharmacists planned vs GP opportunistic
2. Medication-related challenges (mental health particularly, Ach)
3. Experience with digital health tools
4. Design ideas for DynAIRx (timelines)
5. Facilitators and barriers to implementation and adoption

“These were complex people. They weren’t quick, easy medication reviews. They took hours sometimes and sometimes they took several visits to get something quite right.” –Pharmacist

“The blood results are on a different system...And not all of them pull through necessarily, so it can be hit and miss sometimes as well....” –Psychiatrist

“Sometimes you can’t quite work out what medications people have been on. I mean if we talk about SSRIs they may have tried multiple different ones in the past and sometimes its difficult to work out what they’ve been on without having to go through the long, long list searching all the different medications that are SSRIs that they’ve tried.” –GP

PROMPTS

Long-term medicines no longer needed; medicines that disrupt sleep; missing medicines based on NICE guidelines; monitoring/due a review; new recommendations (e.g. from MHRA)

VISUALS

Pictorial, easy-to-digest info; dashboard of recent bloods; graphs for narrow therapeutic range; BNF interactions/matrix of polypharmacy interactions; pictorial NNT for deprescribing; physical observations; patient interface

TIMELINES

Medication; diagnosis; deprescribing; social prescribing; divisions by BNF chapter; links with letters; changes over time; who diagnosed

INTEGRATION INTO DIFFERENT HEALTHCARE SETTINGS

Accessible to secondary care; linking with community pharmacy to check what medicines/brands available

DESIGN

Minimally intrusive; user-friendly; efficient; not error prone; does not slow system; not cluttered; intuitive; accessible; avoids window modality; reduced pop-up fatigue

OUTCOMES

Evidence-based outcomes at practice level; evidence of increased work process efficiency; outcome data on big data level for medication reviews; outcome data in relation to demographic and socio-economic status

TASK COMPLETIONS

Recommendations adhered to; avoid round tripping (lack of follow-up leading to potential harm)

CALCULATORS

Anticholinergic burden; Serotenergic burden; opioid conversion tool; eGFR; risk calculators

QUICK ACCESS TO RESOURCES

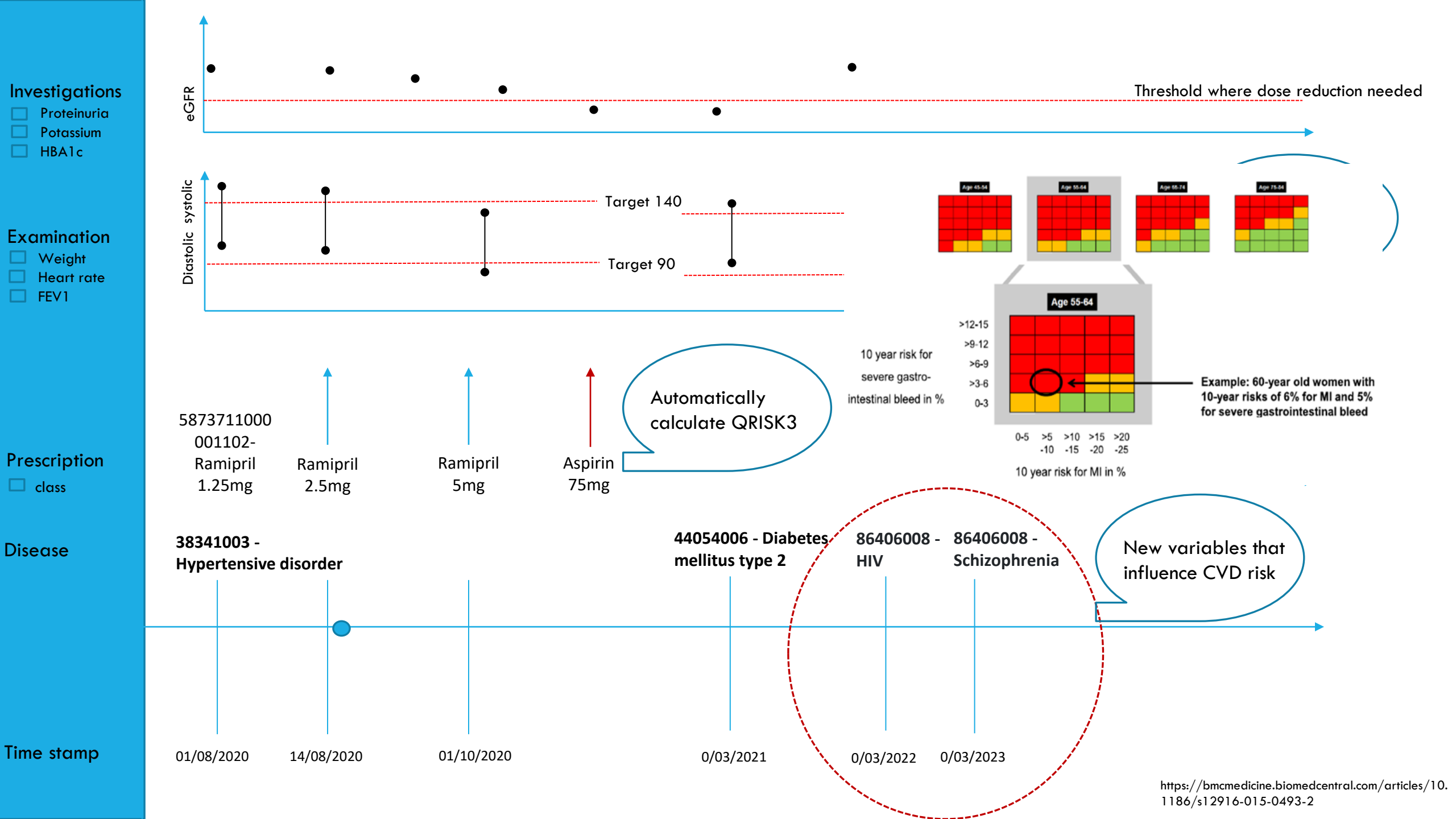
Relevant letters alongside diagnosis; BNF pages for relevant medicines; previous investigation results; prescribing guidelines

LINKAGES USING AI

Integrating QoL when optimising medicines for polypharmacy patients; identify prescribing gaps; soft data to identify potential risks (e.g. safeguarding); patient suspecting possible ADR; linking side-effects to drug list or disease; population level data during clinical decision-making

RISK PREDICTION MODELS

Real-time risk calculators; falls/delirium; diabetes; cardiovascular



FUNDED BY

NIHR

National Institute for
Health and Care Research



DynAIRx

Thank you!



Deep dive: exploring data platform and research initiatives

Prof Tim Hubbard

Associate Director
HDR UK London

Prof William Dixon

Chair in Digital Epidemiology,
Division of Musculoskeletal &
Dermatological Science
University of Manchester

Chair: Natalie Chishick

Director
Silver Buck

Dr Lauren Walker

Senior Clinical Lecturer in
Clinical Pharmacology &
Therapeutics and Internal
Medicine
University of Liverpool