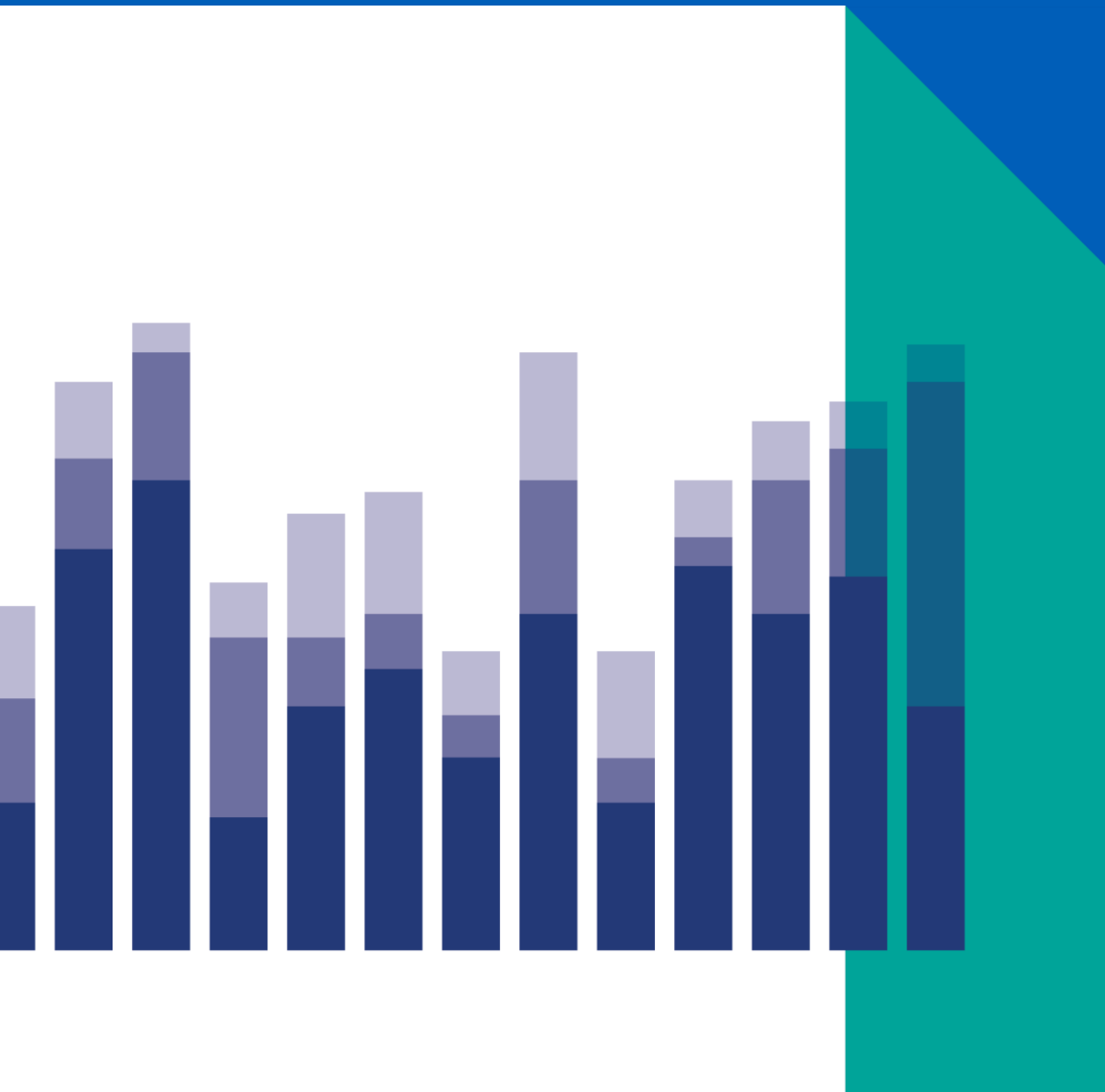


Using data in the South West SDE for research





Eilis Hannon

Associate Professor in
Bioinformatics,
University of Exeter

Agenda

Introduction – how to design a research project using SDE data

Prof Eilis Hannon, Associate Professor in Bioinformatics, University of Exeter

How do we move from site-specific data to standardised infrastructure for research?

Dr Michelle Leadbetter, Senior Research Data Engineer, South West SDE

Data in the NHS – electronic patient records

Dr Nick Kennedy, Chief Research Information Officer, Royal Devon and Exeter Hospital

Developing a data specification

Kane O'Beirne, Head of Operations, South West SDE

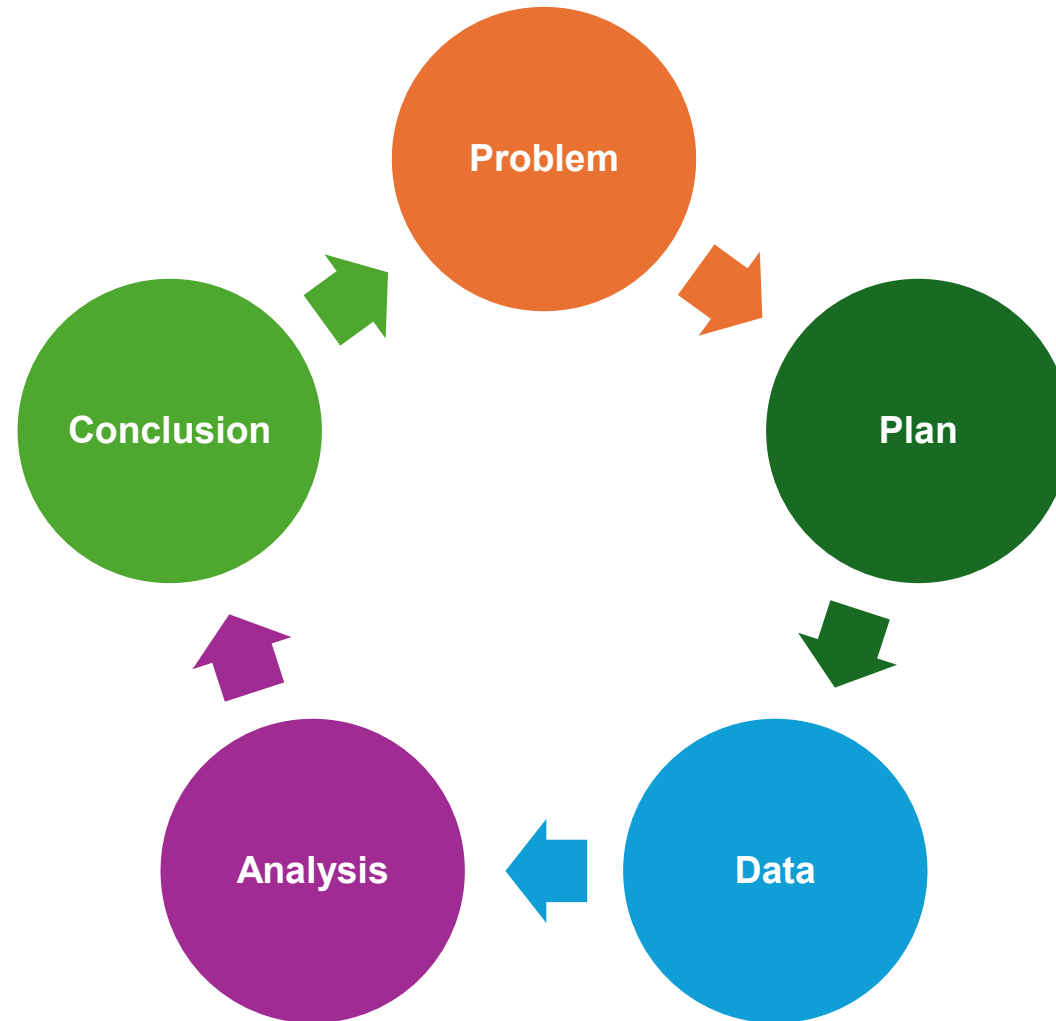
Housekeeping

Please use the Q&A function to ask questions

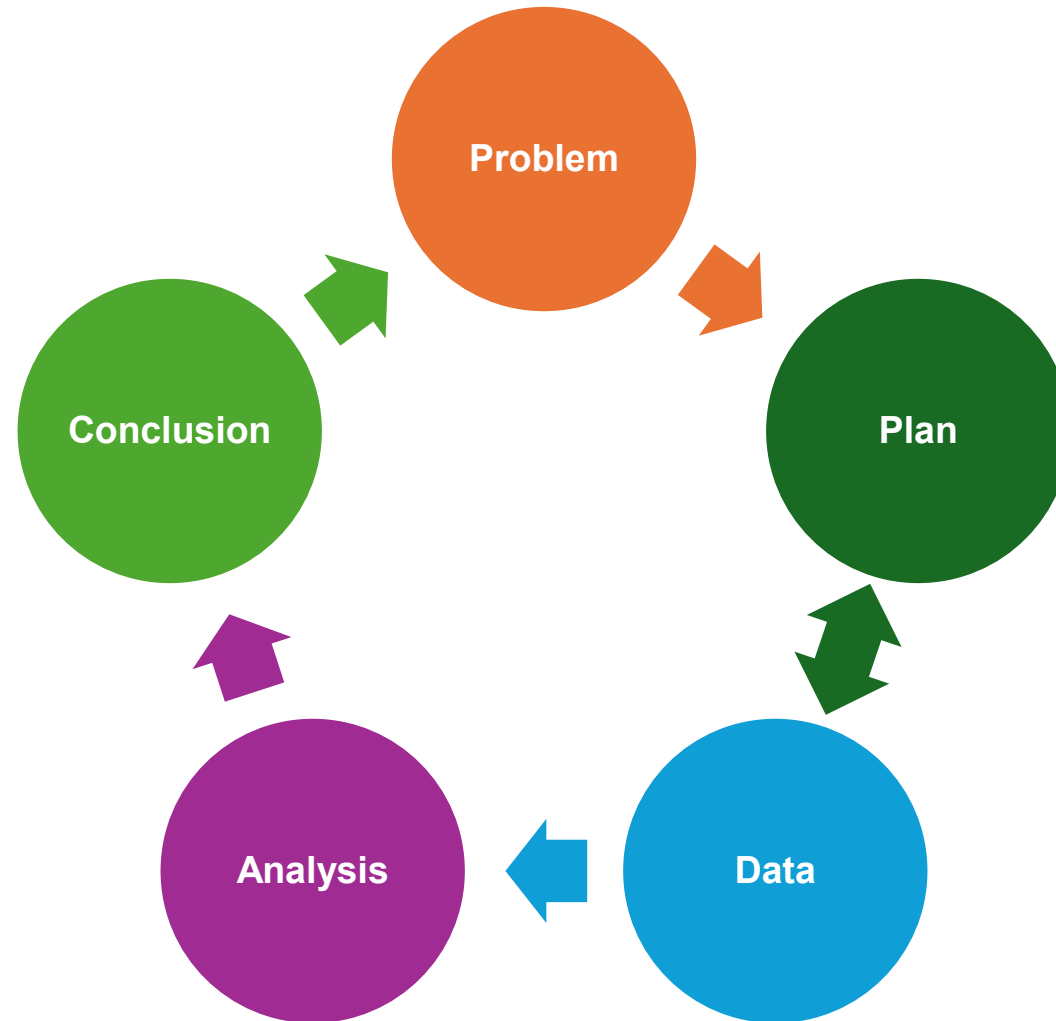
You can upvote other people's questions

How to design a research project using SDE data

Designing research studies: the PPDAC cycle



Designing research studies: the PPDAC cycle



Defining your data upfront



SDE approach

Define question

Specify variables

Justify need

Get access

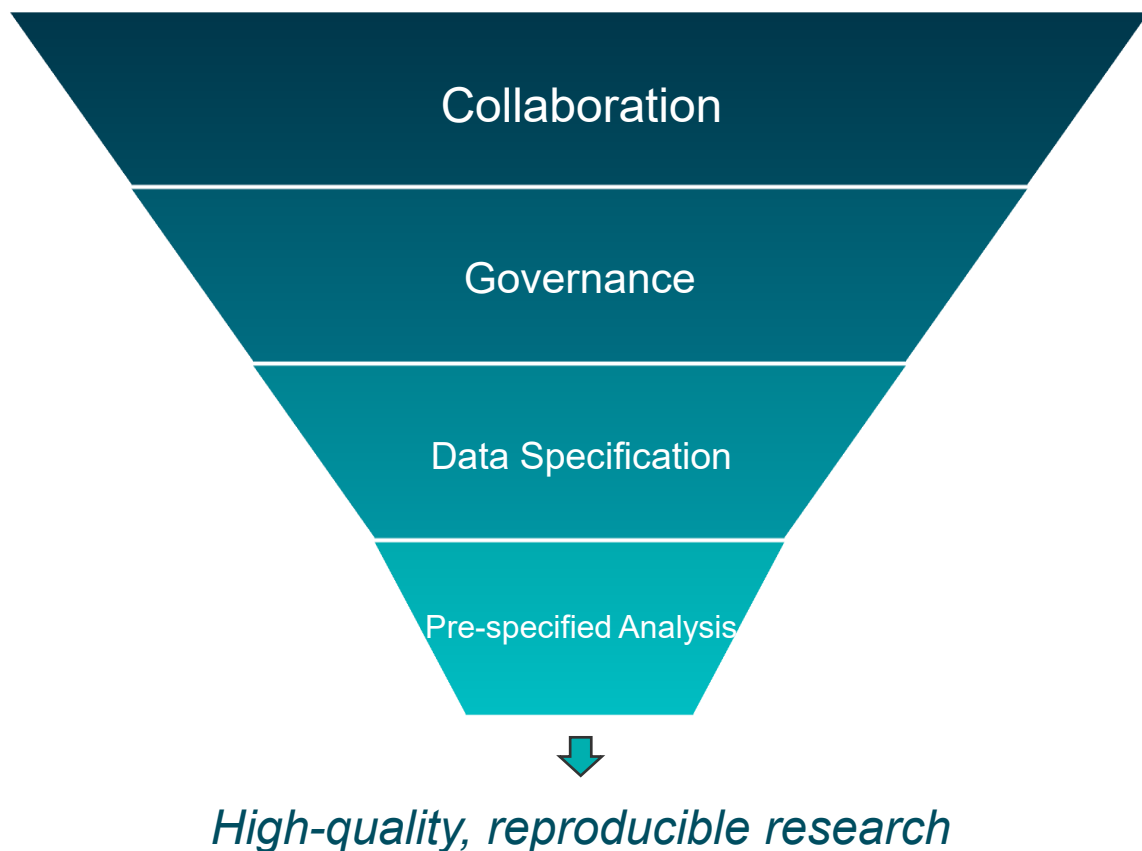
Collaboration for understanding the data

Exploratory analysis is now a conversation



- Engage clinical partners early: they are collaborators, not data collectors
- Understand what the data is before you see it
- The SDE central team can help you find the right people
- Be open to challenge: it leads to better projects

Better research by design



- Interdisciplinary by nature: the system fills your knowledge gaps
- High barrier to entry filters for commitment and quality
- Upfront decisions lower the temptation to move the goalposts
- Aligns with best practice in reproducible research

Open research and reusability

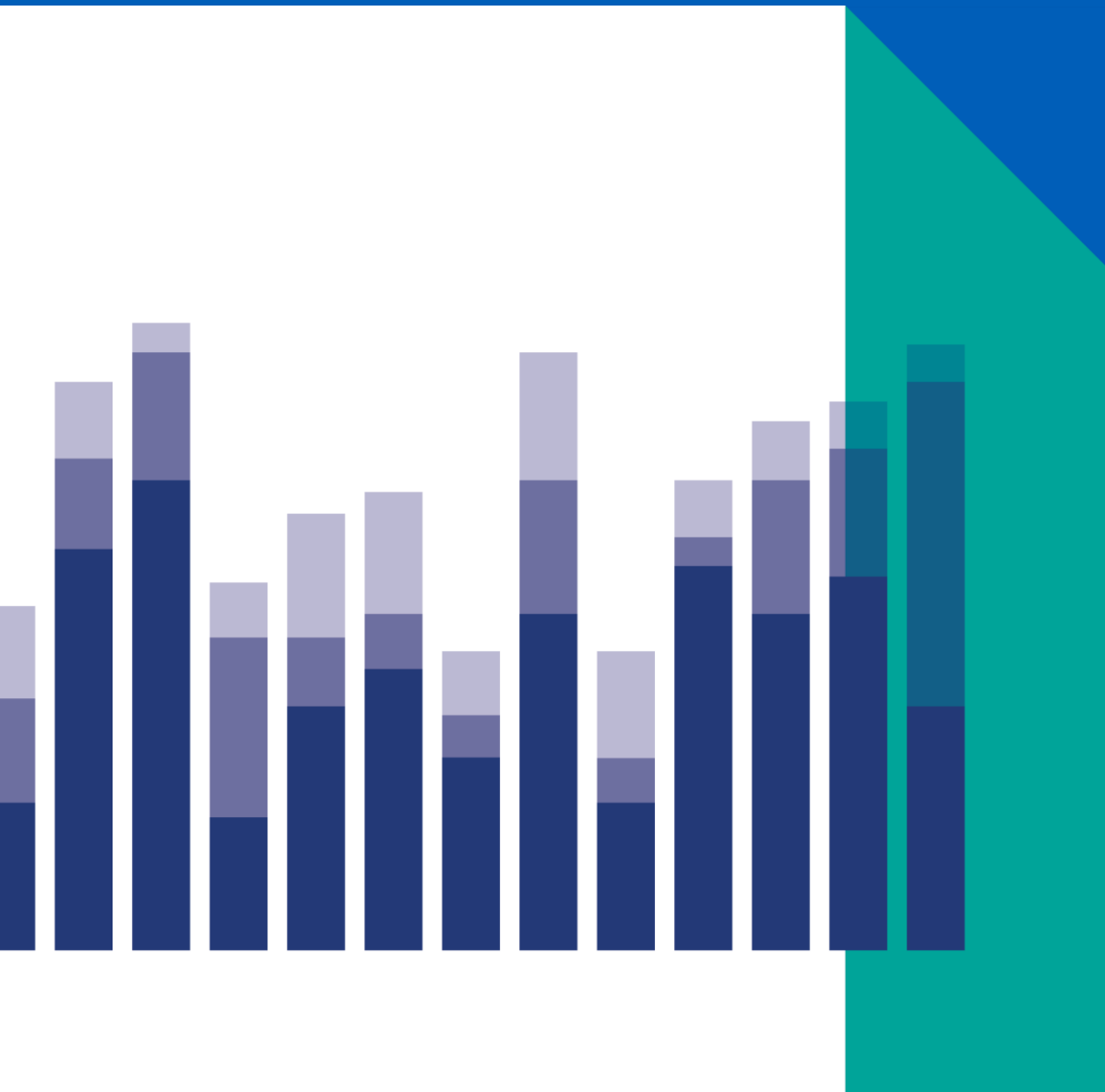


- Code can be shared and reused across projects
- Transparent methods, reproducible pipelines
- Phenotype definitions and data extracts become shared assets
- Your work feeds a growing ecosystem

From patient data back to patient care



- Real-world clinical data = direct relevance to patient care
- Research questions grounded in clinical need
- Results feed back into NHS practice and policy
- A clear line from your research to public benefit



Michelle Ledbetter

Senior Research Data Engineer,
South West SDE

Using data in the South West SDE for research

How do we move from site-specific data to standardised infrastructure for research?

- developing established data pipelines
- standardisation (i.e. OMOP)

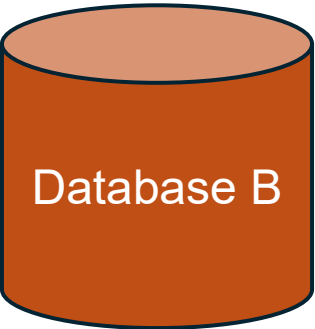
Why do we need research data pipelines?

- Preparing and transforming data can be a time-consuming part of research.
- We want researchers to focus on answering research questions, not repeatedly preparing data.
- We want data transformation to be reproducible, consistent and scalable.

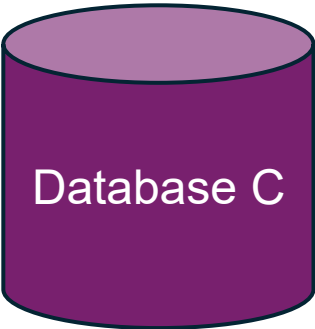
Challenge: data is inherently bespoke



patient id	date_time	measurement	value
1	2026-09-01	Blood Pressure	120/80
2	2026-09-01	Blood Pressure	135/85
3	2026-09-02	Blood Pressure	110/70
4	2026-09-02	Blood Pressure	128/82
5	2026-09-02	Blood Pressure	140/90

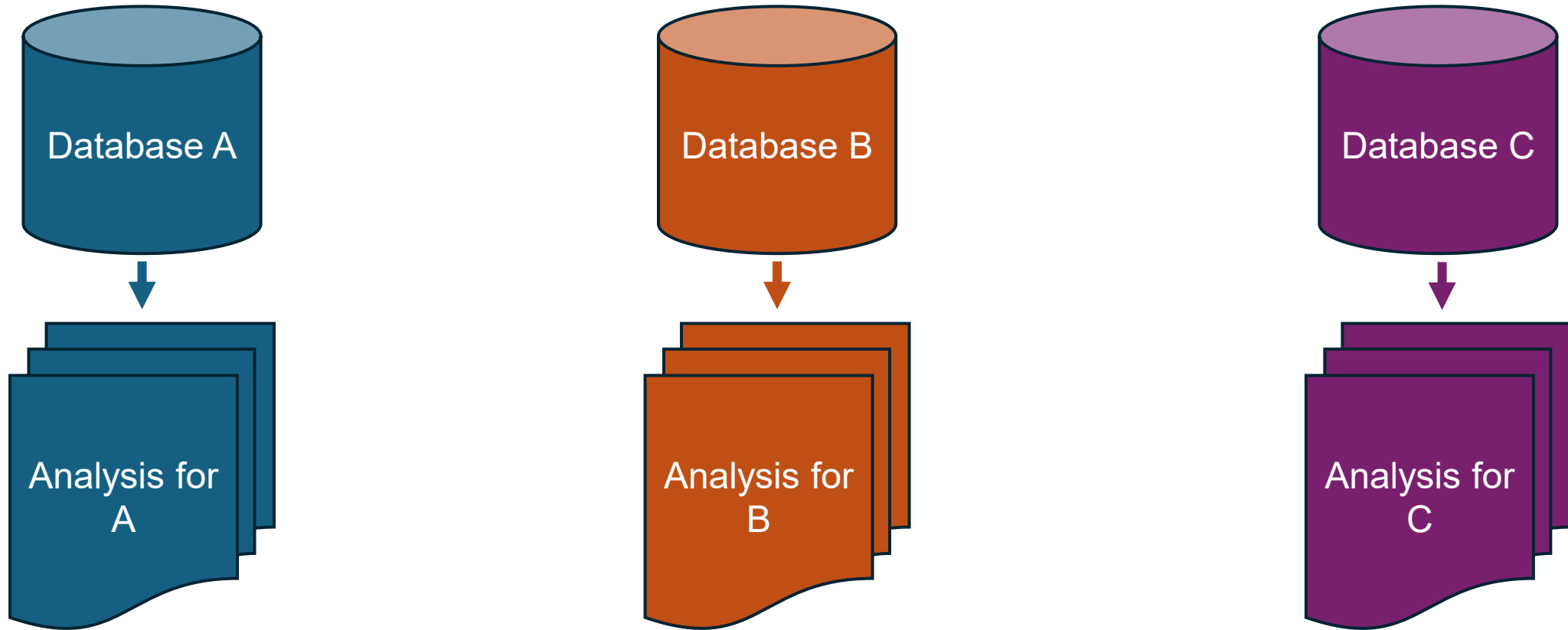


Patient number	Date of measurement	BP value
1	2026-09-01	120/80
2	2026-09-01	135/85
3	2026-09-02	110/70
4	2026-09-02	128/82
5	2026-09-02	140/90



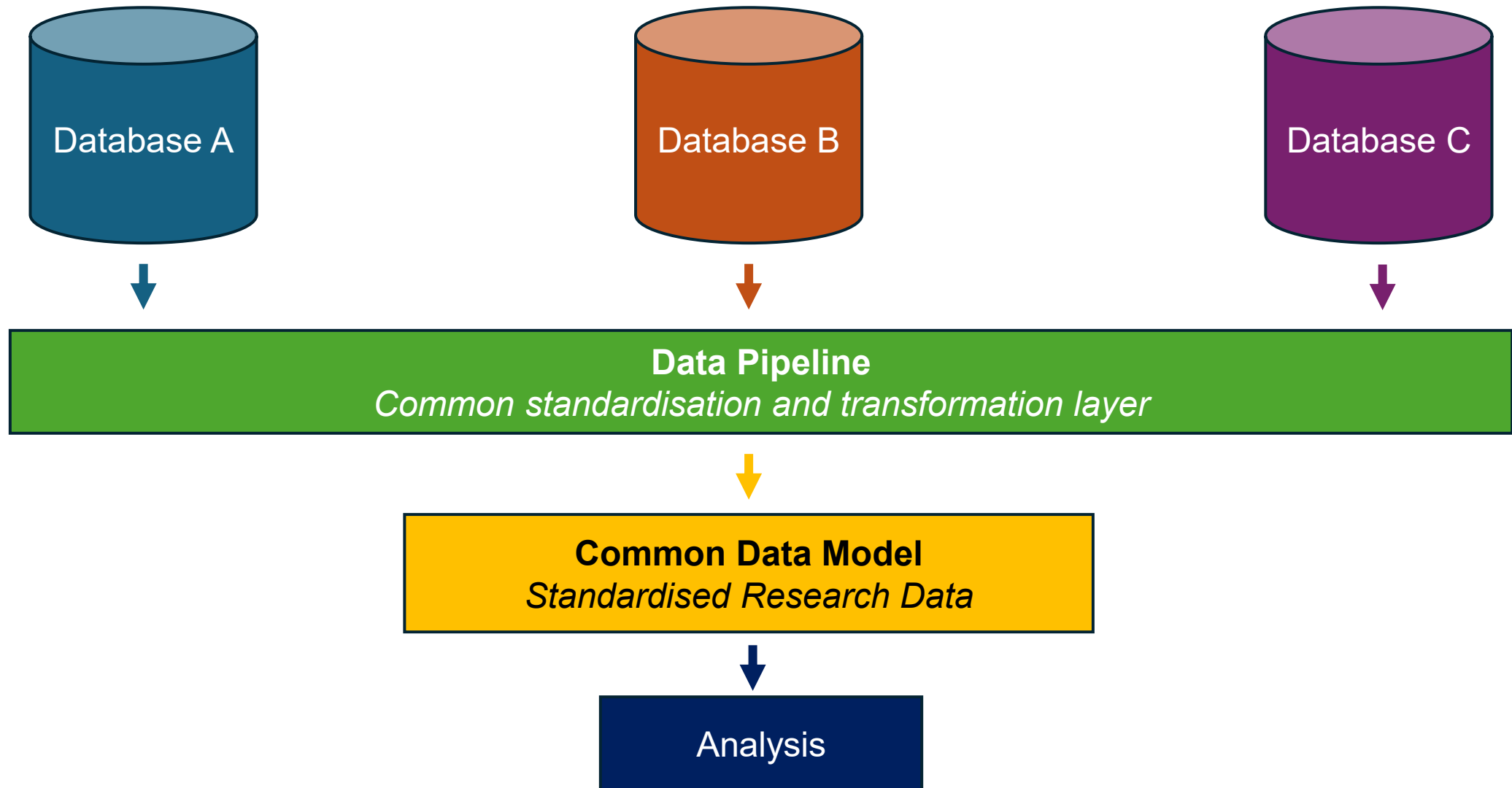
Unique identifier	date	Description	Result
1	2026-09-01	Systolic	120
1	2026-09-01	Diastolic	80
2	2026-09-01	Systolic	135
2	2026-09-01	Diastolic	85
3	2026-09-02	Systolic	110

Consequence: analysis becomes bespoke too

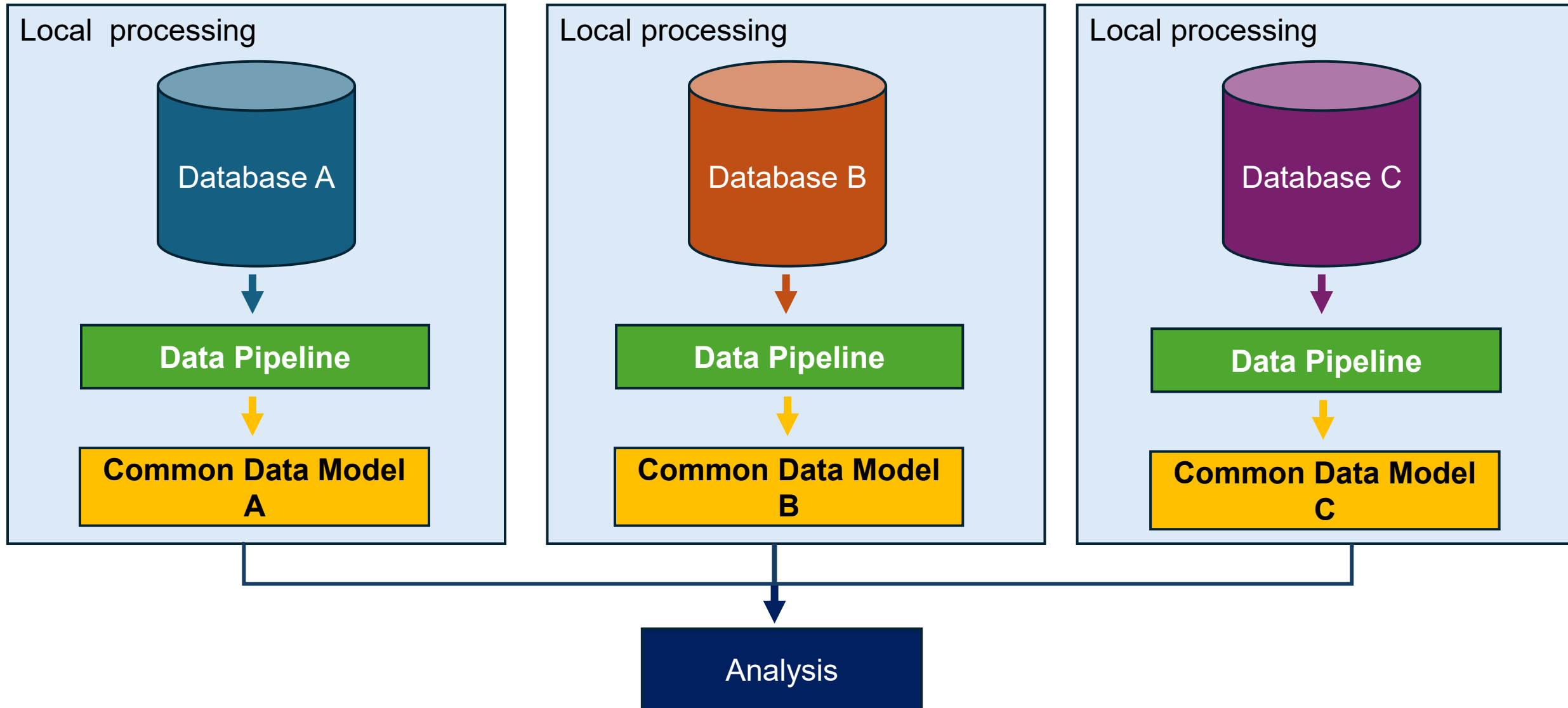


Same research question needs to be solved multiple times

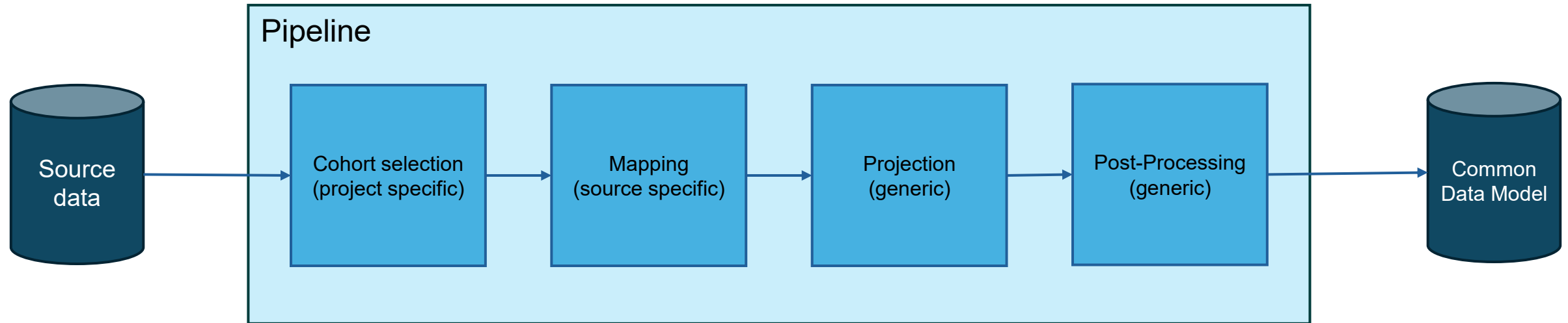
Creating reproducible data pipelines



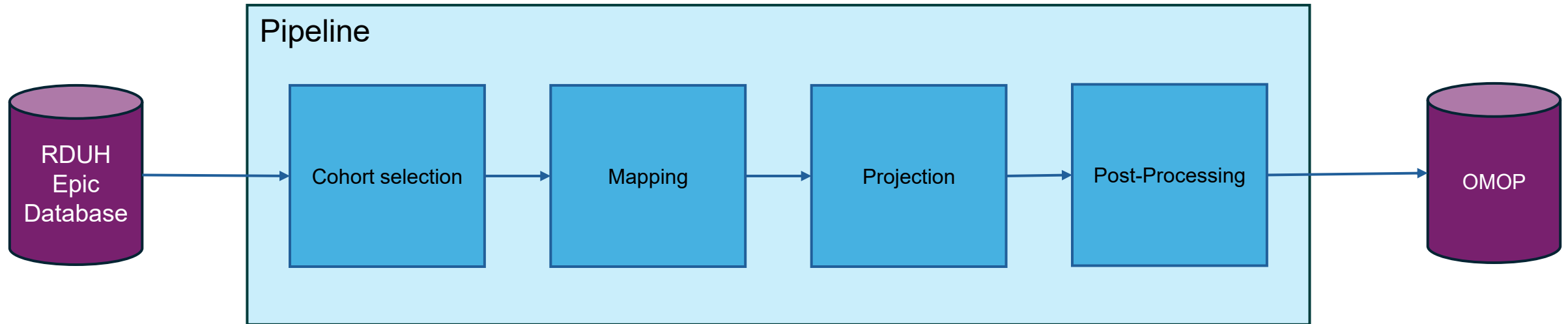
Different data architectures: federated analysis



The stages of a data pipeline



OMOP-ES pipeline

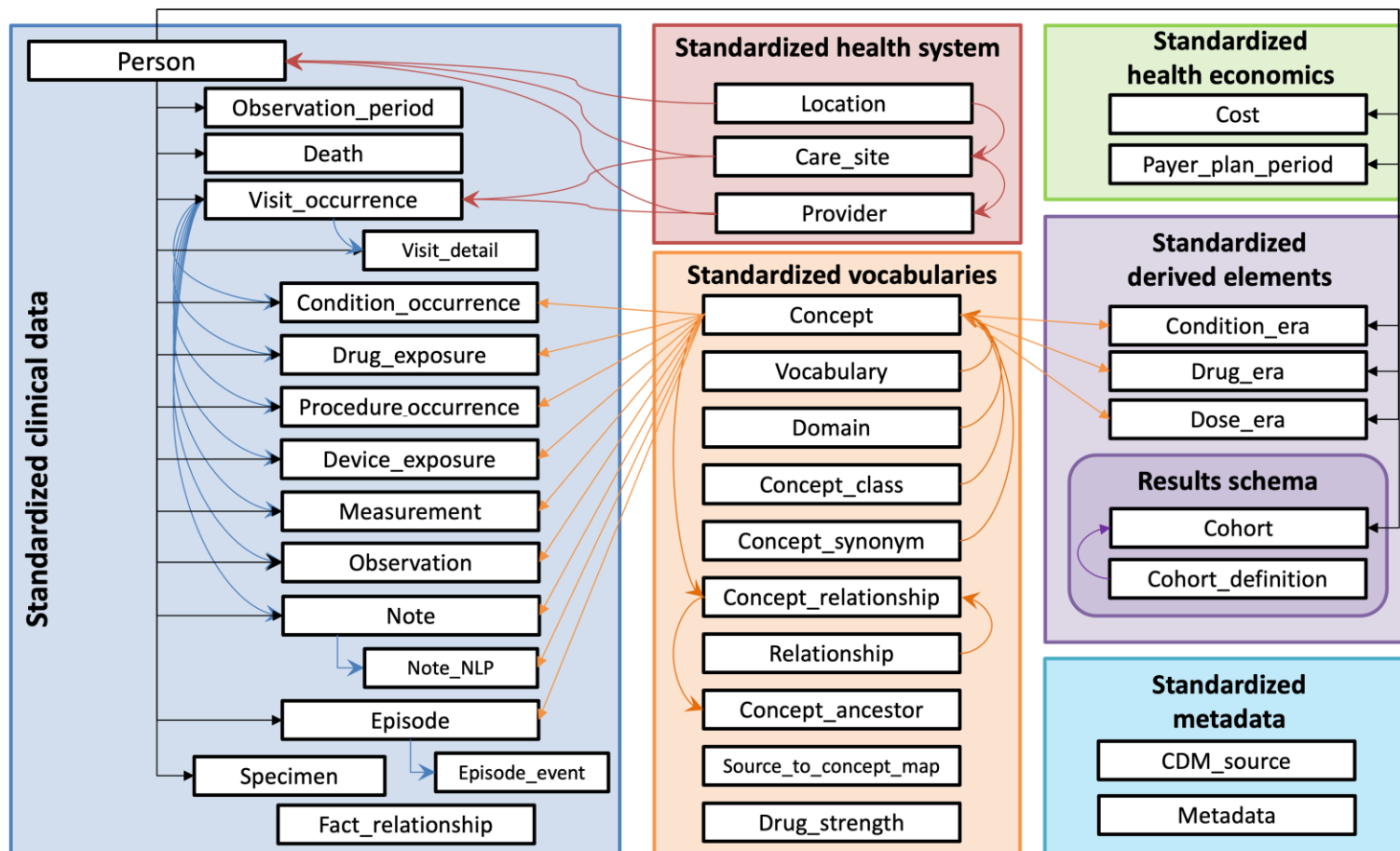


- The data pipeline 'OMOP-ES' is developed by UCLH and adapted by us for use at RDUH

- Used on Cardiovascular Frailty and HELP Flag (cancer risk project)

OMOP CDM

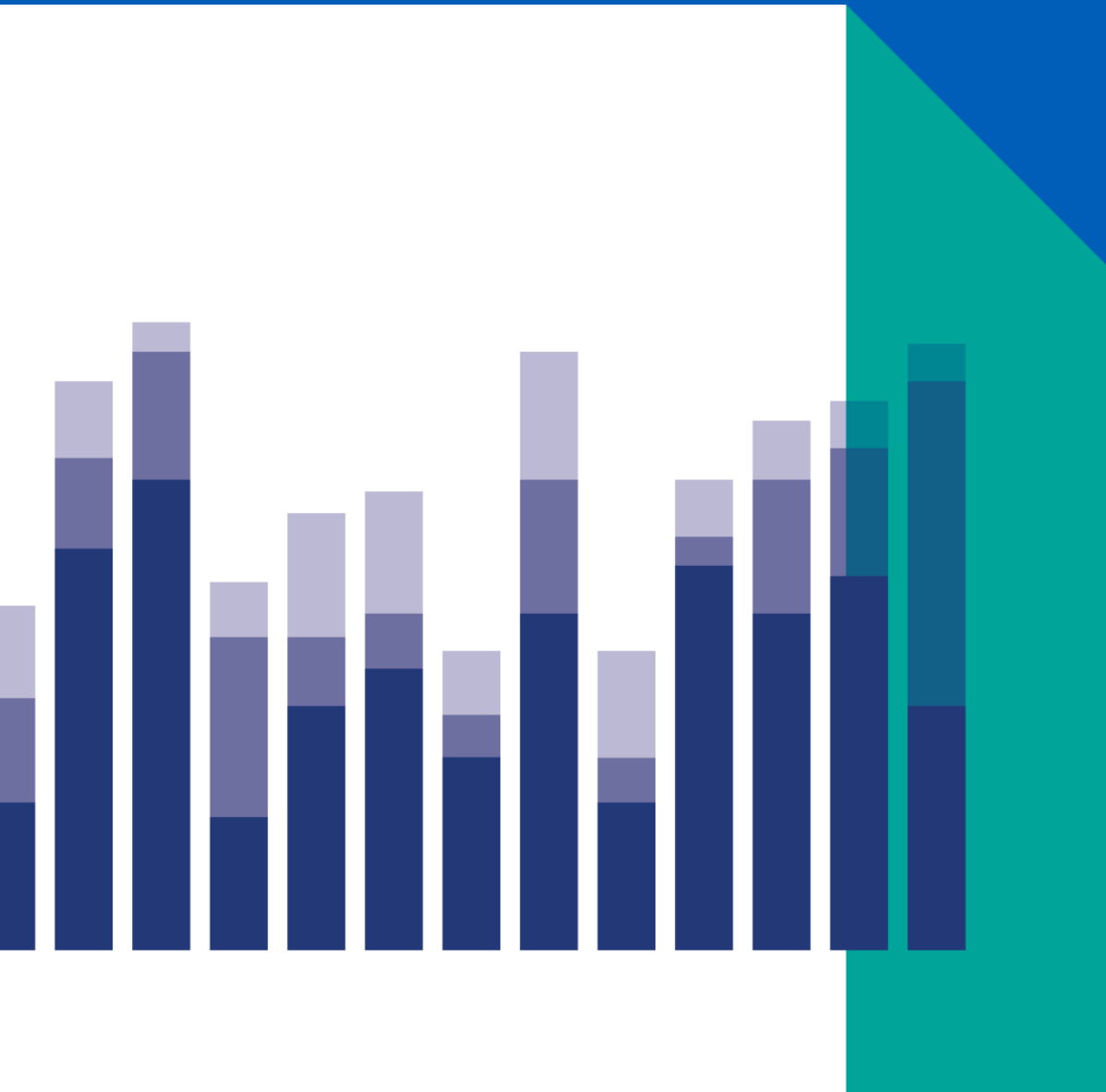
- **O**bservational **M**edical **O**utcomes **P**artnership **C**ommon **D**ata **M**odel
- Designed to standardize the structure and content of observational data



- OMOP facilitates the integration of data from different sources, systems and locations to be combined into one consistent format

From bespoke data to reusable research

- Handle complex and inconsistent data
- Reduce duplication of effort
- Improve quality and consistency
- Make research more reproducible
- Enable research across multiple sites



From one EPR to a regional research asset

Dr Nick Kennedy

Chief Research Information Officer,
Royal Devon and Exeter Hospital

The scale is already substantial

Caboodle data through August 2026

Scale of clinical data held in Epic				
Caboodle data through August 2026				
Domain	Total records	Records in last 12 months	Records in latest month	Patients in latest month
Flowsheet observations	1,929,761,134	609,211,941	37,837,974	57,976
Laboratory tests	53,246,070	14,957,567	711,794	82,333
Outpatient encounters	16,210,828	4,234,643	212,210	90,714
Questionnaire responses	3,698,412	2,298,676	214,348	38,821
Laboratory results	2,011,177	514,539	32,372	4,153
Problem list entries	1,760,379	435,109	18,443	12,039
Admissions	1,536,202	337,182	23,720	14,888
Surgical procedures	256,340	61,639	3,826	3,731
Anaesthetics	191,999	44,529	2,955	2,854
Pregnancies	43,594	11,095	204	204
Source: Epic Caboodle. Counts represent rows in selected Caboodle fact tables. One record has a different meaning in each domain.				

1.93bn

flowsheet observations

609.2m

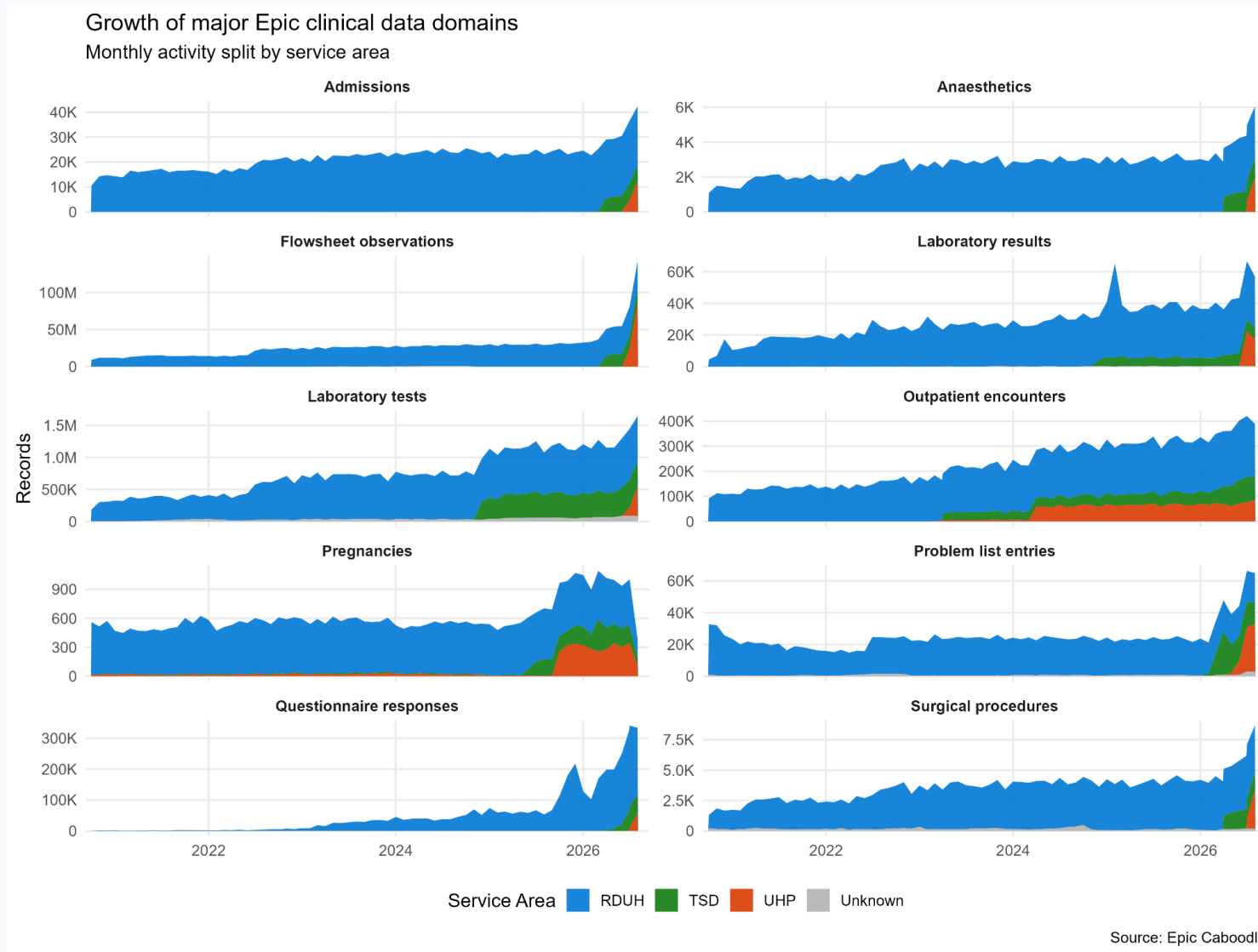
added in the last 12 months

The value is in the breadth:
encounters + observations +
tests + procedures +
questionnaires

Counts are rows in selected Caboodle fact tables; record meaning differs by domain.

The resource is longitudinal, and still growing

Monthly activity by clinical domain and service area



Six years

East Devon data from October 2020, with later expansion across North Devon and new One Devon sites.

Multiple modalities

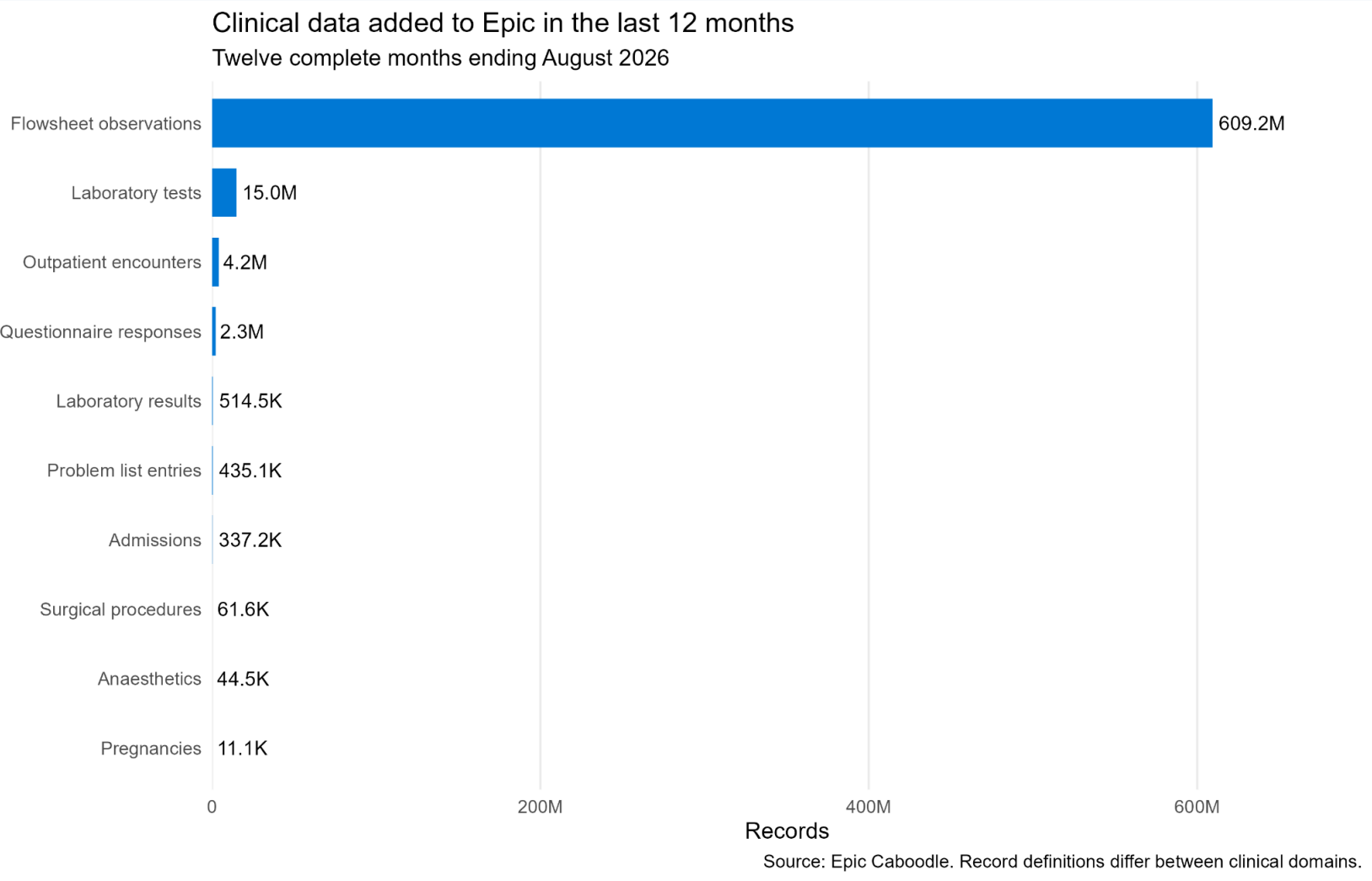
Clinical activity, observations, diagnostics, procedures and patient-entered information.

Research implication

Questions can follow pathways over time, not just count isolated events.

Every year adds another major research asset

Twelve complete months ending August 2026



609.2m

flowsheet observations

15.0m

laboratory tests

4.2m

outpatient encounters

**This is a renewable asset,
not a one-off extract.**

RDUH data alone can answer clinically useful questions

Infection and antimicrobial use

Link antibiotic choice and indication to cultures, susceptibilities, timing and outcomes.

Acute hospital at home

Combine observations, clinical contacts and integrated wearable data to understand deterioration and recovery.

Chronic disease pathways

Reconstruct treatment sequences, monitoring, escalation and outcomes, for example in inflammatory bowel disease.

Adding research and complex datasets

- Genomic data:
 - Exeter 10,000 cohort
 - Disease-specific cohorts, including within Exeter Tissue Bank
- Peninsula-wide links for radiological imaging via PenRad
- Whole slide imaging/distal histopathology

On-premise AI can unlock the information trapped in text

A locally controlled route from narrative to validated, research-ready variables



Southampton collaboration: IBD phenotype extraction for genomic association.

A single Epic footprint changes the unit of research

From Royal Devon to Devon-wide acute and specialist community pathways



One longitudinal record

Royal Devon

East Devon from 2020; North Devon from 2022.

Torbay and South Devon

Epic launched in April 2026.

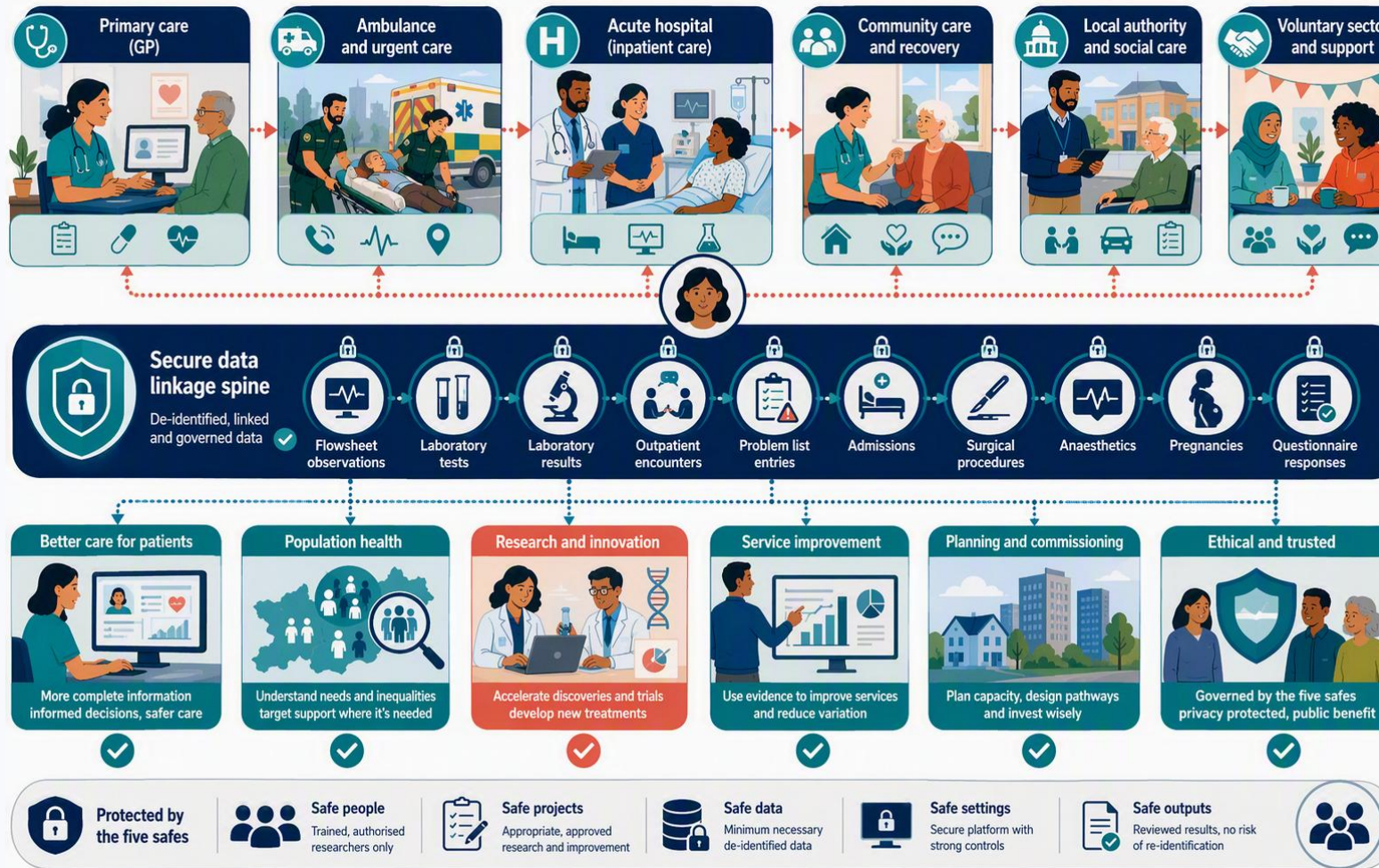
University Hospitals Plymouth

Live since July 2026

Research can study variation and pathways across organisations without first reconciling three unrelated EPRs.

The biggest questions cross organisational boundaries

Link the pathway, while keeping access controlled and purpose-specific



Cancer diagnostic pathways

Primary care presentation → tests → referral → diagnosis → treatment.

Acute MI and stroke

Ambulance observations and timings → acute care → rehabilitation.

Frailty across settings

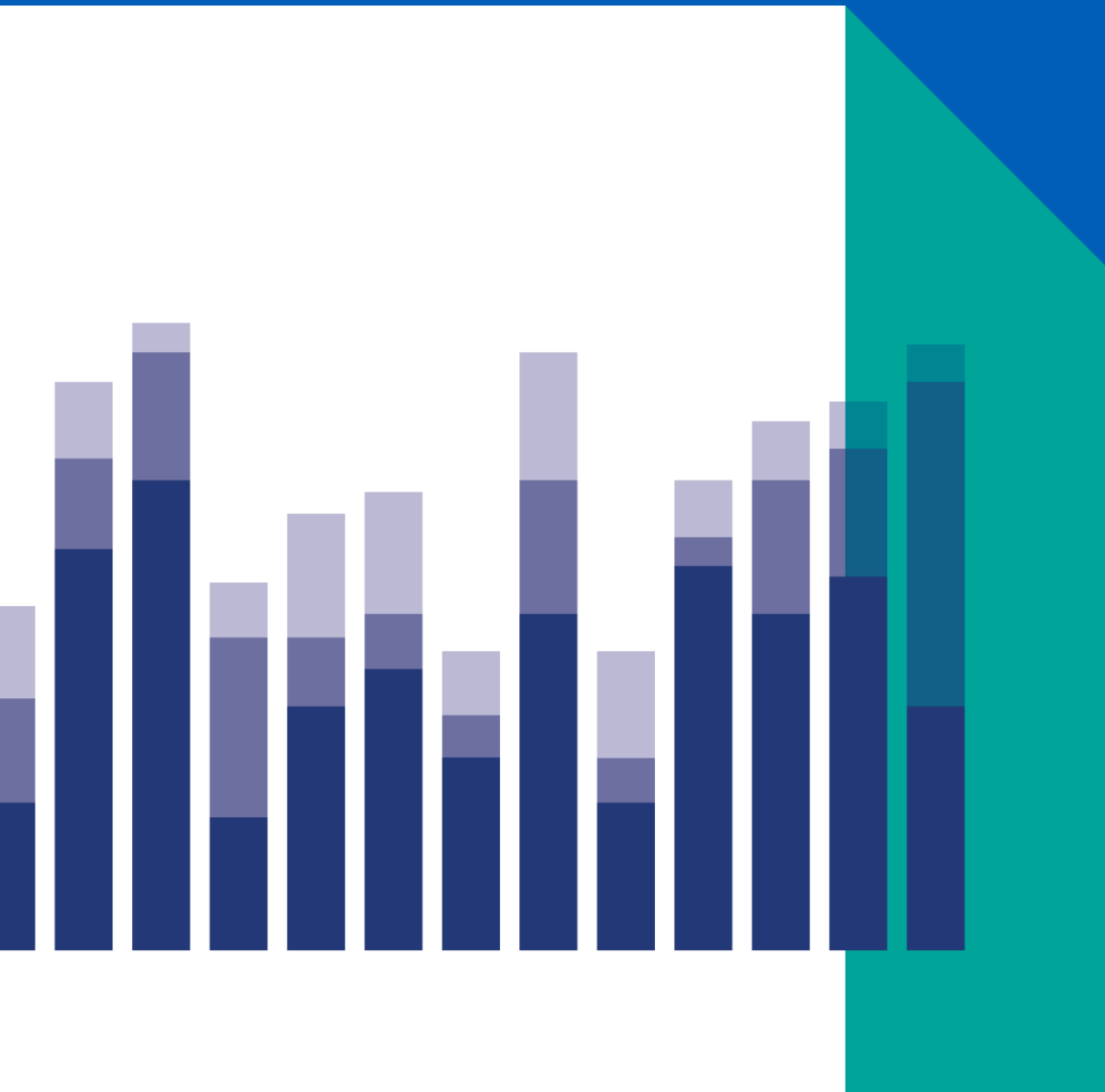
Hospital + community + local authority + voluntary-sector support.

Linkage turns fragments into a pathway.
The SDE keeps the pathway governed.

Work closely together

- NHS organisations define clinically meaningful pathways
- Social care and voluntary partners add the missing context
- Universities bring methods, skills and challenge
- The SDE provides secure, governed access
- Patients and the public shape what is worthwhile and acceptable

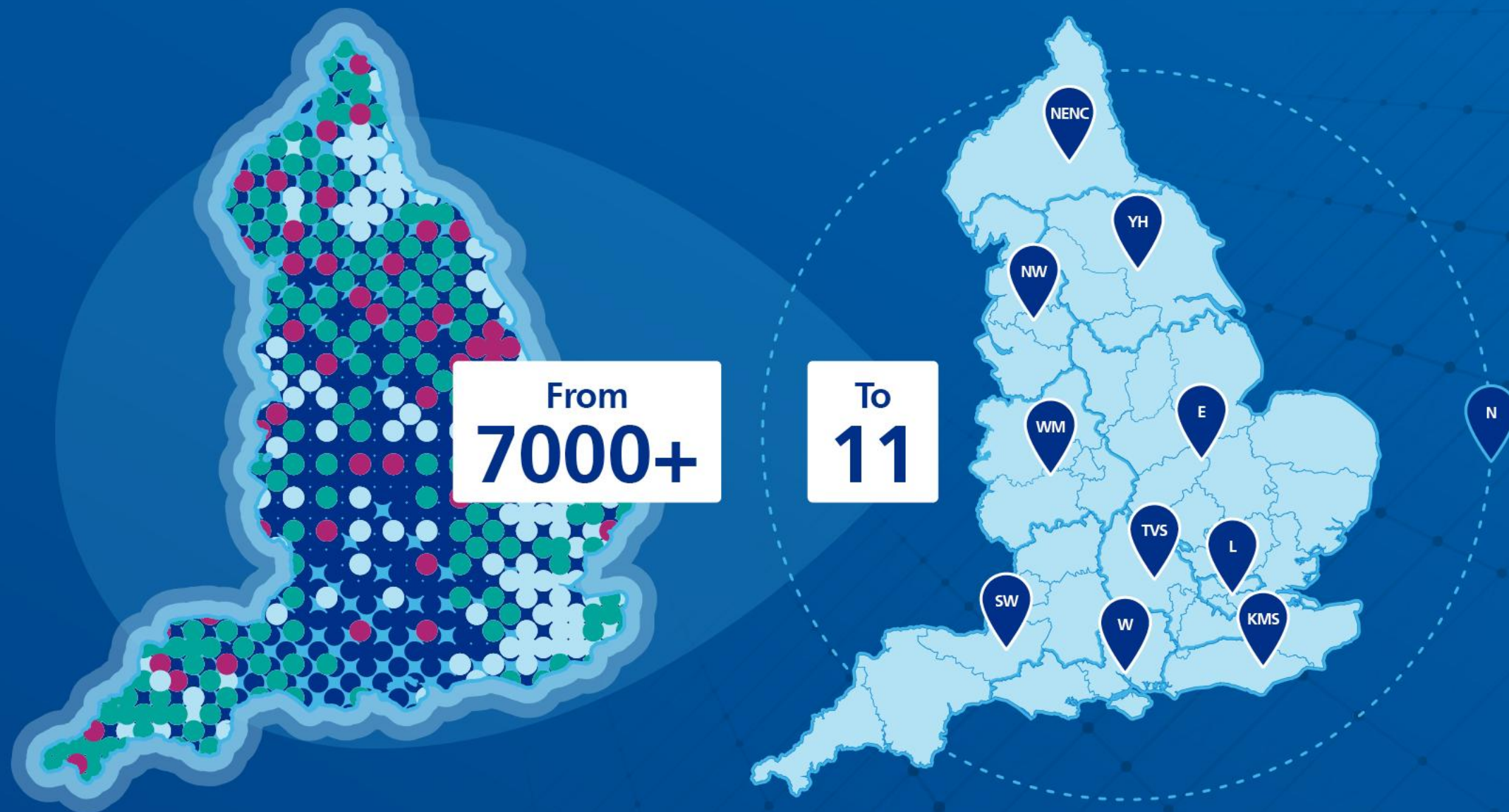
Thank you



Kane O'Beirne

Head of Operations,
South West SDE

Streamlining researcher contact points with data controllers



Data partners



South West
SECURE DATA
ENVIRONMENT

Part of the
**NHS Research Secure Data
Environment Network**

Acute
(4.4m/5.5m)

80%

Community
(2.5/5.5m)

45%

Primary
care
(1m/5.5m)

18%

Acutes (estimated catchment):

- Royal United Hospitals Bath (500,000 population)
- University Hospitals Plymouth (475,000 population)
- Royal Devon University Healthcare (615,000 population)
- Gloucestershire Hospitals NHS Foundation Trust (GHFT) (600,000)
- Great Western Hospitals (270,000)
- Salisbury Foundation Trust (270,000)
- North Bristol Trust (750,000)
- University Hospitals Bristol and Weston (500,000)

Primary Care and Community:

- BrisDoc Health Services - out of hours service for (1 million population)
- BNSSG primary care charter (c.50 practices)
- Avon and Wiltshire Mental Health Partnership (1.6m population)
- Gloucester County Council (GCC)
- Gloucestershire Health and Care NHS Foundation Trust (GHC)
- Integrated Care 24 (IC24)

VSCE, Academic:

- Caafi health
- Centre for Sustainable Energy
- University of Bristol
- Extend (10,000 consented cohort)
- Exeter Genomics Lab

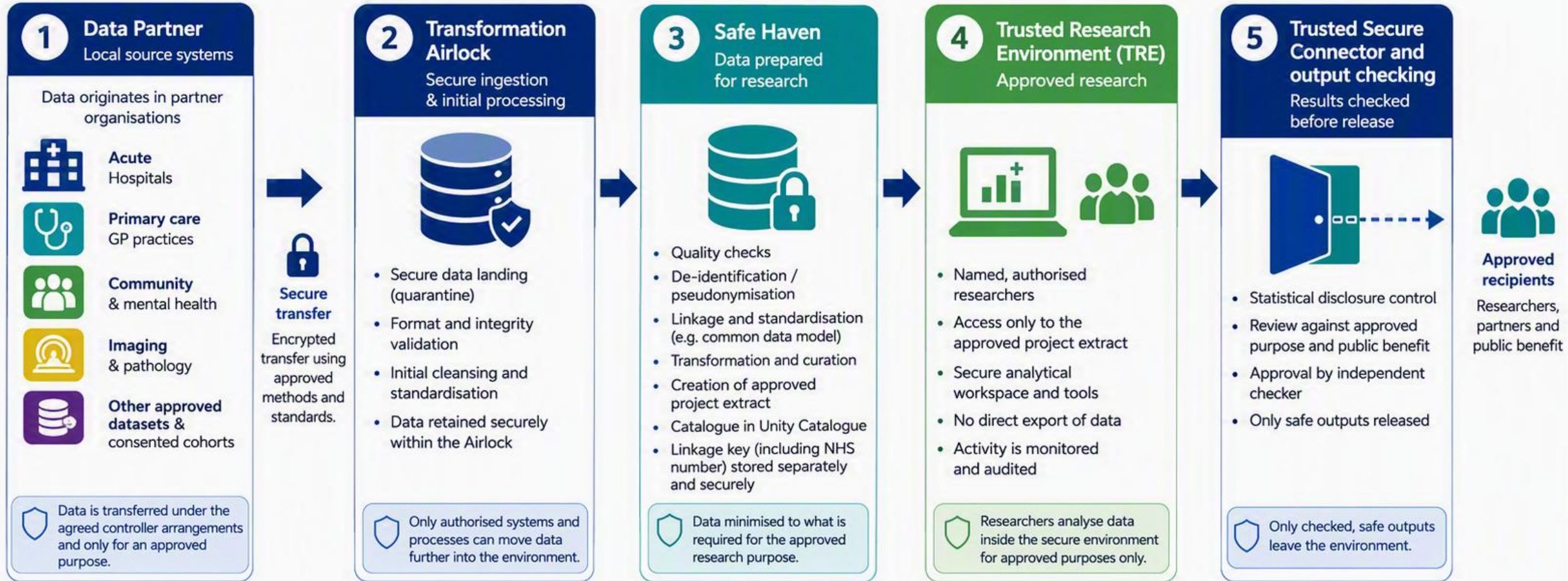
DSAs being reviewed:

- Royal Cornwall Hospitals Trust
- Somerset Foundation Trust
- PenRAD and Peninsula Pathology

Estimated population coverage where a data-sharing agreement or existing framework is in place.

How the South West Secure Data Environment (SWSDE) works

Our data journey: from source to safe outputs



Project approval

DAC, Leadership Board and local approvals



The Five Safes

Safe projects, people, settings, data, outputs



Governed throughout

Access controls

RBAC, least privilege and segregation



Monitoring & audit

Comprehensive logging, monitoring and audit trails



Security assurance

SATRE, DSPT, penetration testing and external audit



Data is brought into the SWSDE only for an approved purpose, prepared securely, analysed securely within the TRE, and only checked outputs are released.

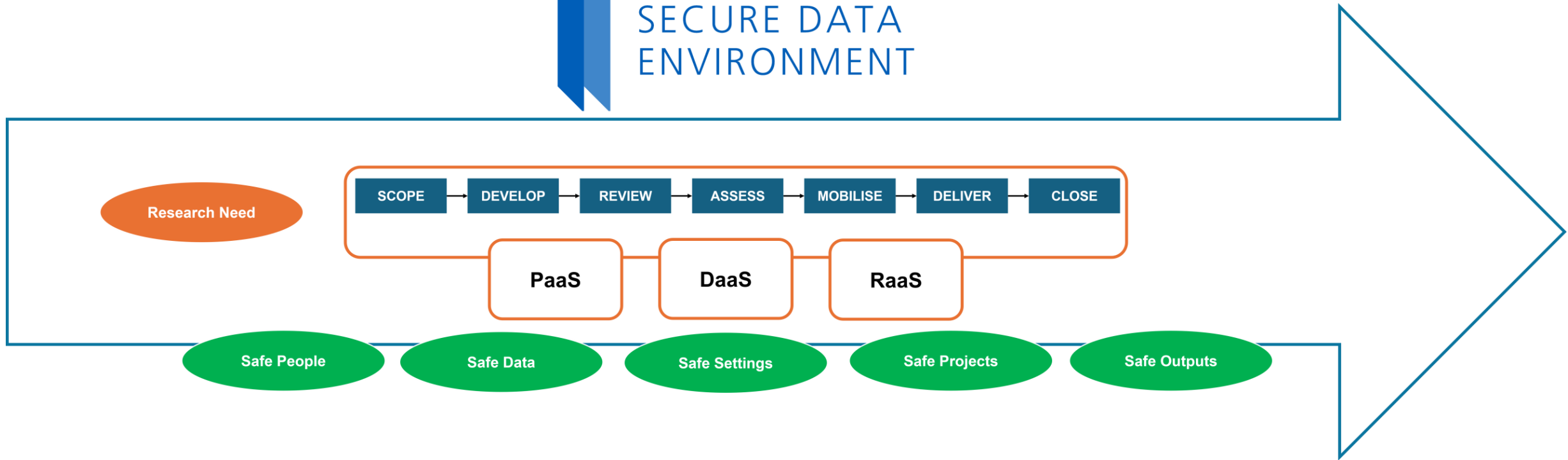


South West
Secure Data Environment



South West

SECURE DATA ENVIRONMENT



Contact us – southwestsde.nhs.uk



Part of the
**NHS Research Secure Data
Environment Network**



Search this we



[Home](#)

[What is an SDE?](#)

[Public involvement](#)

[Research](#)

[News](#)

[About us](#)

[About us](#) > [Contact us](#)

Contact us

Researcher enquiries

If you are a researcher interested in accessing data through the South West SDE, visit our [research page](#) for more information, or [complete this request form](#).

Readiness Assessment Matrix

Priority Type / Score	0	1	2	3	4	5
Strategic Alignment: Does this project align with SWSDE, NHS, or national research priorities?	Unknown or insufficient information to assess alignment	Very weak alignment to any recognised SWSDE, NHS or national priority	Limited alignment to niche, local or low-priority themes only	Aligns to at least one recognised SWSDE, NHS, or national priority	Strong alignment to multiple SWSDE and/or NHS or national priorities	Directly aligned to a key national or SWSDE priority programme or flagship initiative
Public Benefit: Does this project clearly benefit patients, the NHS, or population health?	Unknown or insufficient information to assess public benefit	Minimal or poorly articulated benefit	Limited or indirect benefit with unclear outcomes or impact	Moderate benefit to a defined patient group, population cohort or service area	Clear and meaningful benefit to patients, NHS, or population health	Significant, scalable, or high-impact public benefit aligned to national or system priorities
Feasibility: Can we realistically deliver this with available data, systems, and technical capability?	Unknown or insufficient information to assess feasibility	Major feasibility issues with no clear solution	Significant gaps in data, systems and/or technical approach	Feasible with some dependencies or unknowns, for example dataset not yet available but expected in time	Largely feasible using mostly available data and standard SDE tools, with minor clarifications needed	Fully feasible using existing data assets, tools and established delivery approaches
Financial: Is funding in place to support delivery of the project?	Unknown or insufficient information on funding position	Very limited or uncertain funding with no clear route to secure sufficient resources	Partial funding identified but gaps remain; funding approach not fully defined	Funding expected or in progress but not yet confirmed or secured	Funding largely secured with minor gaps or dependencies	Full funding secured, including clear allocation for SDE costs and delivery
Operational Impact: How much effort will this require from the SDE?	Unknown or insufficient information to assess operational impact	Extremely high effort requiring significant new capability, multiple dependencies, or major coordination	High effort requiring multiple teams and coordination and some additional work required	Moderate effort across teams with manageable coordination and some additional work required	Manageable within existing processes with limited additional effort	Low effort and straightforward to deliver within existing capacity and workflows
Risks: Are there legal, ethical, IG, or reputational risks?	Unknown or insufficient information to assess risk	Very high risk with major unresolved concerns or contentious elements	High risk requiring significant mitigation	Moderate risk with manageable mitigations	Low risk with minor considerations	Minimal or no identifiable risk

Consultation form

South West Secure Data Environment: Consultation Form

Introduction. This is the list of questions our team need to understand to fully scope and cost your project. We will talk through each on the call but please fill in as much information ahead of time as you can.

Questions

Project Title	Provide a clear, descriptive title that reflects the scope and purpose of the project.
Lead Researcher	Provide the name of the individual responsible for overall project delivery and coordination.
Lead Organisation	Provide the name of the organisation accountable for the project and primary contractual engagement.
Lead Researcher Email	Provide the primary contact email for project communications.
Project Summary	Provide a concise overview of the project, including its purpose, scope, key activities, and intended outcomes.
Primary Research Question	Clearly state the main question the project aims to answer.
Project Objectives	List the specific, measurable objectives that support the primary research question.
Expected Public Benefit	Describe the anticipated societal, healthcare, or economic benefits that will result from the project.
Collaborating Organisation(s)	List all partner organisations involved in the project and briefly describe their roles.
What is the study design / methodology?	Describe the study design and methodological approach (e.g. cohort study, case-control study, randomised controlled trial). Where relevant, please also outline any analytical or computational approaches (e.g. statistical modelling, machine learning) that will be used.

What is the population being studied?	Define the target population and any relevant demographic or clinical characteristics.
What is the inclusion criteria?	Specify the criteria used to include subjects or data in the study.
What is the exclusion criteria?	Specify the criteria used to exclude subjects or data from the study.
What is the estimated cohort size?	Provide the approximate number of records or participants, including any expected growth over time.
What time period is required?	Specify the historical and/or prospective time range of data required for the study.
Are data refreshes required?	Please describe how frequently the data changes and how often updates should be reflected in the platform (e.g. real-time, hourly, daily, weekly, or ad hoc), including the acceptable latency for updates (minutes, hours, or days) and whether updates will be full refreshes or incremental. Please also note any dependencies on external data providers or schedules.
What data sets are required?	For each dataset, please provide its name and source, format (e.g. tabular, imaging, genomics, text), approximate size and number of files or tables, and any known structure (e.g. relational tables, time series, partitioning). Please also indicate whether datasets need to be linked and how this linkage will be performed.
Who are the data controllers?	Please identify the Data Controller(s) and, where applicable, the Data Processor(s), and clarify responsibilities for data access approvals, governance, compliance, and data deletion or retention decisions in line with applicable regulations such as General Data Protection Regulation.
Will you need to link multiple datasets?	Indicate whether data linkage is required and, if so, describe the method (e.g. deterministic or probabilistic matching) and any identifiers that will be used.
Will you be bringing your own data?	If yes, please describe the data source and ownership, format and size, sensitivity (e.g. anonymised, pseudonymised, identifiable, or special category data), and how the data will be ingested (e.g. secure file transfer, API, cloud storage). Please also indicate whether this will be a one-off upload or a recurring process.
Are special data types required (e.g. imaging, genomics)?	Specify any non-standard data types required and any associated storage, processing, or tooling considerations.
What is the minimum and maximum data required?	Define the smallest and largest expected data volumes to help optimise storage and compute design.

What type of analysis will be conducted?	Please describe the type of analysis (e.g. descriptive analysis, statistical modelling, machine learning or AI), how frequently it will be performed, and the expected workload pattern (e.g. batch processing, interactive analysis, or long-running training jobs). Where possible, include expected runtimes and any performance requirements (e.g. near real-time or offline processing).
What software tools will be required?	Please list the required tools and environments (e.g. Python, R, SQL), including any key libraries or packages, custom or non-standard dependencies, and whether these need to be centrally managed rather than installed by users. Please also include any required integrations with external systems or services.
What is the estimated data size?	Please provide the current total data size along with any expected growth over the course of the project, including, where possible, a breakdown between raw, processed, and output data.
Is GPU required?	If GPU acceleration is required, please describe the use case (e.g. deep learning or large-scale modelling), the tools or frameworks being used, how frequently GPUs will be needed, and any indication of the expected scale of usage. What type of VM required and expected time of use per week.
How many users require access?	Please provide the number of users by role (e.g. data engineer, data scientist, analyst, principal investigator), including their required access levels (e.g. read-only, write, admin) and whether access requirements are expected to change over time.
What storage is required?	Please estimate storage requirements, including the distinction between active and archival data, expected growth over time, any performance requirements (e.g. high-speed access vs long-term storage), and any known retention requirements.
What is the anticipated start date?	Provide the expected project start date, including any dependencies or prerequisites.
What is the anticipated end date?	Provide the expected project end date and note any flexibility or extension considerations.
Are there any key deadlines?	Include any milestones or deadlines that may impact prioritisation or delivery timelines.
Do you have funding in place?	Confirm whether funding has been secured for the project.
What is the funding source?	Provide details of the funding body, programme, or source of financial support.
Is this academic or commercial work?	Indicate whether the project is academic, commercial, or a combination of both.
Are there commercialisation plans?	Describe any plans to commercialise outputs or develop intellectual property arising from the project.
Has ethics approval been obtained?	Provide the status of ethics approval, including any reference numbers if available.

Have data controllers been engaged?	Confirm whether data controllers have been contacted and whether approvals or agreements are in progress.
What outputs will be produced?	Please describe the expected outputs (e.g. datasets, reports, dashboards, models), including their format, estimated size, and frequency of generation, and whether outputs will need to be exported outside the environment or be subject to approval or governance processes.
Will outputs be published?	Indicate whether outputs will be published and describe any restrictions or approval requirements.
Is background Intellectual Property being brought in?	Describe any pre-existing intellectual property that will be used within the project.
Are there plans for licensing or commercialisation?	Provide details of any licensing strategy or intended commercial pathways.
Is there anything else you would like to add to assist the team in delivering your project?	Please include any additional technical, governance, or operational considerations not covered above, such as data sensitivity or classification, data export requirements (including frequency, format, and approvals), or end-of-project plans for data handling (e.g. archiving, deletion, or transfer).

Internal assessment

Data Access and Quality Assessment (DAQA). This considers data availability, completeness, and quality, the feasibility of linkage and integration, technical architecture requirements, and any processing complexity associated with delivery. It also identifies key assumptions, dependencies, and risks that may impact delivery.

Information Governance Assessment (IGA). This document outlines the legal basis for processing, required approvals (such as REC, CAG) and any material IG risks or constraints which could affect delivery timelines.

Commercial Assessment (CA). This summarises the commercial viability of the project and informs costing strategy. It includes confirmation of funding position, key cost drivers and complexity considerations, any commercial risks or dependencies, and relevant contractual considerations.

Data specification

Annex A: Data Specification, Linkage and Technical Onboarding Schedule

A.1 Project Cohort and Refresh Logic

A.1.1 Project Cohort

Cohort Definition	Adults aged 18 years or older who were discharged alive from an emergency hospital admission at the fictional Example University Hospitals NHS Foundation Trust between 1 April 2021 and 31 March 2025, where the first consultant episode within the hospital provider spell records heart failure as the primary diagnosis. The cohort will be derived from admitted patient care data. Emergency care records will be linked to describe urgent and emergency care <u>activity</u> after discharge. The final research dataset will contain project-specific pseudonymous identifiers only.
--------------------------	--

A.1.2 Inclusion Criteria

Rule ID	Description	Logic
INC-01	Adult at index admission	Age derived from PERSON BIRTH DATE and START DATE (HOSPITAL PROVIDER SPELL) is ≥ 18 years.
INC-02	Finished spell in study window	DISCHARGE DATE (HOSPITAL PROVIDER SPELL) between 2021-04-01 and 2025-03-31 inclusive.
INC-03	Emergency admission	ADMISSION METHOD code is within the protocol-approved emergency admission code set, supplied as a version-controlled <u>codelist</u> .
INC-04	Heart failure primary diagnosis	PRIMARY DIAGNOSIS (ICD) on the first consultant episode begins I50; exact ICD-10 <u>codelist</u> stored with specification version.
INC-05	Discharged alive	METHOD OF DISCHARGE (HOSPITAL PROVIDER SPELL) is not code 4 (patient died) or 5 (stillbirth).
INC-06	Linkage quality	Valid NHS NUMBER and acceptable identifier quality for deterministic <u>linkage</u> , or approved alternative linkage route documented before extract.

A.1.3 Exclusion Criteria

Rule ID	Description	Logic
EXC-01	Same-day apparent transfer	Exclude a subsequent spell starting on the index discharge date where transfer/discharge destination logic indicates continuation of care.
EXC-02	Invalid chronology	Exclude where discharge date precedes admission date or required dates cannot be resolved.
EXC-03	No follow-up eligibility	Exclude index spells where the available data cut does not provide 30 complete days of follow-up.
EXC-04	Duplicate spell	Retain one record after deduplication using provider, spell identifier and admission/discharge dates; unresolved conflicts quarantined.
EXC-05	Overlapping index periods	Where a person has multiple eligible spells within 90 days, retain the earliest as index and treat later spells as outcomes.

A.3 Phase and Refresh Schedule

Phase	Target Date	Datasets	Refresh Schedule	Operational Notes
Discovery	Month 1	APC and ECDS metadata/sample schema	None	Confirm availability, local mapping, <u>codelists</u> and volumes.
Initial extract	Month 2	APC + ECDS	One-off	Identifiable transfer to trusted linkage function; no researcher access.
Curation	Months 2–3	Linked pseudonymised tables	Iterative internal builds	Data-quality issues returned to provider; specification version controlled.

Research release	Month 3	Approved analytical views	Release 1	Only approved pseudonymised fields made visible in project workspace.
Planned refresh	Month 9	APC + ECDS	One refresh	Same schema and logic unless change control approved.
Final freeze	Month 10	Curated analytical dataset	No further refresh	Dataset version locked for final analysis and publication.

A.2. Dataset Register

A.2.1 Dataset 1: Admitted Patient Care

Data Delivery Process						
Dataset	Commissioning Data Set (CDS) admitted patient care, finished general episodes, transformed to spell and episode tables.					
Data controller	Example University Hospitals NHS Foundation Trust (fictional).					
Data engineering lead	Provider Data Warehouse Team, named contact to be confirmed.					
Source system	Provider patient administration system and clinical coding warehouse.					
Data location	Provider-managed secure SQL data warehouse.					
Transfer method	Encrypted approved transfer into the SWSDE landing zone; method confirmed during technical onboarding.					
Status	In discovery (illustrative).					
Extraction window	1 April 2020 to 30 June 2025, providing look-back and outcome coverage.					
Analytical purpose	Derive index heart-failure admissions, demographic and clinical covariates, prior use and readmission outcomes.					
Volume / file size	Invented planning estimate: 1.8 million episodes, 4 GB compressed. Provider to confirm.					
Format	Parquet preferred; UTF-8 CSV accepted by agreement. Dates in ISO 8601. One row per episode plus spell-level table.					
Grain and keys	Episode table: one row per consultant episode. Spell table: one row per hospital provider spell. Provider-local keys retained only in protected linkage staging.					
Field-Level Data Dictionary						
Field Name	Description	Data Type	Format / Coding Standard	Linkage Use	Sensitivity	Notes
obs_number	NHS NUMBER	Unique person identifier used only in protected linkage staging.	numeric string	n10	Linkage	Direct identifier
person_birthday	PERSON BIRTH DATE	Derive age at admission and support linkage quality checks.	date	CCYY-MM-DD	Linkage / derivation	Direct identifier
person_stated_gender_code	PERSON STATED GENDER CODE	Describe cohort and adjust analysis where justified.	string	an1 national code	Analysis	Special category / quasi-identifier
ethnic_category	ETHNIC CATEGORY	Equity analysis using self-stated ethnicity.	string	an2; national code in first character	Analysis	Special category
postcode_usual_address	POSTCODE OF USUAL ADDRESS	Derive approved geography and deprivation measures at source/controlled staging.	string	UK postcode	Geography derivation	Direct identifier
organisation_identifier_provider	ORGANISATION IDENTIFIER (CODE OF PROVIDER)	Identify provider responsible for activity.	string	min an3 max an8	Analysis	Operational
hospital_provider_spell_id	Local Hospital Provider Spell Identifier	Join episodes belonging to the same spell.	string	provider-local	Internal linkage	Confidential identifier
episode_number	Episode sequence number	Order consultant episodes within a spell.	integer	positive integer	Cohort / analysis	Operational
spell_start_date	START DATE (HOSPITAL PROVIDER SPELL)	Date of admission and start of index/risk spells.	date	CCYY-MM-DD	Cohort / outcome	Quasi-identifier

~~spell_discharge_date~~	DISCHARGE DATE (HOSPITAL PROVIDER SPELL)	Index date and end of hospital provider spell.	date	CCYY-MM-DD	Cohort / outcome	Quasi-identifier
~~admission_method~~	ADMISSION METHOD	Identify emergency versus elective admission.	string	national code	Cohort / outcome	Clinical
~~method_of_discharge~~	METHOD OF DISCHARGE (HOSPITAL PROVIDER SPELL)	Exclude deaths and support transfer logic.	string	an1 national code	Cohort	Clinical
~~discharge_destination~~	DISCHARGE DESTINATION	Support transfer and pathway interpretation.	string	national code	Cohort / analysis	Clinical
~~primary_diagnosis_icd~~	PRIMARY DIAGNOSIS (ICD)	Identify heart failure index diagnosis.	string	ICD-10 code; min an4 max an6	Cohort / analysis	Clinical sensitive
~~secondary_diagnosis_icd~~	SECONDARY DIAGNOSIS (ICD)	Derive comorbidities recorded during episode.	string	ICD-10	Analysis	Clinical sensitive
~~primary_procedure_opcs~~	PRIMARY PROCEDURE (OPCS)	Describe major procedures during index admission.	string	OPCS-4	Analysis	Clinical sensitive
~~care_prof_main_specialty~~	CARE PROFESSIONAL MAIN SPECIALTY CODE	Describe responsible consultant specialty.	string	an3 national code	Analysis	Operational

A.3 Platform Access and Analytical Requirements

A.3.1 Platform Users

First Name	Surname	Email	Job Title	Organisation	Role	Access Requirements
Sarah	Williams	sarah.williams@example.ac.uk	Senior Research Fellow	University of Exeter	Principal Investigator	Full access to approved pseudonymised datasets, SQL, R and Python environments, ability to create analytical outputs for disclosure review
James	Patel	james.patel@example.ac.uk	Research Data Scientist	University of Bristol	Data Analyst	Access to approved datasets, Python and R environments, ability to develop and run analytical workflows and statistical models
Emily	Thompson	emily.thompson@example.nhs.uk	Clinical Research Associate	Royal Devon University Healthcare NHS Foundation Trust	Clinical Researcher	Read-only access to approved datasets, ability to review cohort characteristics and contribute to validation and interpretation activities

A.3.2 Analysis and Compute Requirements

Analysis	Frequency	User	Tooling	Packages/ Libraries	Current execution Environment	Compute Requirement	GPU Required	Notes
Cohort validation and descriptive analysis	Repeated during setup	Statistician / analyst	SQL, R	~~tidyverse~~ , ~~data.table~~ , arrow	Standard: 4 vCPU, 16 GB RAM	No	Package versions fixed in reproducible environment file.	Cohort validation and descriptive analysis
Regression modelling	Weekly during analysis	Statistician	R	survival, ~~glmnet~~ , broom	High memory: 8 vCPU, 32 GB RAM	No	No automated clinical decision support.	Regression modelling
Data quality reports	Each release	Analyst	Python, SQL	pandas, ~~pvarrow~~	Standard	No	Reports saved with dataset version.	Data quality reports

Next events

Commercial partnerships and the South West SDE

Wednesday 30th September, 1-2pm (online)

In-person workshops at Sandy Park, Exeter

Exploring data available through the South West SDE

Thursday 22nd October, 10am – 4pm

Sandpit: data driven health and care research in the South West

Wednesday 18th November (details to be confirmed)

www.southwestsde.nhs.uk/explore-connect-collaborate-autumn-research-events/



**You will also find the
recording and slides
from the first webinar
on this web page.**

Questions



Contact us

bnssg.swsde@nhs.net

www.southwestsde.nhs.uk

LinkedIn:

[@south-west-secure-data-environment](#)

Bluesky:

[@southwestsde.bsky.social](#)



Part of the
**NHS Research Secure Data
Environment Network**

Thank you