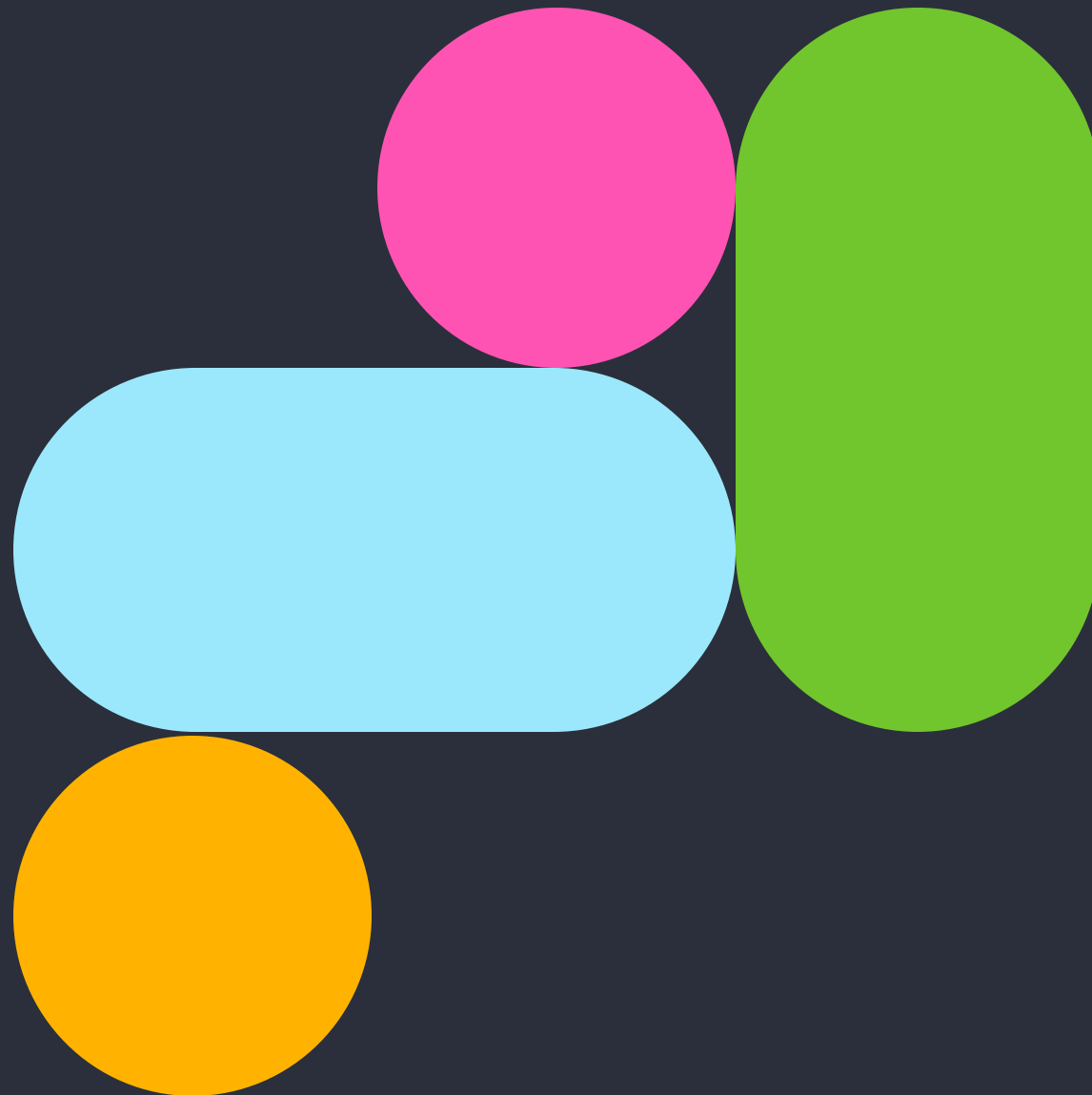


Getting medical history for participants

Emily Perry

9th September 2025



Data security

- This training session will include data from the GEL Research Environment
- As part of your IG training you have agreed to not distribute these data in any way
- You are not allowed to:
 - Invite colleagues to watch this training with you
 - Take any screenshots or videos of the training
 - Share your webinar link (we will remove anyone who is here twice)
- We will record this training and distribute the censored video afterwards

Questions



All your
microphones
are muted



Use the Zoom
Q&A to ask
questions



Upvote your
favourite
questions: if we
are short on
time we will
prioritise those
with the most
votes

Questions



Jonathan Gosling
Data Wrangler -
Data Release
Team



Laura Kerr
Data Manager



Gemma Richards
Data Wrangler -
Data Release
Team

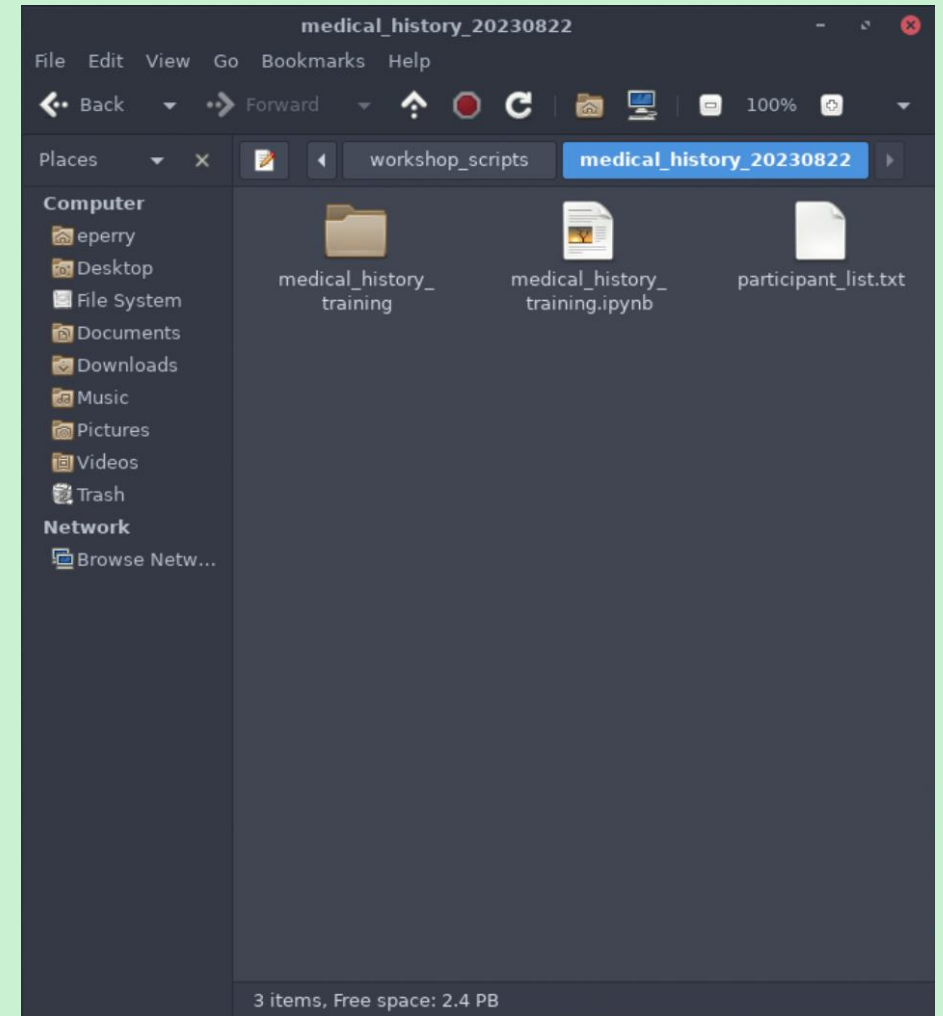
Agenda

- 1 Introduction and admin
- 2 NHS England data in the RE
- 3 Mental health data in the RE
- 4 Accessing and comparing NHS England data with Participant Explorer
- 5 LabKey tables
- 6 Accessing medical history programmatically
- 7 Help and questions

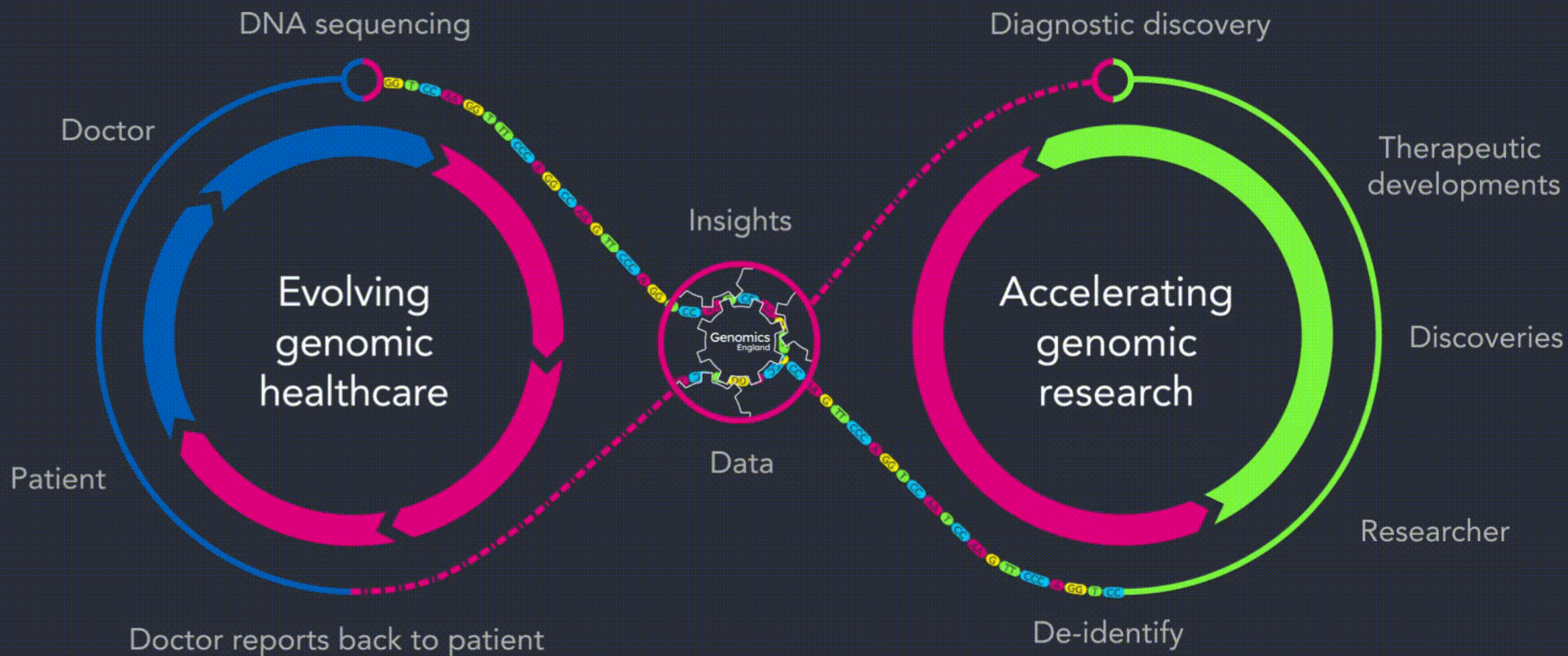


Materials

- Slides and video will be sent out to you after the session
- Scripts available in
`/gel_data_resources/example_scripts/workshop_scripts/medical_history_2025`



2. NHS England data in the RE



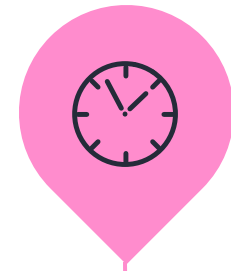
Genomics England data ingestion



Birth



Genomics England
registration

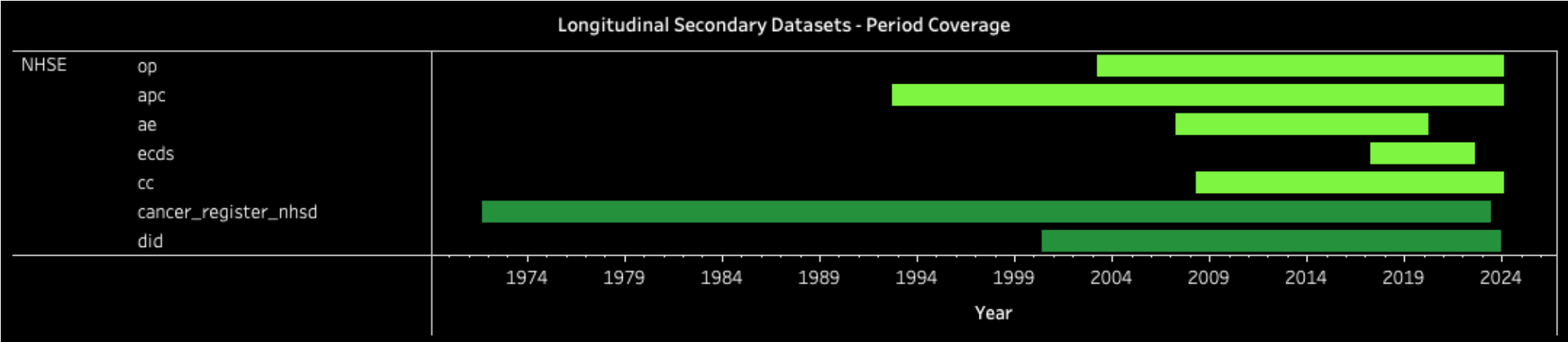


Now

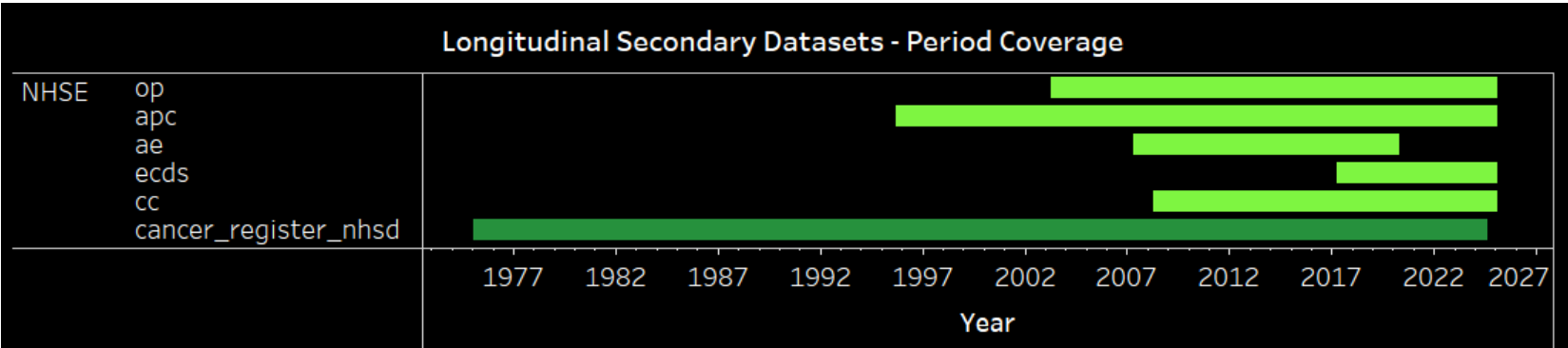


Medical history coverage over time

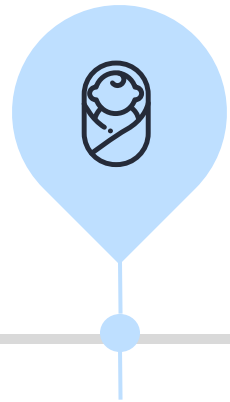
100kGP



NHS
GMS



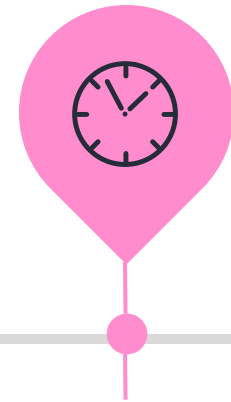
Genomics England data ingestion



Birth



Genomics England
registration



Now



Hospital episode statistics

Out-patients
op

Planned day
appointments in
hospital

Admitted
patient care

apc
Overnight hospital
stays

Critical care
cc

Time on life
support

Accident
and
emergency
ae

Unplanned
emergency visits –
walk-in or
ambulance

Outpatients **op**

Planned day appointments at a hospital

- Date/participant age
 - at appointment: **apptage**, **apptdate**
 - start/end of full series of appointments: **rttper***
- Diagnosis codes ICD10 : **diag***
- Procedures: **opertn***
- Clinician speciality and type: **tretspef**, **mainspef**, **stafftyp**
- Referral circumstances: **firstatt**, **refsourc**, **refdate**
- Location/demographics/admin for NHS charging

Admitted patient care **apc**

Hospital overnight stays

- Date/age at admission/discharge
- Admission circumstances: **admimeth**, **admisourc**, **admistat**, **epikey**, **aekey**, **referorg**
- Discharge circumstance: **disdate**, **disdest**, **dismeth**
- Diagnosis codes ICD10: **diag***
- Treatments/procedures/investigations: **domproc**, **opertn***, **operdate***, **operstat**
- Clinician speciality and type: **mainspef**, **tretspef**
- Admission-type specific: **maternity**, **alcohol**, **psychiatric**
- Critical care periods
 - Link to cc: **susrecid**
 - Admission: **apcsour***, **apcplan***
 - Type of care: **acploc***, **apcspef***, **orgsup***
 - Discharge: **acpdisp***, **acpend***
 - Timing: **acpn***, **apcstar***, **intdays***
 - Death: **apcout***
- Location/demographics/admin for NHS charging

Critical care **cc**

Time on life support

- Link to apc: **susrecid**
- Date admission/discharge: **ccstart***, **ccdismatch**
- Days spent under types of care: ***days**
- Admission (**ccadm***) and discharge (**ccdismatch***) circumstances
- Unit type: **ccunitfun**

Accident and emergency **ae**

Unplanned emergency hospital attendance – walk-ins and ambulance

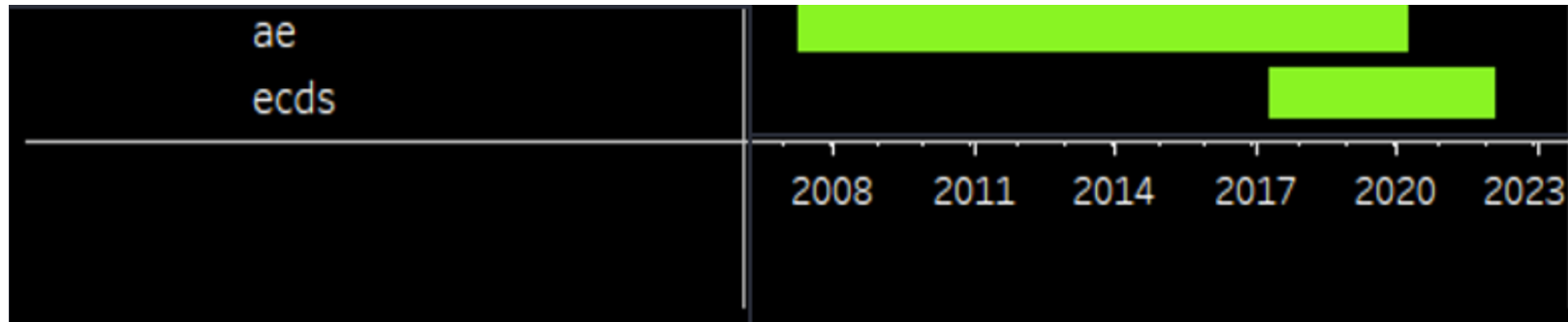
- Dates/age admission/discharge
- Diagnosis codes: ICD10 (**diag_***), type of injury (**diag2_***), anatomical area (**diaga_***, **diags_***) - ICD10
- Treatments/procedures/investigations: **treat***, **invest***
- Arrival circumstances: **aearrivalmode**, **aeincloctype**
- Discharge or department/ward admitted to **epikey**, **aekey**, **aeattenddisp**
- Location/demographics/admin for NHS charging

Emergency care data set **ecds**

Unplanned emergency hospital attendance – walk-ins and ambulance

- Dates/age admission/discharge
- Clinician speciality and type: **care_professional_tier***, **department_type**
- Diagnoses: **diagnosis_code***, **injury***, **comorbidities***
- Treatments/procedures/investigations: **investigation***, **treatment***
- Arrival circumstances: **arrivalmode**
- Discharge or department/ward admitted to: **discharge***, **referred_to_service***

ae/ecds



Hospital episode statistics

AutoSave OFF r17_data_dictionary

Home Insert Draw Page Layout Formulas Data Review View Automate Acrobat Tell me

Paste Arial 12 B I U Alignment Number Conditional Formatting Format as Table Cells Editing Analyse Data Sensitivity Create and Share Adobe PDF

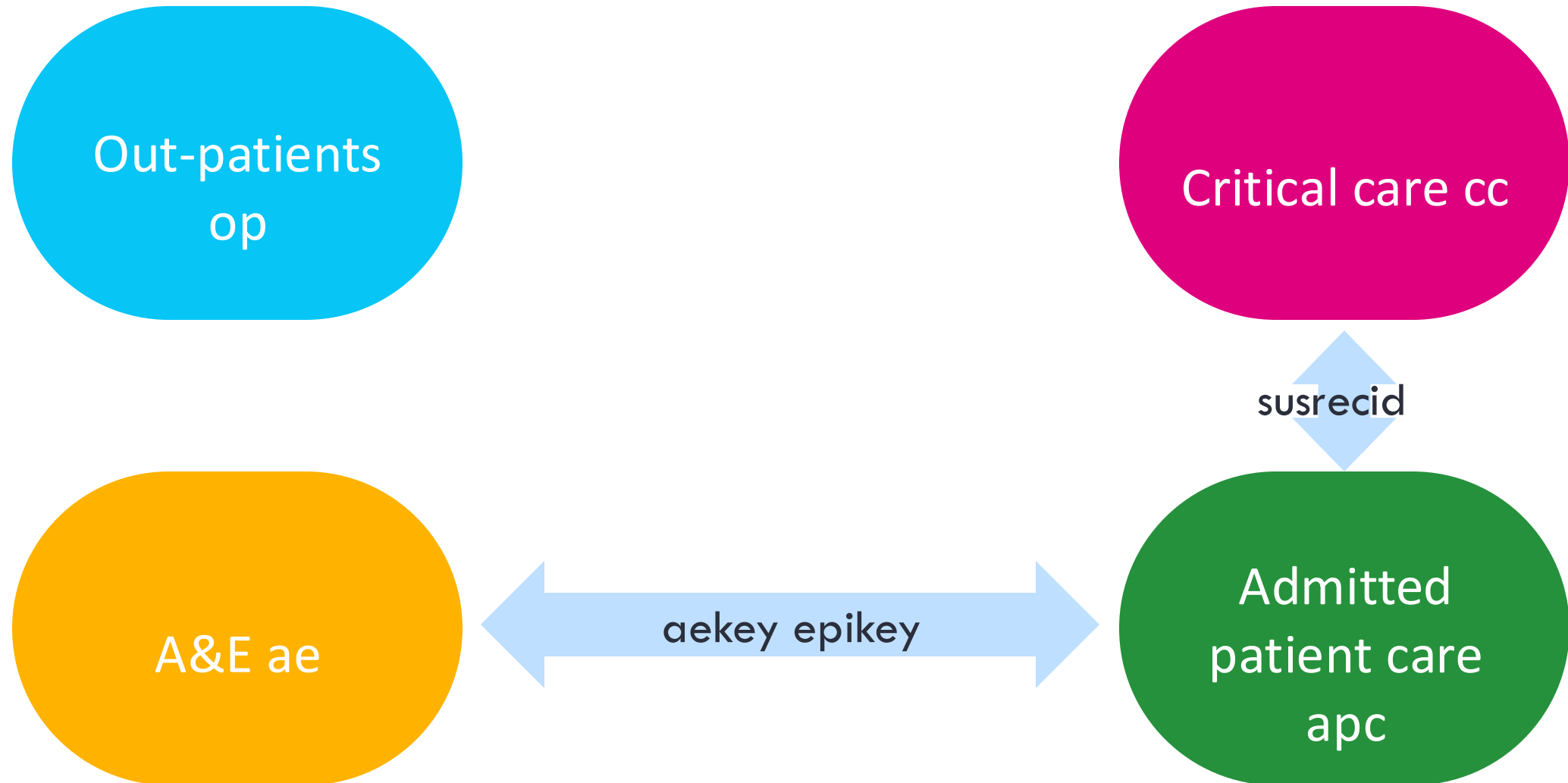
A2 fx cancer_registry

	A	B	C	D	E
1	Table	Field	Category	Short Name	Description
2	cancer_registry	participant_id			Participant's Identifier (supplied by Genomics England)
3	cancer_registry	cancer_registration_year			Cancer Registration year
4	cancer_registry	event_date			Date of cancer diagnosis
5	cancer_registry	cancer_site			ICD10 code of cancer site from cancer network register
6	cancer_registry	cancer_type			ICD-O3 morphology code from cancer network register
7	cancer_registry	cancer_behaviour			Cancer behaviour
8	did	did_date1	Diagnostic Test Dates	Diagnostic Test Request Date	The date as submitted that the diagnostic provider received the referral request. It must be in the correct CCYY-MM-DD format and meet the following validation rules: Date submitted must be: <= Diagnostic Test Request Received Date If date submitted >1 year before Diagnostic Test Date an alert will be raised.
9	did	did_date2	Diagnostic Test Dates	Diagnostic Test Request Received Date	The date as submitted that the diagnostic provider received the referral request. It must be in the correct CCYY-MM-DD format and meet the following validation rules:
10	did	did_date3	Diagnostic Test Dates	Diagnostic Test Date	The date as submitted that the diagnostic test took place. It must be in the correct CCYY-MM-DD format and meet the following validation rules: Date of test must be: >= Diagnostic Test Request Received >= Diagnostic Test Request Date Date <= Service Report Issue Diagnostic Test Date is mandatory and cannot be left blank
11	did	did_date3_month	Diagnostic Test Dates	Month of Test	Based on submitted field Diagnostic Test Date, returning month part only
12	did	did_date3_year	Diagnostic Test Dates	Year of Test	Based on submitted field Diagnostic Test Date, returning year part only
13	did	did_date4	Diagnostic Test Dates	Service Report Issue Date	The date as submitted that the diagnostic provider issues the test report. It must be in the correct CCYY-MM-DD format and an alert raised if >1 month after Diagnostic Test Date
	did	did_ethcat	Ethnic Category	Ethnic Category Code	Code as per submission. Field undergoes hard validation during data submission process. Field must contain one of the nationally defined codes. If the code supplied

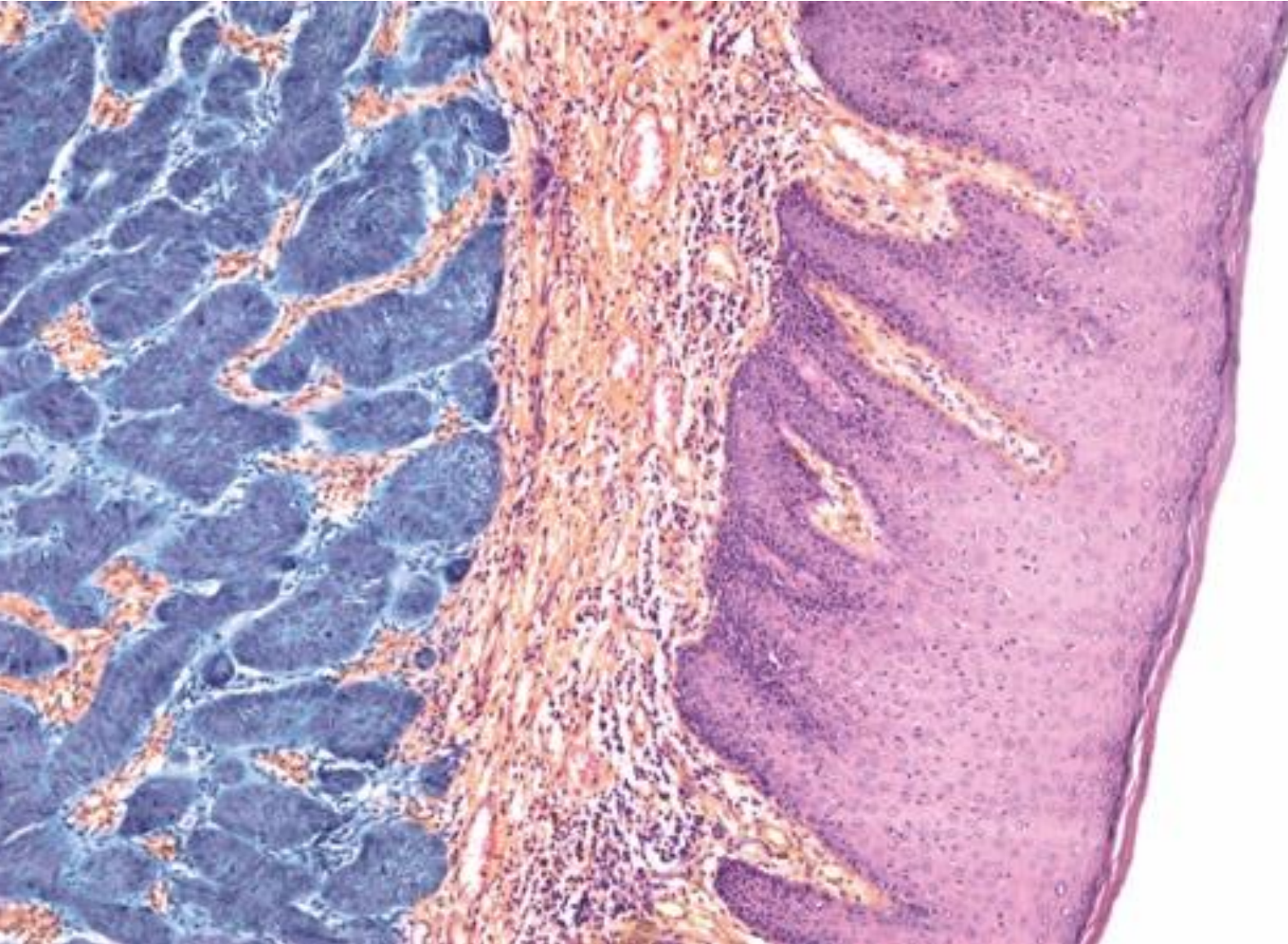
Long Read Sequencing Research Community Provided Data Secondary Data - PHE NCRAS Secondary Data - NHS Secondary Data - MHSDS +

Ready Accessibility: Investigate 100%

Hospital episode statistics



cancer_registry



participant_id
cancer_registration_year
event_date
cancer_site
cancer_type
cancer_behaviour

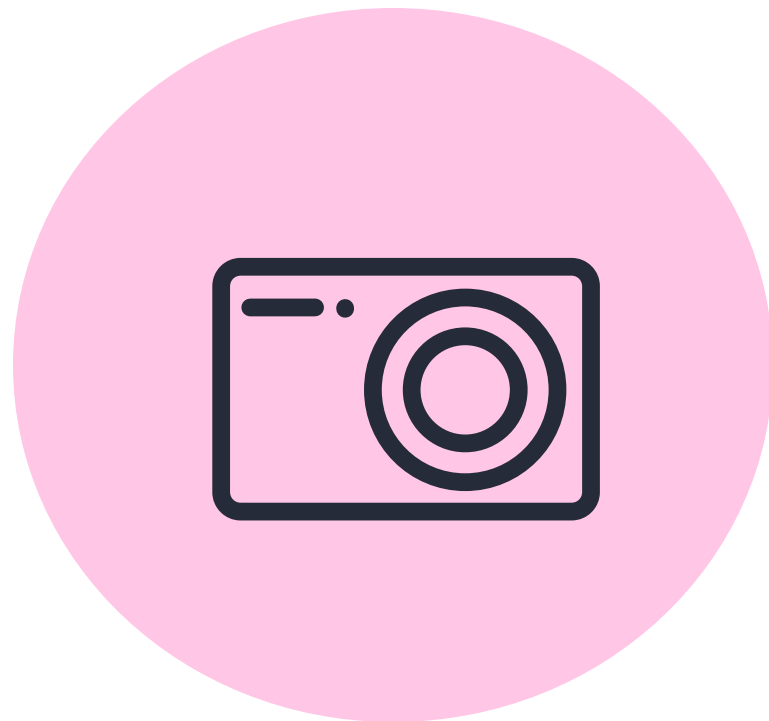
We don't have...



Free text



GP data



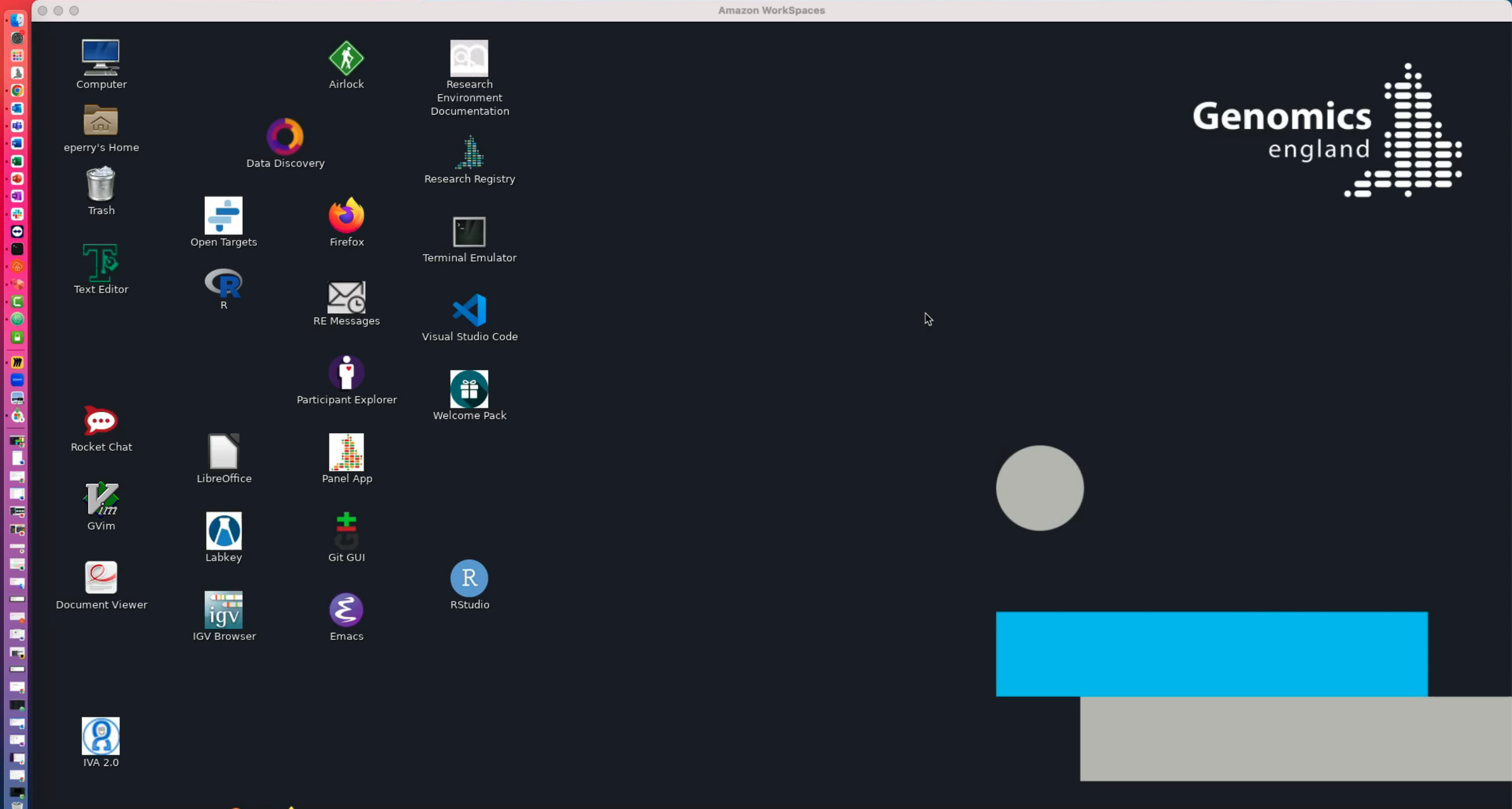
diagnostic
imaging

Diagnostic imaging data **did** (100k only)

Imaging metadata (no images)

- Dates/age: **did_date3**, **ic_age***
- Imaging type: **did_nicip_code**, **did_snomedct_code**
- Location in body: **ic_region***, **ic_sub_region***, **ic_system***, **ic_sub_sys***, **ic_morph***
- Related procedures: **ic_modality***, **ic_sub_modality***
- Source of referral: **did_patsource_code**, **did_refer_orgcode**, **ic_reftype_desc**
- Diagnosis: **ic_cancer_desc**

Contact physician demo



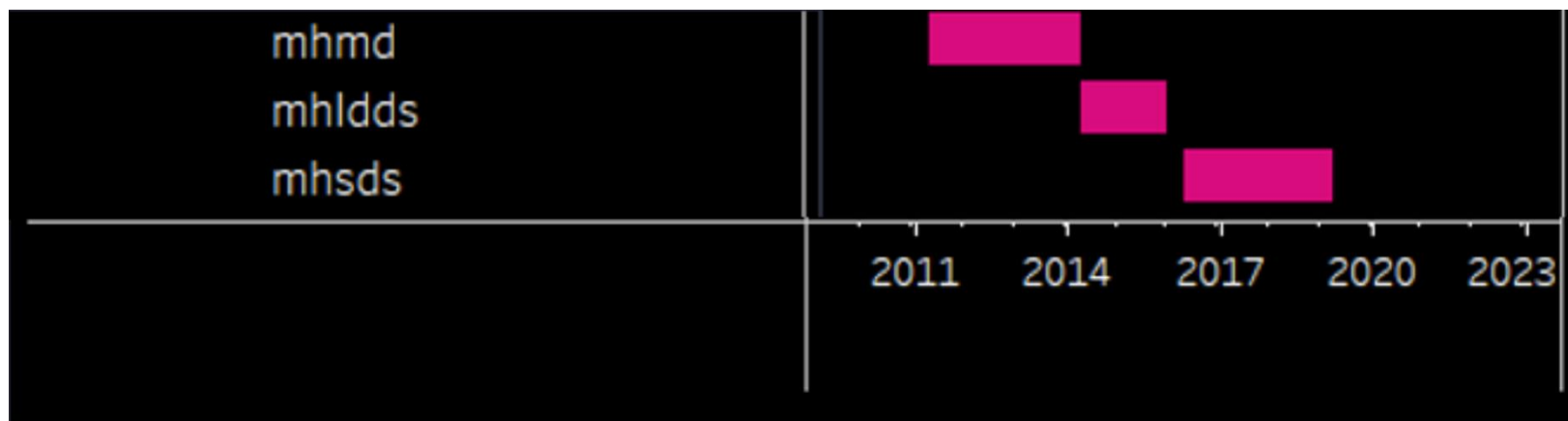
3. Mental health data in the RE

Mental health data over time (100k only)

mhmd (Mental Health Minimum Dataset) – 2011-2014

mhlds (Mental Health Learning Disabilities Dataset) – 2014-2016

mhds (Mental Health Services Dataset) – 2016 onwards



Mental health data tables (100k only)

mhmd – 2011-2014

- mhmd_v4_episode
- mhmd_v4_event
- mhmd_v4_record

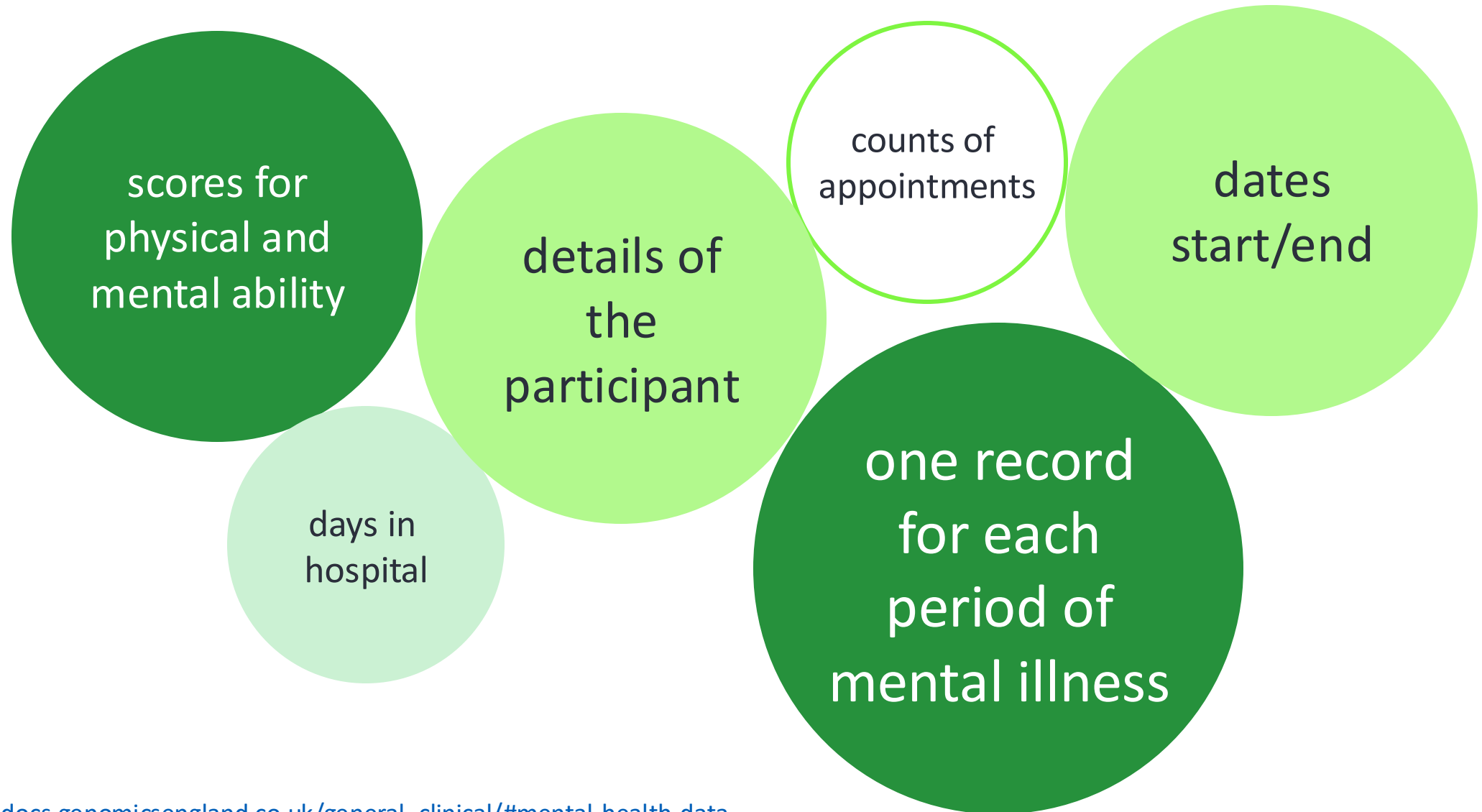
mhlds – 2014-2016

- mhlds_episode
- mhlds_event
- mhlds_record

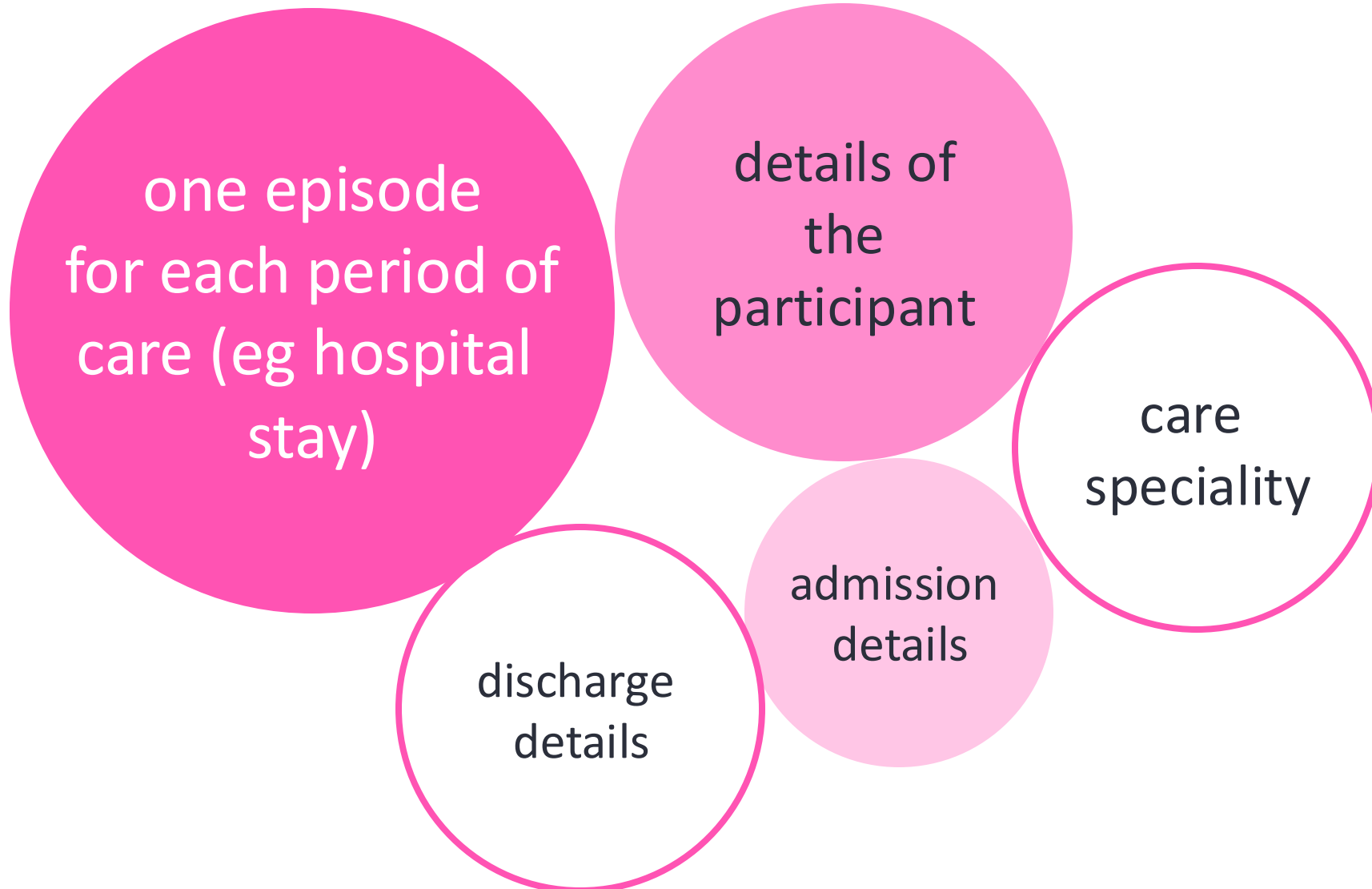
mhsds – 2016 onwards

Completely different schema

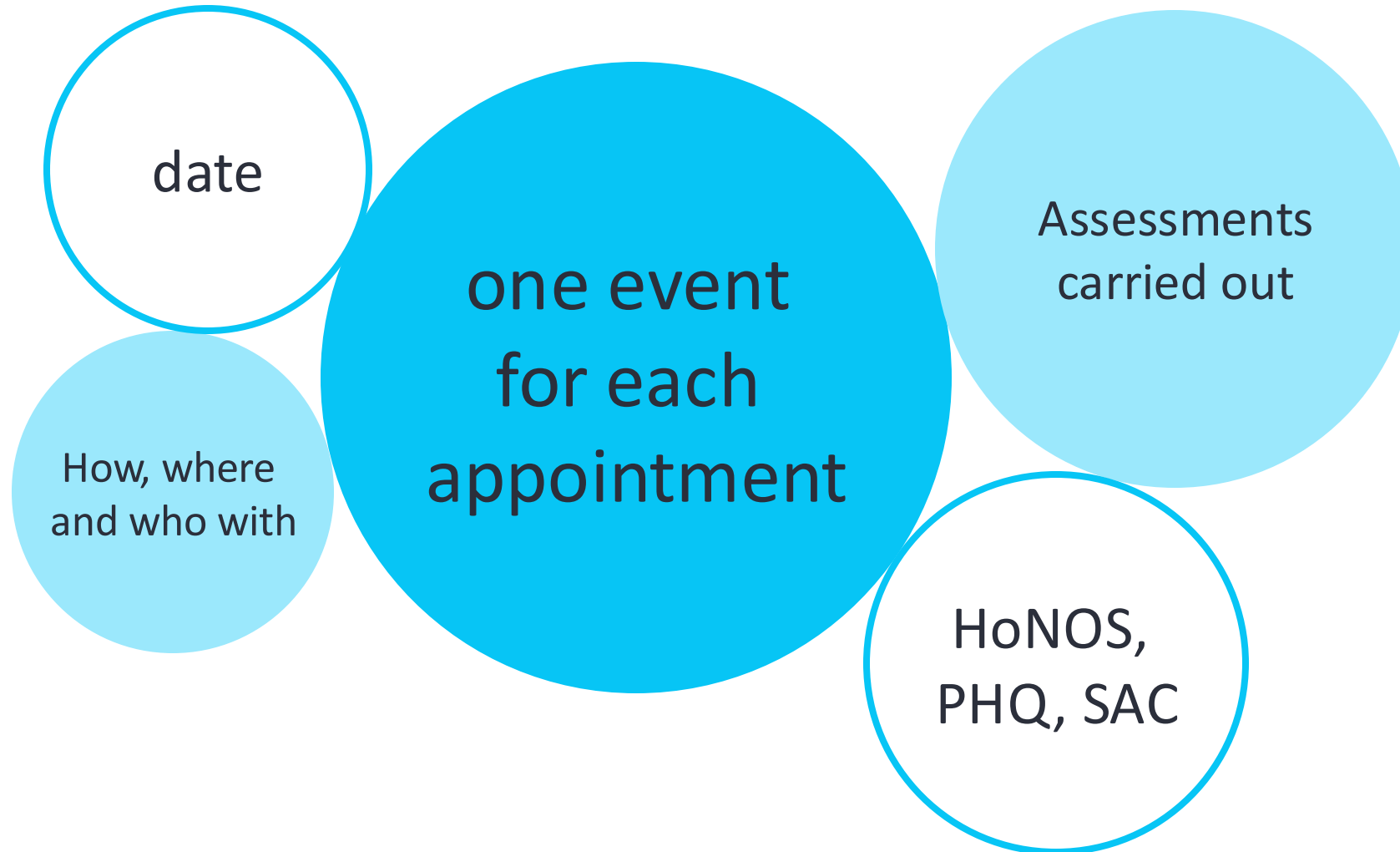
record – **mhmd** and **mhlds** (100k only)



episode – **mhmd** and **mhlds** (100k only)



event – **mhmd** and **mhlds** (100k only)



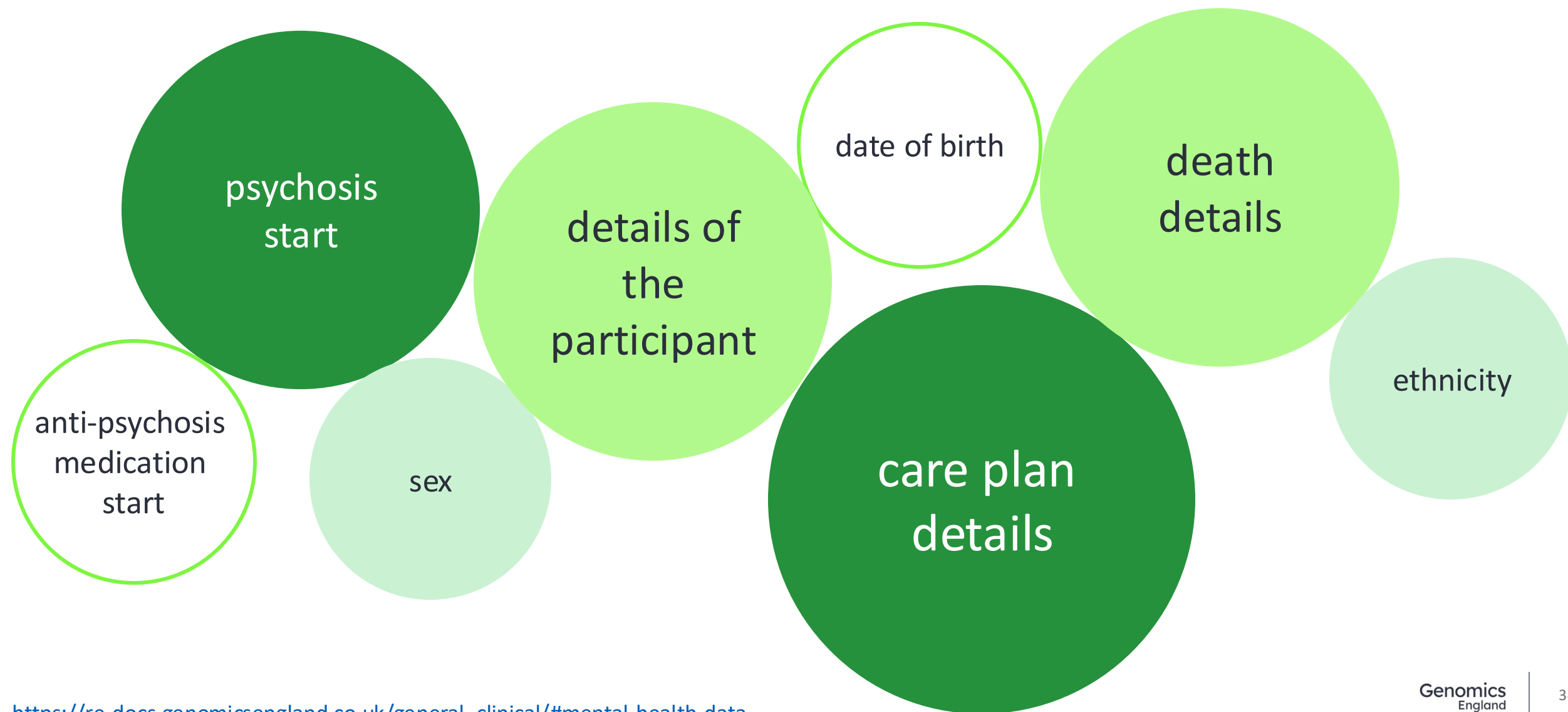
mhsds – 2016 onwards (100k only)

- mhsds_bridge
- mhsds_master_patient_index
- mhsds_gp_practice_registration
- mhsds_patient_indicators
- mhsds_care_coordinator
- mhsds_care_plan_type
- mhsds_crisis_plan
- mhsds_care_plan_agreement
- mhsds_service_or_team_referral
- mhsds_service_or_team_type_referred_to
- mhsds_other_reason_for_referral
- mhsds_referral_to_treatment
- mhsds_onward_referral
- mhsds_discharge_plan_agreement
- mhsds_care_contact
- mhsds_care_activity
- mhsds_other_in_attendance
- mhsds_indirect_activity
- mhsds_responsible_clinician_assignment
- mhsds_hospital_provider_spell
- mhsds_ward_stay
- mhsds_assigned_care_professional
- mhsds_delayed_discharge
- mhsds_hospital_provider_spell_commissioner
- mhsds_medical_history_previous_diagnosis
- mhsds_provisional_diagnosis
- mhsds_primary_diagnosis
- mhsds_secondary_diagnosis
- mhsds_coded_scored_assessment_referral
- mhsds_coded_scored_assessment_act
- mhsds_coded_scored_assessment_cont
- mhsds_care_programme_approach_care_episode
- mhsds_care_programme_approach_review
- mhsds_clustering_tool_assessment
- mhsds_coded_score_assessment_clustering_tool
- mhsds_care_cluster

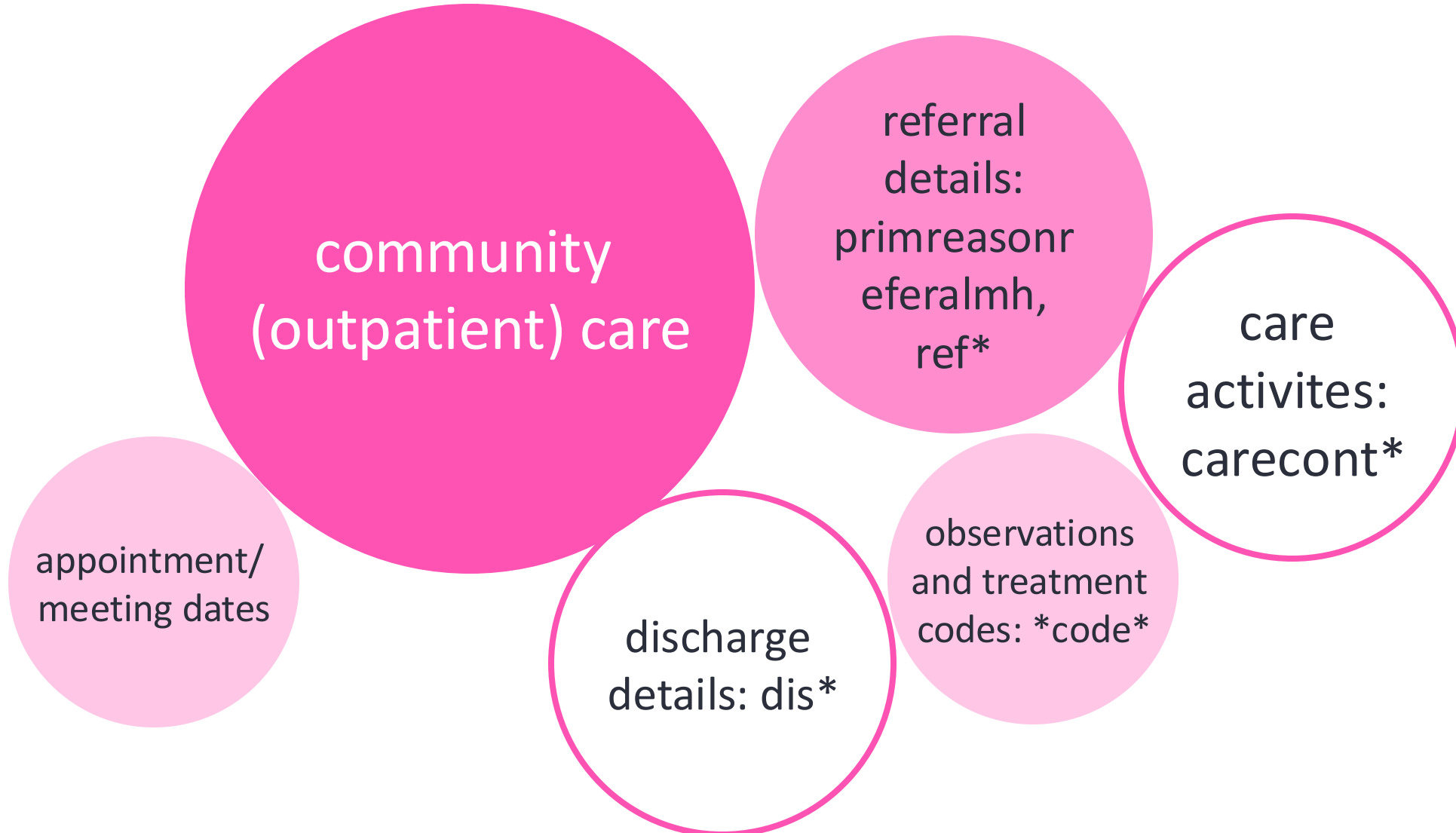
mhsds – 2016 onwards (100k only)



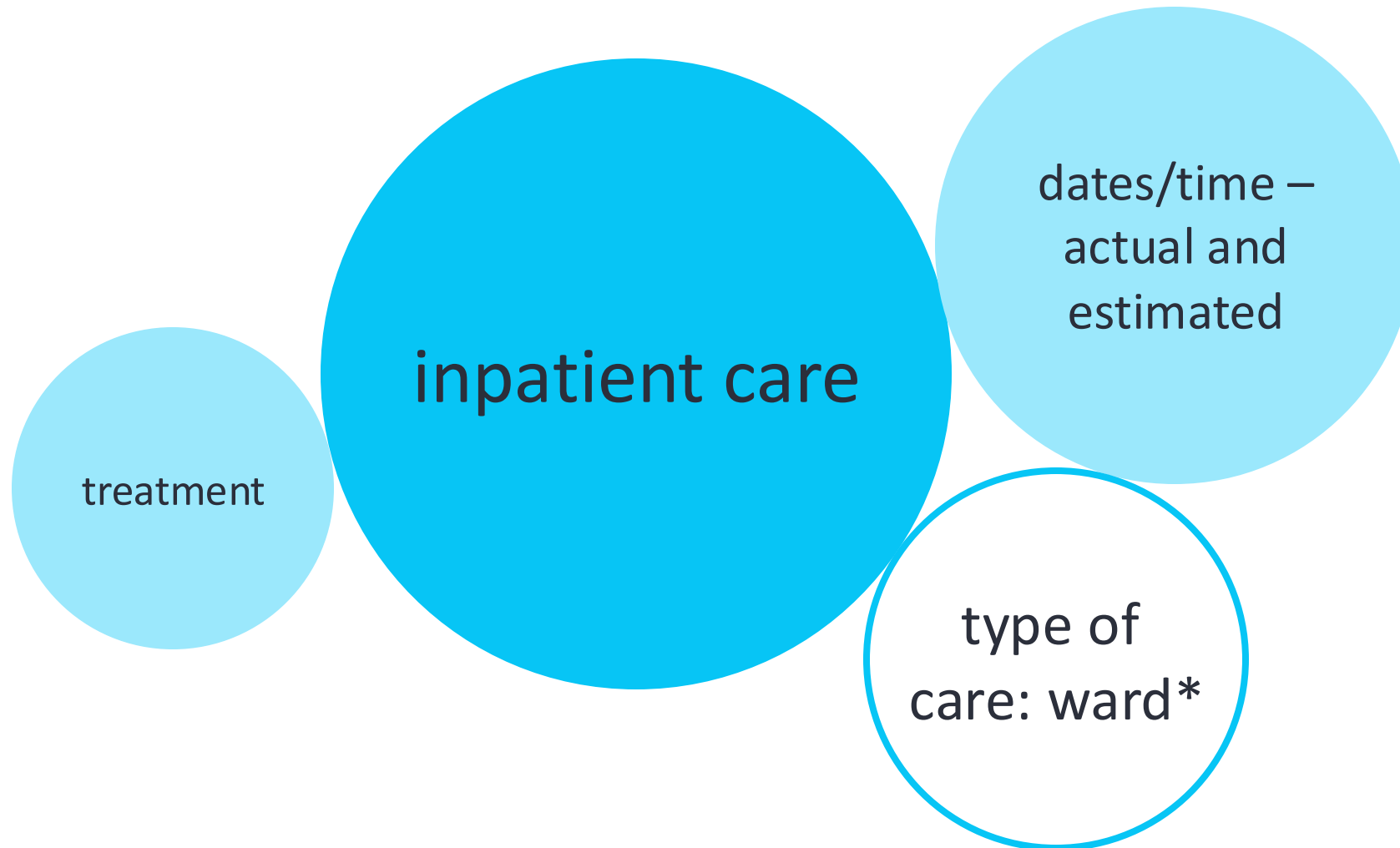
mhsds_curated_participant (100k only)



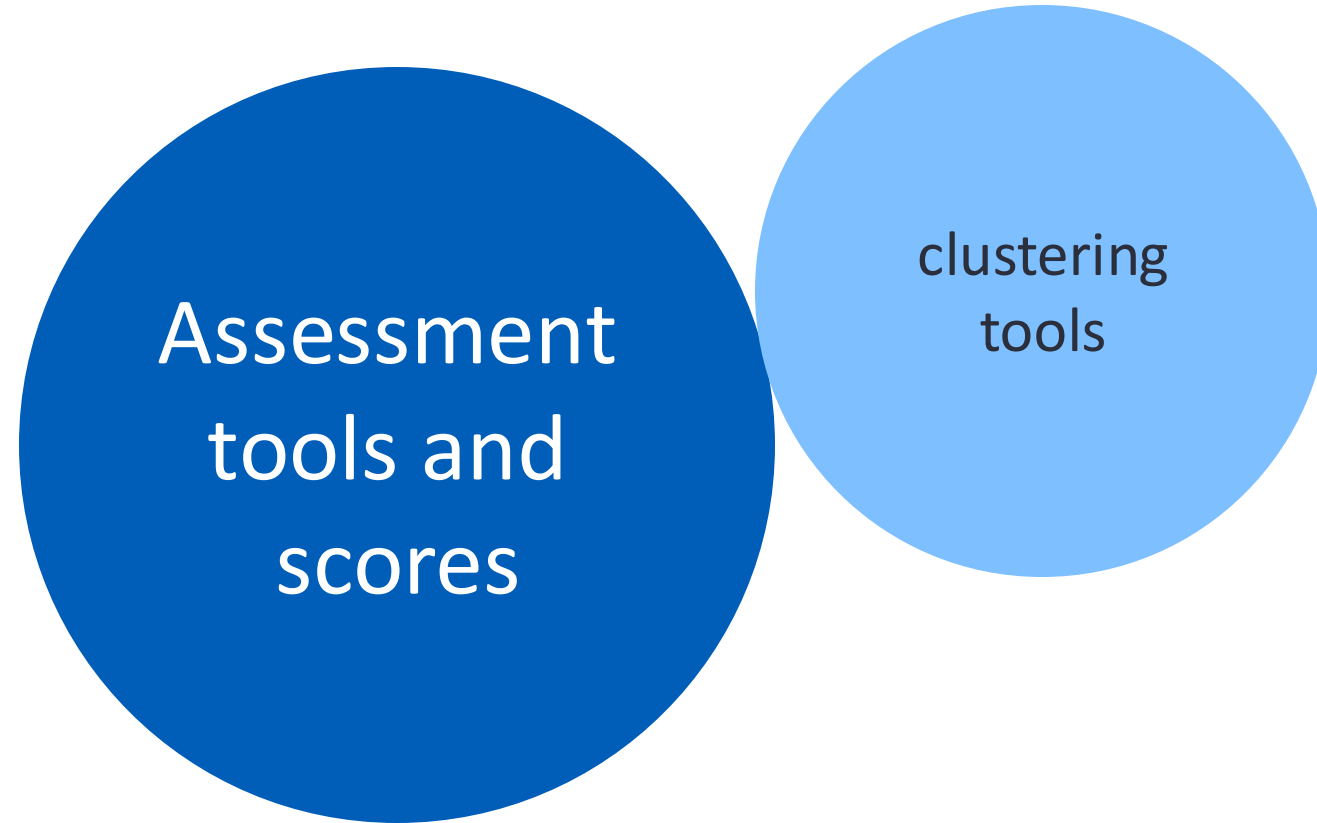
mhsds_curated_community (100k only)



mhsds_curated_inpatient (100k only)



mhsds_curated_assessment_diagnoses_and_cluster
(100k only)



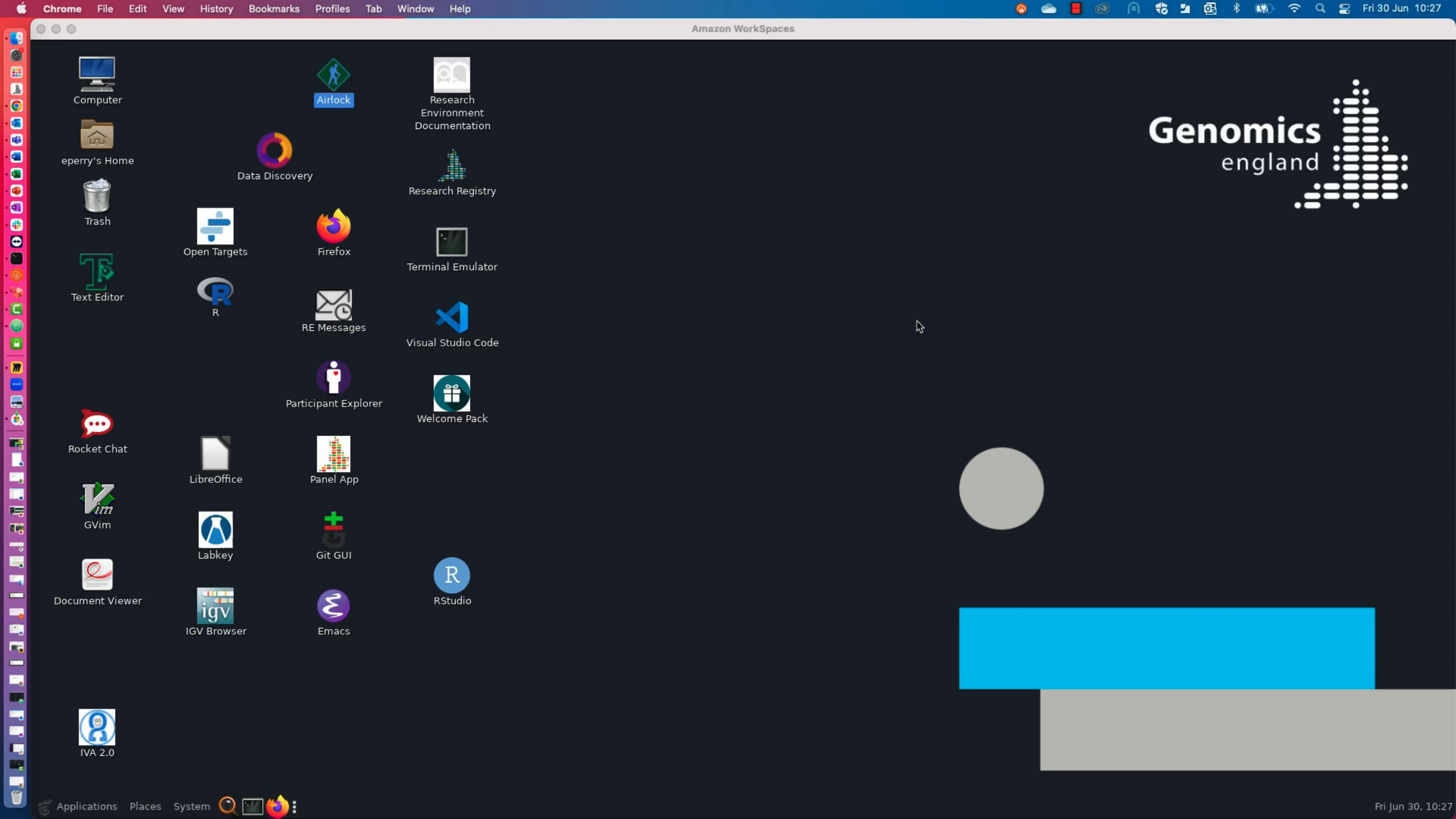
4. Accessing and comparing NHS England data with Participant Explorer

Participant Explorer

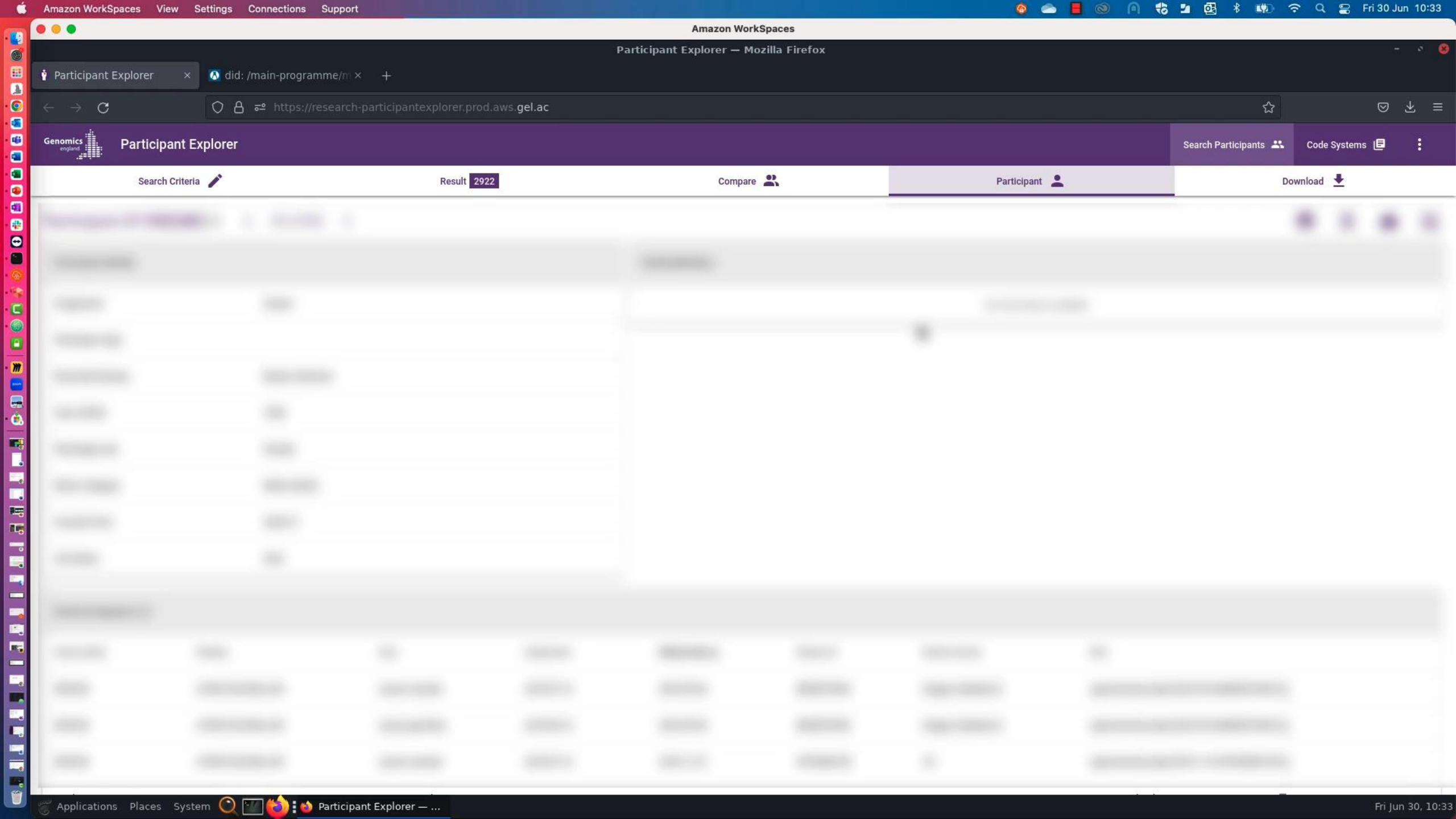
- Search for participants by:
 - Clinical concepts
 - diagnoses
 - treatments
 - Personal details
 - Identifiers
- Ontology-aware
- View/compare medical histories
 - text
 - graphical



View medical
history demo



Compare medical history demo



LabKey

- Participant details and family relationships
- Sample details
- Genomic file locations
- Clinical data
- Bioinformatics analysis results



hes tables demo

ChromeFileEditViewHistoryBookmarksProfilesTabWindowHelp

Homepage - Genomics England x

re-docs.genomicsengland.co.uk

Raise A TicketRE docsCloudRE-docsJIRA ConfluenceMulti Factor AuthChange PasswordbobSharePointRE1RE1 UGbrs_gitProfit.coProduct boardiconsshowcaseNotionGEL Academy

Genomics England Tusted Research Environment User Guide

Search

HomeGetting startedDataToolsCloud-REHigh Performance Cluster (HPC)Scripts, workflows and containersGetting helpData security and youTutorials

Home

2nd June 2023

12th May 2023

3rd May 2023

Genomics England Research Environment

→ Getting started

Data in the Research Environment

Tools in the Research Environment

High Performance Cluster (HPC)

Pre-made scripts and workflows for data analysis

Getting help

Data security and you

Tutorials

ANNOUNCEMENTS

Current data release:

/main-programme/main-programme_v17_2023-03-30

nhs-gms/nhs-gms-release_v2_2023-02-28

2nd June 2023

R packages in the AWS Research Environment

Following the transition to AWS, we encourage you to create and maintain your own personal R package libraries. See Self-service R package installation in the Research Environment for more details.

12th May 2023

Cloud-RE documentation

Documentation for the Cloud-RE has moved location. Documentation describing the CloudOS platform and its functionality is now hosted by LifeBit, while documentation describing the Genomics England data within the Cloud-RE can be found within these pages.

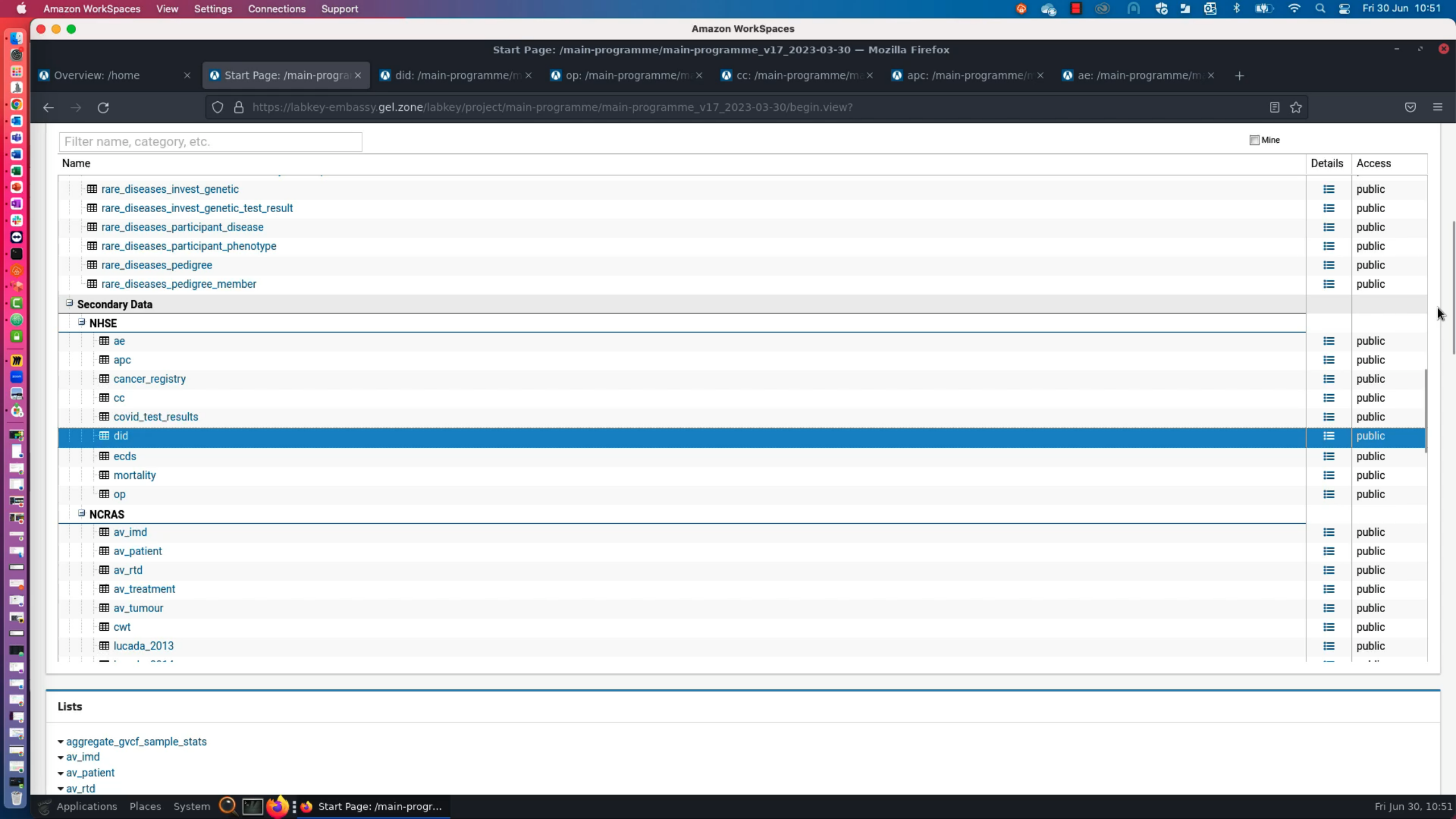
3rd May 2023

Updated workflow

the Small variant 2.0 workflow is a complete re-write of the Gene variant workflow (current version 1.7) from WDL/Cromwell to Nextflow DSL1. The

Last update: March 16, 2023

mental health tables demo



Filter name, category, etc.

Mine

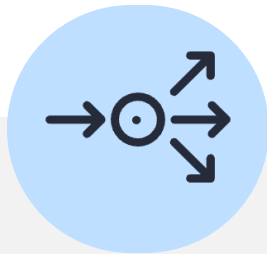
Name	Details	Access
rare_diseases_invest_genetic		public
rare_diseases_invest_genetic_test_result		public
rare_diseases_participant_disease		public
rare_diseases_participant_phenotype		public
rare_diseases_pedigree		public
rare_diseases_pedigree_member		public
Secondary Data		
NHSE		
ae		public
apc		public
cancer_registry		public
cc		public
covid_test_results		public
did		public
ecds		public
mortality		public
op		public
NCRAS		
av_imd		public
av_patient		public
av_rtd		public
av_treatment		public
av_tumour		public
cwt		public
lucada_2013		public

Lists

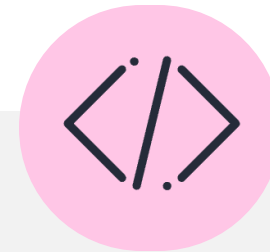
- ▼ aggregate_gvcf_sample_stats
- ▼ av_imd
- ▼ av_patient
- ▼ av_rtd

6. Accessing medical history programmatically

LabKey API



Combine queries between tables



Work in a variety of programming languages
(support for Python and R) using SQL
queries



Replicate queries between releases and
analyses



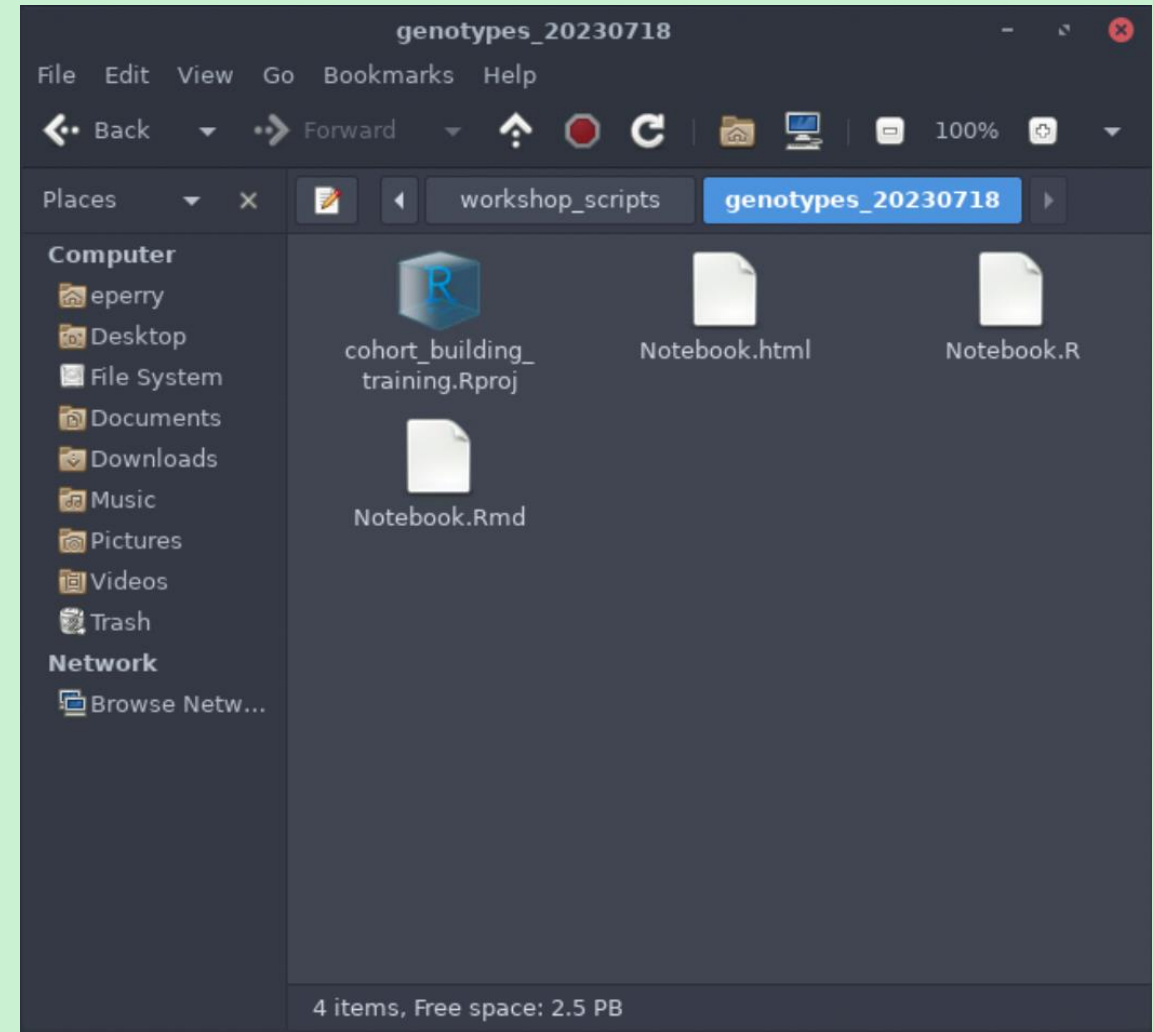
Work locally and on the HPC

LabKey .netrc

- You can access the same data via the LabKey API as you can through other means
- You will need to configure access to the LabKey API with your username and password
 - In your home directory
 - On the HPC
- You do this by editing a file called .netrc

Materials

- Slides and video will be sent out to you after the session
- Scripts available in
/gel_data_resources/example_scripts/workshop_scripts/medical_history_2025



To launch the notebooks



https://re-docs.genomicsengland.co.uk/en/able_rstudio/



https://re-docs.genomicsengland.co.uk/hpc_jupyter/



<https://lifebit.atlassian.net/wiki/spaces/CD/pages/813106134/Query+the+database+to+build+a+cohort>

programmatic access demo

Chrome

File

Edit

View

History

Bookmarks

Profiles

Tab

Window

Help

Amazon WorkSpaces

Lifebit CloudOS

medical_hist... - JupyterL

+

←

→

↺

https://pro.cloud-os.prod.aws.gel.ac/app/interactive-analysis/running

☆

📧

📁

☰

Interactive Analyses

Save snapshot

Save

Stop

Jupyter session: medical_history_training

+

Add time limit

+

Add cost limit

STATUS: RUNNING

🔄

Click to expand details

⚠

>

File

Edit

View

Run

Kernel

Tabs

Settings

Help

Maximise

jovyan@db01e0f21c49: -X

medical_history_CloudOS

+

+

✂

📄

📄

▶

■

🔄

▶▶

Code

⚙

R (libraries-cloudos)

🔍

[]:

📄

↑

↓

📄

🔍

⚙

Getting medical histories for participants

This notebook will walk you through getting medical histories for participants using the LabKey API in R. You are welcome to copy/paste any code from this notebook for your own scripts/notebooks.

Contents:

- Import R libraries you need
- Helper function to access the LabKey API with R
- Get details from a single `hes_*` table for a participant
- Plot full medical history
 - Find all diagnoses
 - Find all investigations
 - Find all treatments
- Link events in medical history
 - Match accident and emergency episodes to overnight admissions
 - Find critical care periods within an overnight stay
 - Find hospital day appointments before or after another type of admission
 - ECDS
- Mental health data
- Optional exercise

Simple

1

1

1

R (libraries-cloudo...

Mode: Com...

Ln 1, ...

medical_history_Cloud...

0

🔔

R Notebook

+

←

→

↺

file:///nas/weka.gel.zone/pgen_int_work/BRS/emily/medical_history_2024

☆

📧

📁

☰

R Notebook

Getting medical histories for participants

This notebook will walk you through getting medical histories for participants using the LabKey API in R. You are welcome to copy/paste any code from this notebook for your own scripts/notebooks.

Contents:

- Import R libraries you need
- Helper function to access the LabKey API with R
- Get details from a single `hes_*` table for a participant
- Plot full medical history
 - Find all diagnoses
 - Find all investigations
 - Find all treatments
- Link events in medical history
 - Match accident and emergency episodes to overnight admissions
 - Find critical care periods within an overnight stay
 - Find hospital day appointments before or after another type of admission
 - Find details of diagnostic imaging within a time period
 - ECDS
- Mental health data
- Optional exercise

Import R libraries you need

library(tidyverse)

— Attaching core tidyverse packages — tidyverse 2.0.0 —
✓ dplyr 1.1.2 ✓ readr 2.1.4
✓ forcats 1.0.0 ✓ stringr 1.5.0
✓ ggplot2 3.4.2 ✓ tibble 3.2.1
✓ lubridate 1.9.2 ✓ tidyr 1.3.0
✓ purrr 1.0.1
— Conflicts — tidyverse_conflicts() —
* dplyr::filter() masks stats::filter()
* dplyr::lag() masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

Applications

Places

System

🔍

🖥

📧

eperry@a-3en38gisq1q...

📁

[medical_history_traini...

🔥

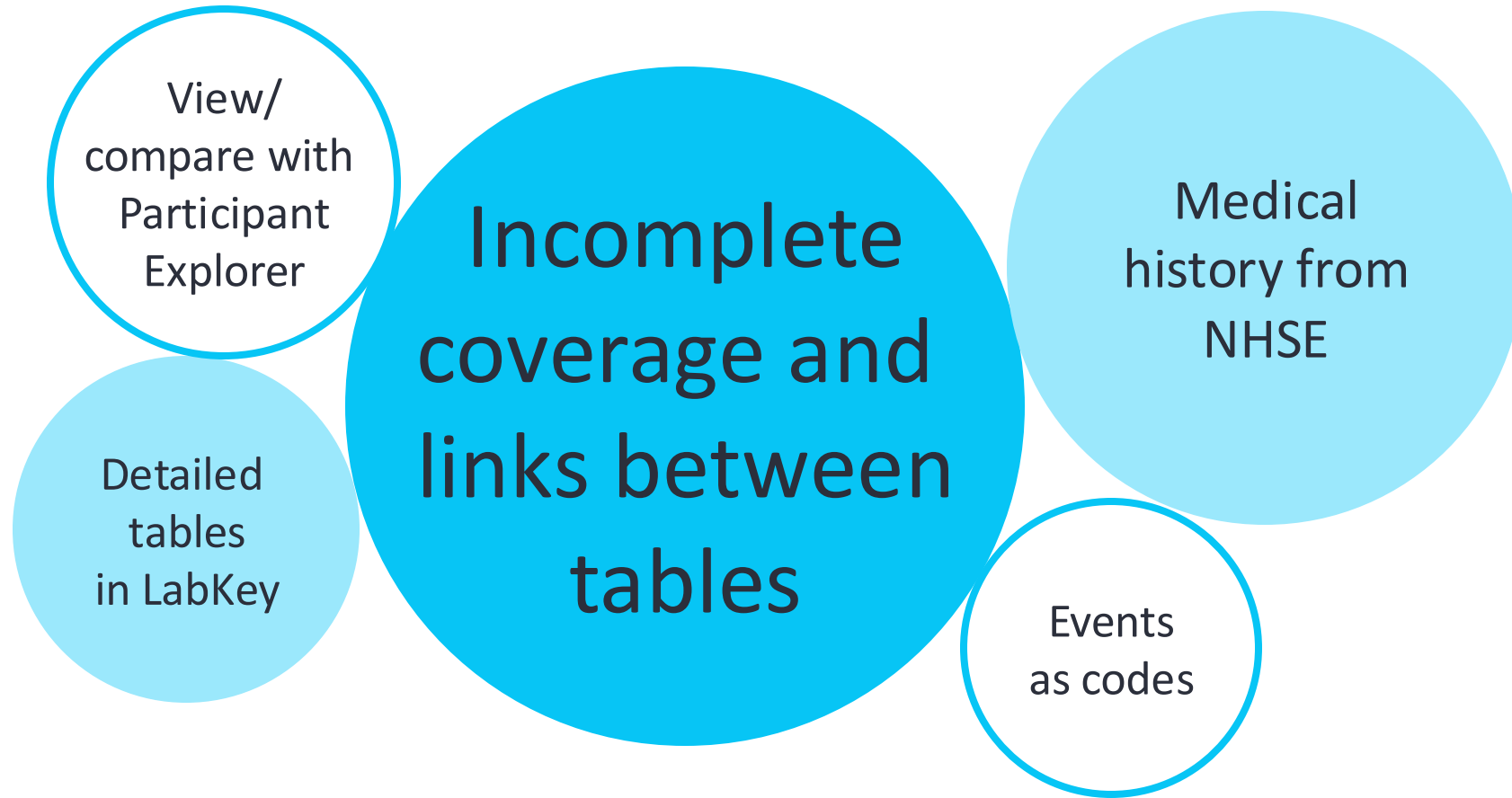
Lifebit CloudOS — Mozil...

🔥

R Notebook — Mozilla Fi...

Mon Jun 17, 11:08

Key takeaways



7. Help and questions

Getting help



Check our documentation:
<https://re-docs.genomicsengland.co.uk/>
Click on the documentation icon in the environment



Contact our Service Desk:
<https://jiraservicedesk.extge.co.uk/plugins/servlet/desk>

Questions



All your
microphones
are muted



Use the Zoom
Q&A to ask
questions



Upvote your
favourite
questions: if we
are short on
time we will
prioritise those
with the most
votes

Training sessions

3rd Tuesday every month

Introduction to the RE

16/9

21/10

18/11

16/12



Materials from
past training
all online

Training sessions

14/10

Using GEL data for publications and reports

11/11

What tools and workflows should I use to fulfil an overall goal?

9/12

Running workflows on the HPC and Cloud



Materials from
past training
all online

Feedback



Thank you

Visit: <https://re-docs.genomicsengland.co.uk/>