

DataLoch: enabling quality research through a secure, pioneering data service

Stuart Dunbar (1), Zengyi Huang (1), Sara Hatam (1), Arlene Casey (1,2), Atul Anand (1,3), Kathy Harrison (1), Nicholas Mills (1,3)

(1) DataLoch, Usher Institute, University of Edinburgh; (2) ACRC, Usher Institute, University of Edinburgh; (3) Centre for Cardiovascular Science, University of Edinburgh

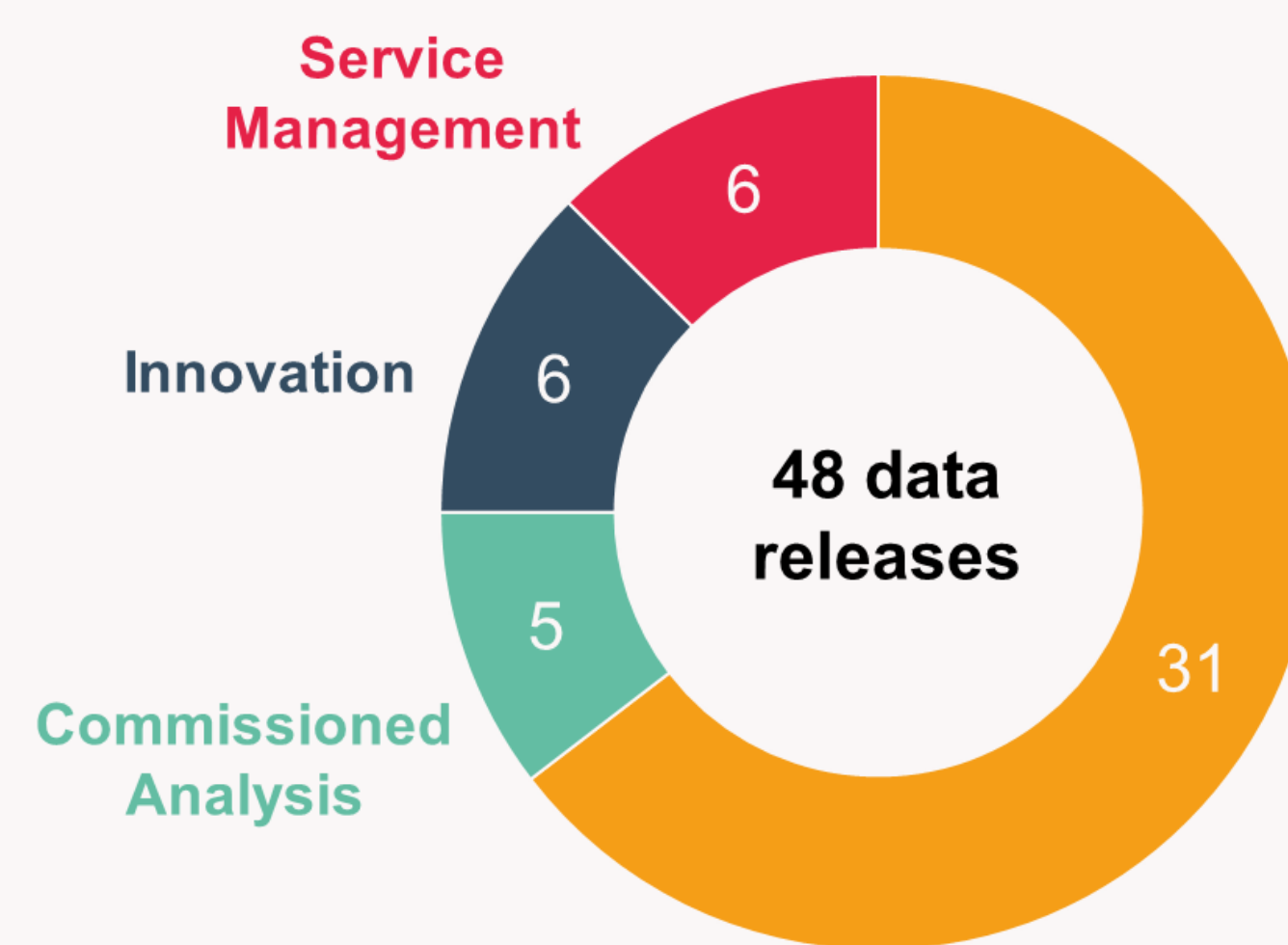
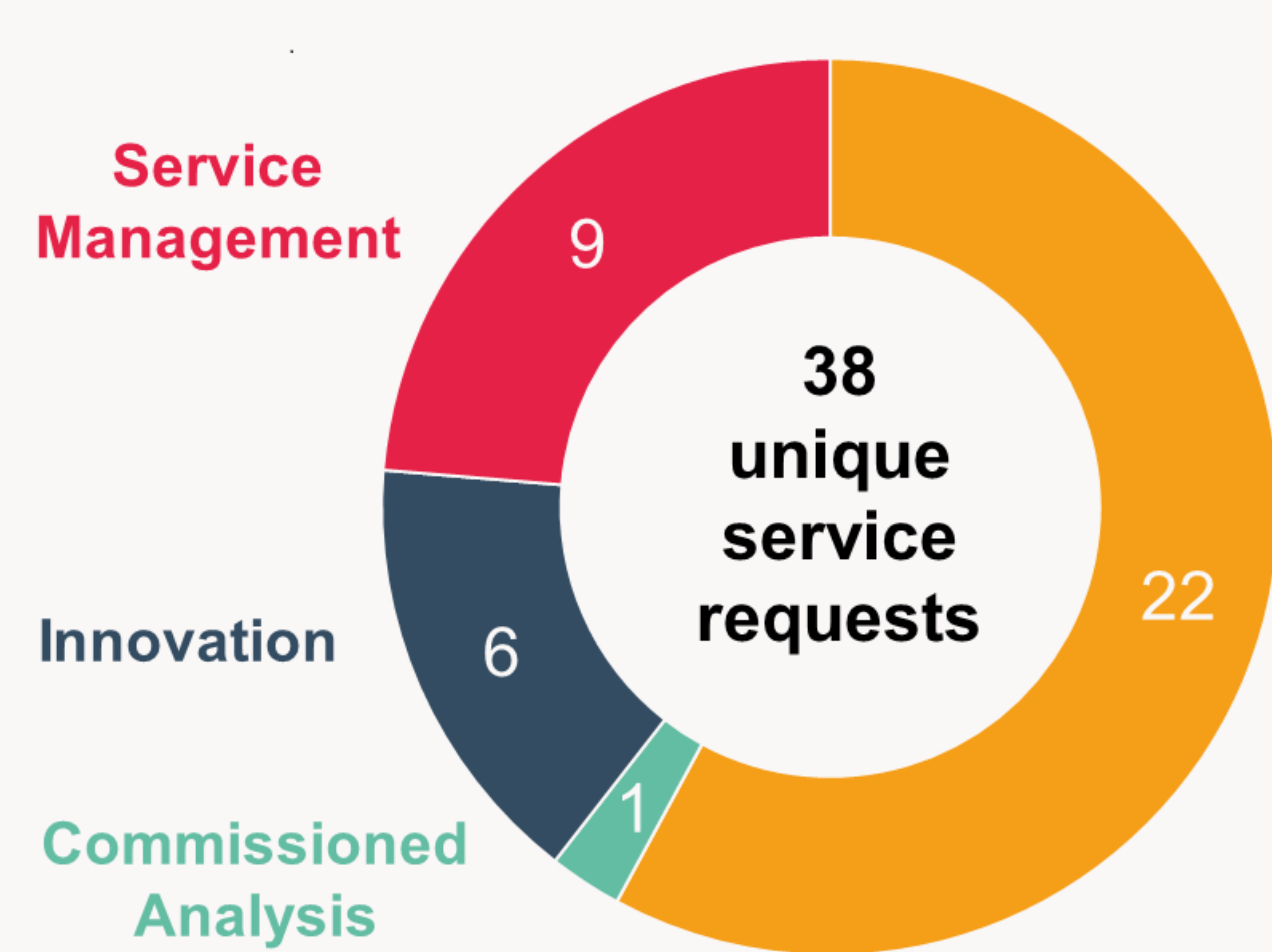
What is DataLoch?

One of the Regional Safe Haven organisations in Scotland, DataLoch is a data service that has been developed in partnership by the University of Edinburgh and NHS Lothian. Our role includes:

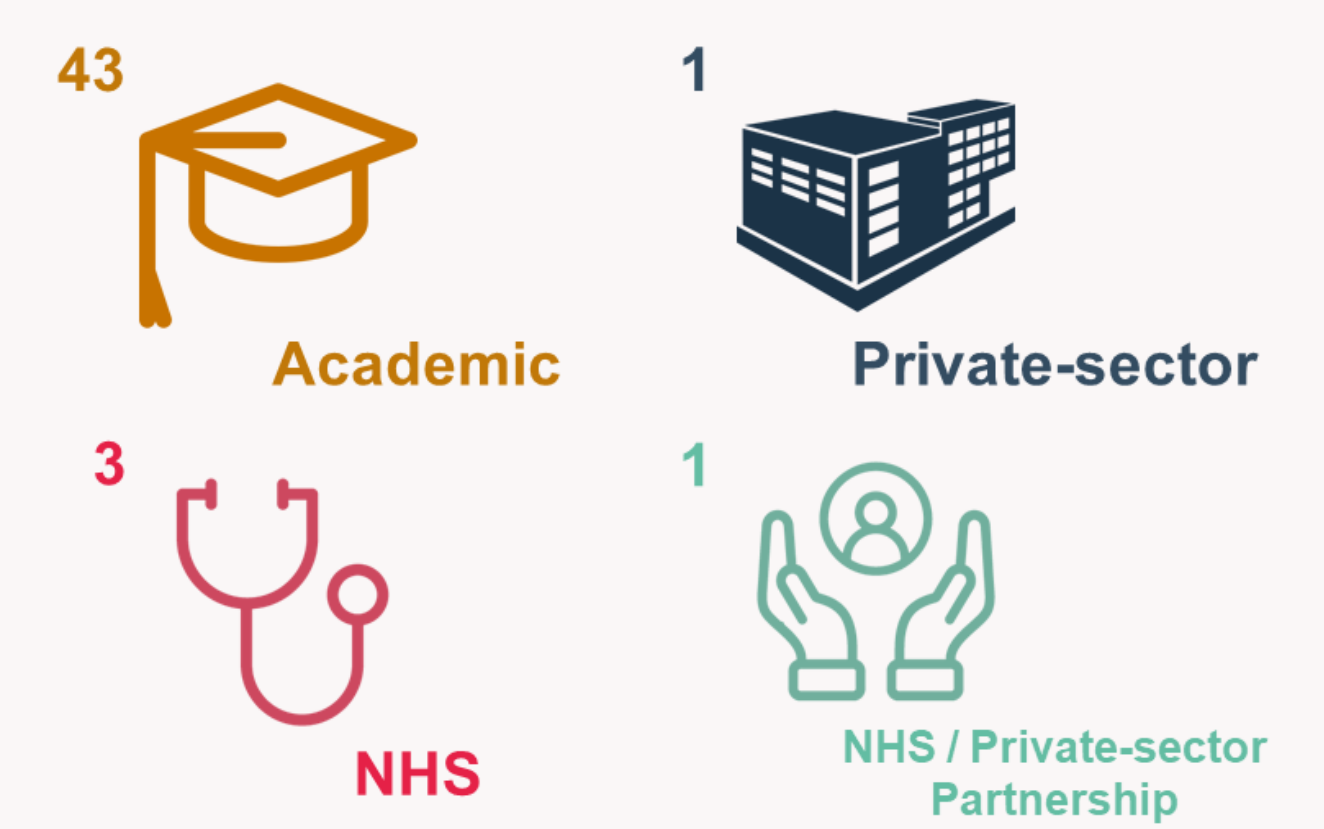
- bringing together health and social care data for the South-East Scotland region;
- working with experts in health and social care to understand and improve this data; and
- providing safe access to data for researchers through our secure data environment.

Working collaboratively, these efforts allow a holistic data-driven approach to the prevention and treatment of different conditions, as well as the provision of health and social care services more broadly.

Our 2022 in numbers



Number of data releases by organisation



Some images sourced from Flaticon

Streamlining data for research

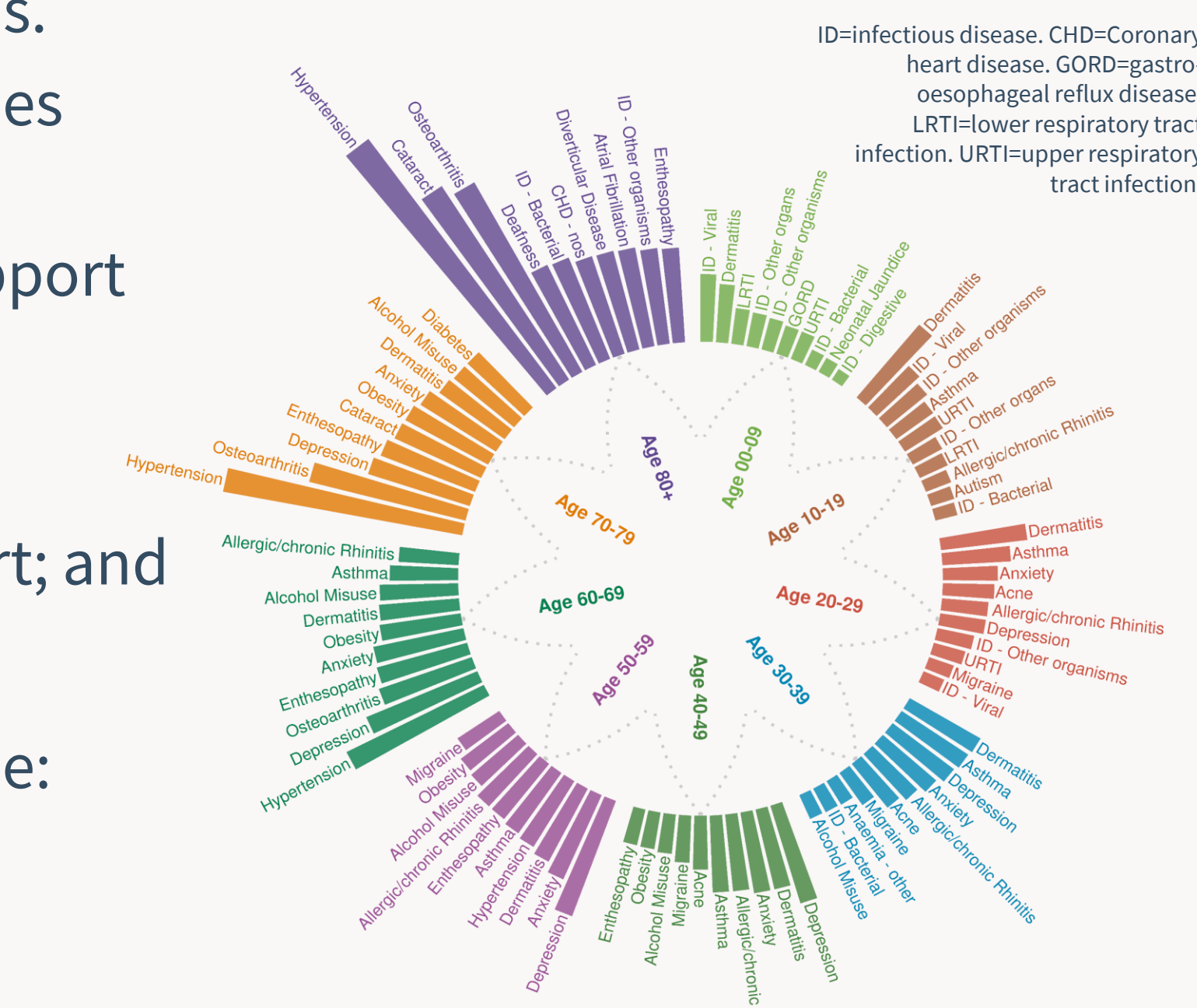
A key element in our work is the transformation of data to enable research to take place more easily, yet still maintain data security and confidentiality.

One example is through incorporating **HDR UK CALIBER phenotyping**. Each phenotype contains codelists of all the possible ways to describe a condition across GP, hospital, and death records.

This implementation harmonises codes across primary and secondary care systems. Our phenotype datasets support researchers to:

- define populations of interest;
- define the baseline data of a cohort; and
- provide outcomes for a cohort.

Explore this work through the QR code:



10 most prevalent conditions in Lothian by age group

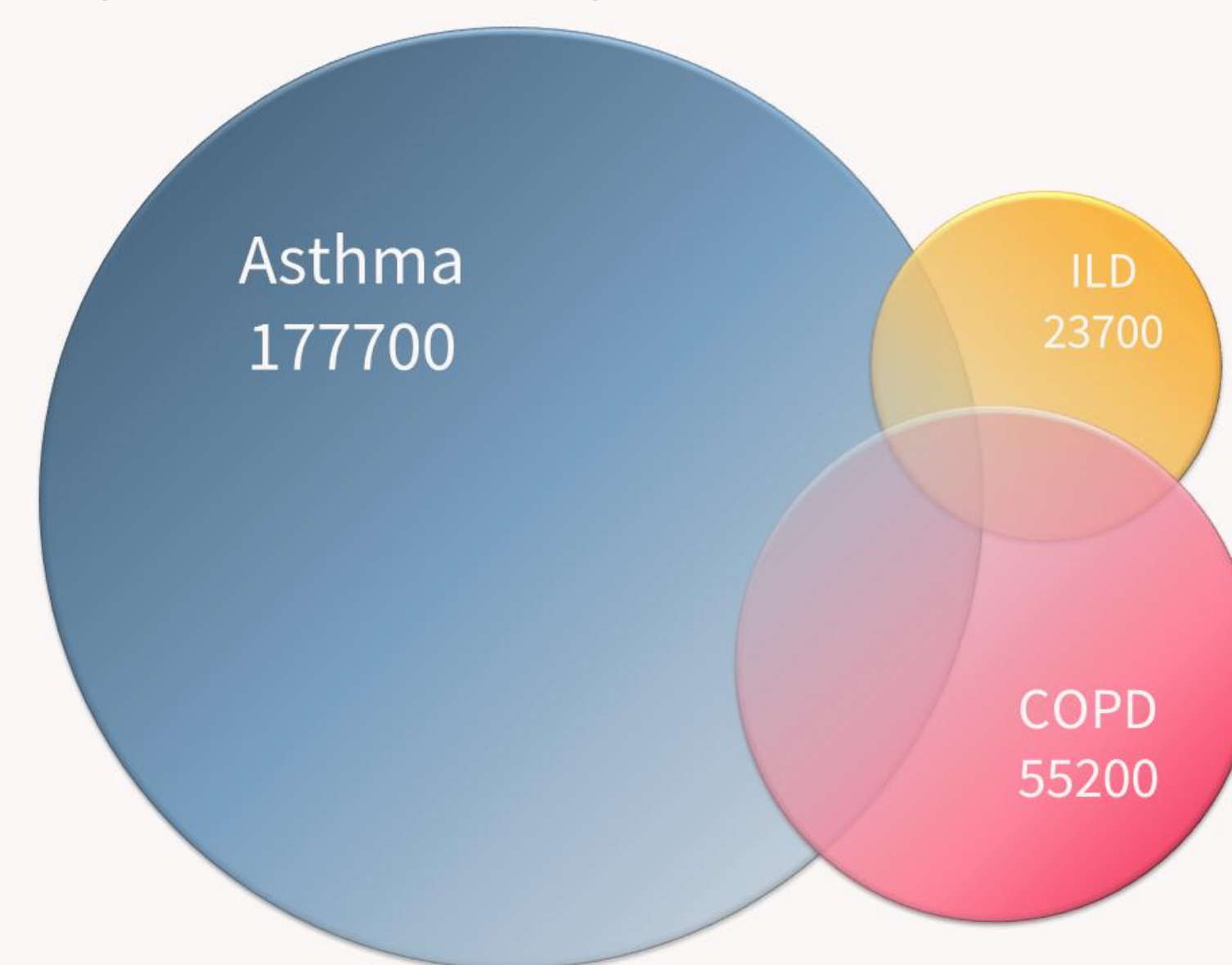
DataLoch Registries: focussed research support

Our registries are theme-specific databases (defined by one or more related conditions or diseases) validated by clinical and academic experts. Pre-defined cohorts streamline the application process without compromising data security.

A result of close collaboration with the BREATHE Hub and Health Data Research UK, the **DataLoch Respiratory Registry** is one of three related registries alongside those in England and Wales, which offers significant opportunities for three-nation research.

Plans are in place for registries on **Heart Disease, Ageing, and Mental Health**.

Discover more about our Respiratory Registry through the QR code:



Primary and secondary care condition cohorts as of May 2023

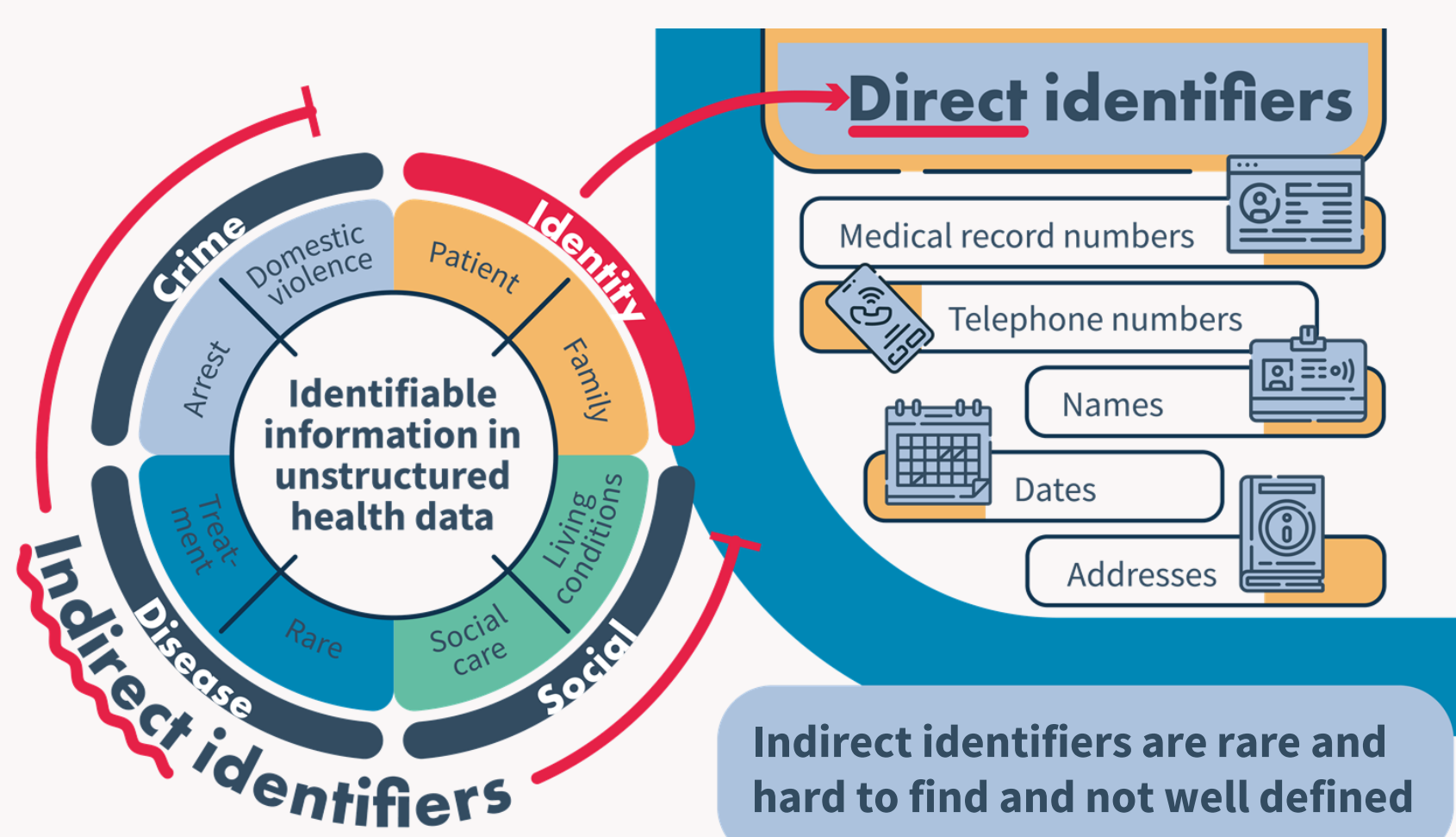


Understanding risk

Privacy risk assessment is vital to ensure that health data made available for research do not contain identifiable information.

However, risk assessments are usually manual and time-consuming. For unstructured data in particular, frameworks for assessing privacy risks are not well defined.

Through the application of Natural Language Processing techniques on hospital discharge summaries, we are exploring how privacy risks can accumulate as a basis for **developing a privacy risk dashboard**. This dashboard will support information governance teams by making the process of minimising privacy risks more robust. For further information, scan the QR code:



Indirect identifiers are rare and hard to find and not well defined

Public Perspectives

Since our inception, we have prioritised the inclusion of public perspectives on health data research.

Our **Public Reference Group (PRG)** continues to offer important feedback on our development and next steps.

More specifically, PRG members instigated and co-designed **the Public Value Assessment** process with us. For any application requiring an ethics review, there is also a Public Value Assessment which must secure approval.

Beyond this group of committed local residents, we run **public consultations** with Ipsos Scotland on specific questions where public perspectives are critical.

This approach has been core to our privacy risk assessment work (see left). Furthermore, in advance of our full launch in 2022, we sought public views to inform the expansion of our information governance framework to include researchers beyond academic and clinical settings (see graph above).

