DataLoch: supporting the responsible and secure use of data to improve health and social care

Pamela Linksted (1), Kathy Harrison (2), Atul Anand (2), Arlene Casey (2), Stuart Dunbar (2), Zengyi Huang (2), Sara Hatam (2), Franz Gruber (2), Elvina Gountouna (2), Derek O'Sullivan (2), Stephen Powell (2), Nicholas Mills (2).

(1) NHS Lothian, (2) University of Edinburgh, contact: dataloch@ed.ac.uk

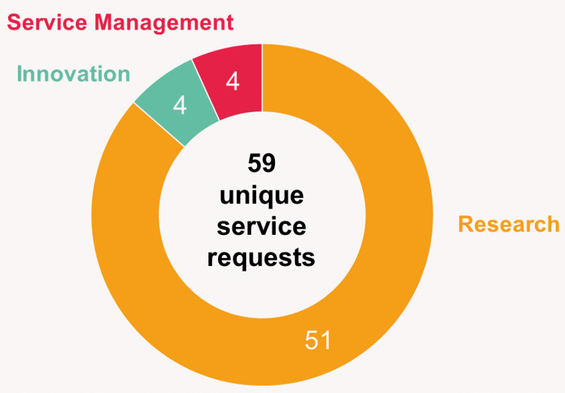
What is DataLoch?

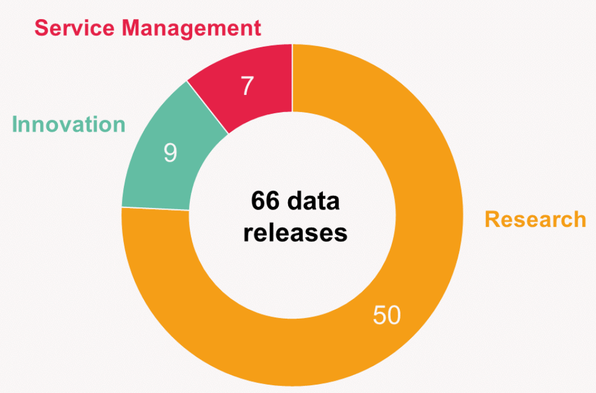
One of Scotland's Regional Safe Havens, DataLoch is a data service developed in partnership by the University of Edinburgh and NHS Lothian, underpinned by public involvement and engagement. We:

- bring together health and social care data for the South-East Scotland region;
- work with experts in health and social care to understand and improve these data; and
- provide safe access to data for health data research through our Secure Data Environment and for clinical studies.

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. Ultimately, we support projects that improve the health and wellbeing of the nation.

Our 2024 in numbers





Some images sourced from Flaticon

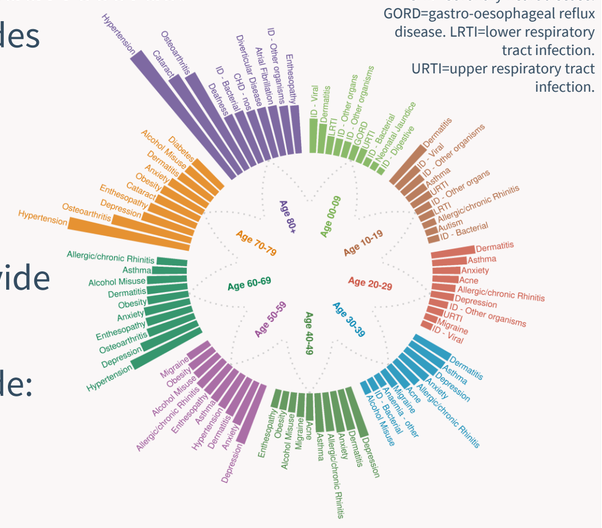
Streamlining data for research

A key element in our work is to securely curate research-ready data, while maintaining individual confidentiality.

One example is through incorporating **Health Data Research (HDR) UK CALIBER phenotyping**. Each phenotype contains codelists that 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: a) define populations of interest; b) define the baseline data of a cohort; and c) provide outcomes for a cohort.

Explore this work through the QR code:

ID=infectious disease. CHD=Coronary heart disease. GORD=gastro-oesophageal reflux disease. LRTI=lower respiratory tract infection. URTI=upper respiratory tract infection.



10 most prevalent conditions in Lothian by age group

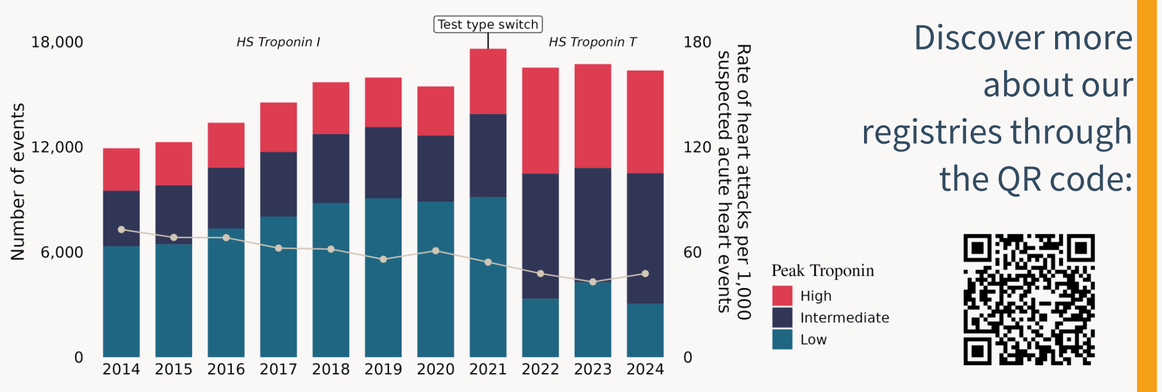
DataLoch Registries

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

Through close collaborations with clinicians and researchers, we have recently developed registries focused on **Heart Disease** and **Respiratory** conditions, which accelerate the opportunities for innovative studies in these specialist areas. Plans are in place for registries on **Ageing** as well as **Mental Health**.

Cardiac troponin testing in NHS Lothian Emergency Departments

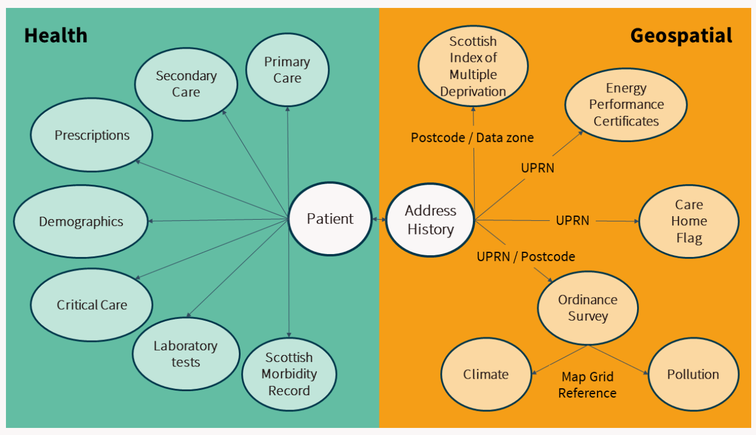
Total: 166,400 suspected acute heart events; 99,100 patients



Discover more about our registries through the QR code:

Geospatial data

At DataLoch, we are exploring the linkage of health data with various geospatial datasets to provide opportunities to investigate health impacts of:



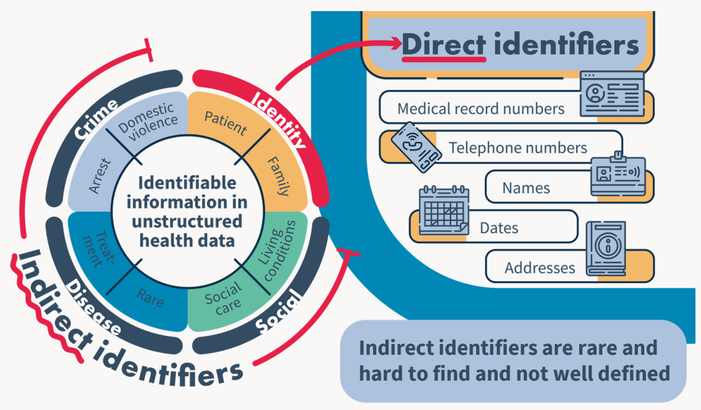
1) different types of **home heating, insulation,** etc. from the household Energy Performance Certificate dataset;

2) various **pollutants and climate conditions** from publicly available pollution and climate datasets; and

3) living in a **care home**.

This work is based on accurate identification of addresses, which are then linked to Unique Property Reference Numbers (UPRN) and Grid Map References. Data are de-identified before researcher access.

Natural Language Processing (NLP)



In comparison with coded data, clinical free-text holds a broader spectrum of patient experiences. However, it also includes a multitude of potential **direct** and **indirect identifiers** (see graphic), which currently prevents the use of free-text extracts within health research and innovation.

By applying NLP techniques on free-text, we are developing the processes to remove direct and indirect identifiers so that the de-identified data can be made available for research without compromising confidentiality.

We are also developing methods to convert free-text into coded, structured data that are research ready. Find out more through the QR code.

Download this poster:

NHS Lothian | The University of Edinburgh | usher institute