Managing and Sharing Access to Linked Datasets: the CALIBER Programme

Kenan Direk
Farr Institute of Health Informatics Research
University College London
Outline

– UCL Farr Institute
  • Organisation
  • What we do
  • Partners

– CALIBER
  • The resource
  • Data access management
  • Information Governance

– Data Safe Haven
  • Governance
  • Infrastructure
  • Processes
UCL Farr Institute

Harnessing Data for Health Science and e-Health Innovation

Farr Institute created in 2013 (4 distributed hubs across UK)

UK-wide research collaboration involving 21 academic institutions and health partners in England, Scotland and Wales

Institute aims to optimize the use of health records in research and address the UK’s capacity building requirements to support a sustainable health informatics research base
Partnerships and collaborations

Public & Patients
- General Public
- Patient Advocacy Groups
- General & Nurse Practitioners
- Third Sector

Industry
- Strategic Multipliers: ABPI, ABHI, One Nucleus etc
- Corporate IT
- Small to Medium Sized Innovators
- Big Pharma
- Medical Devices

Data Custodians
- Genomics England
- UK Biobank
- NHS Digital
- Clinical Practice Research Datalink
- Department of Health
International partnerships
CALIBER

- Research platform and resources linking national EHR across healthcare settings in ~10m patients
- Established 2009

Aims
- Enable use of national EHR for research across the translational pathway
- Stimulate a community of methodological and technical tool developers
- Facilitate reproducible and replicable research

65 ISAC protocols approved, >20 in preparation
- ~75 peer-reviewed journal publications
- >500 phenotype components, ~30 disease phenotypes

>130 registered Data Portal users contributing phenotypes
- >10 countries
- ~13 HEIs
- ~25 research groups
- Industry: Astra Zeneca, Evidera, [...]

-
Linked EHR across care settings

Clinical Practice DataLink (CPRD)

- Provides anonymised data to public health researchers.
- Ability to link primary care data to other health datasets, i.e. secondary care and disease registries (e.g. cancer)
The fun part:

- Linked EHR across care settings
- Patient identifiers
- EHR data
- De-identified EHR Data (DED)
- Research Ethics Committee approval
- NHS Information Governance Toolkit
- CPRD
- De-identified Other Data (DOD)
- Other health data
- Encrypted Linker Key
- Anonymised data
- Public health researcher
- Section 251 approval
- Independent Scientific Advisory Committee
- Contractual compliance with CPRD
Independent Scientific Advisory Committee

- Researcher submits:
  - Application form
  - Research protocol with objectives, methods and analysis plan
  - CV
  - Data specification form
- Approval/feedback from CPRD in ~ 4-6 weeks
Data security incident trends

What action we've taken in Q1, what you've reported to us and what you can do to stay secure

Data security incidents, which are breaches of the seventh data protection principle or personal data breaches reported under the Privacy and Electronic Communications Regulations, are a major concern for those affected and a key area of action for the ICO. We have published this information to help organisations understand what we’re seeing and help them to take appropriate action.

What action we've taken in Q1

**Five fines**

We fined:

- *Construction Materials Online Ltd*, £55,000;
- *Greater Manchester Police*, £150,000;
- *Basildon Borough Council*, £150,000;
- *Gloucester City Council*, £100,000; and
- *Boomerang Video Ltd*, £60,000

Royal Free - Google DeepMind trial failed to comply with data protection law

*Date:* 03 July 2017

*Type:* News

The ICO has ruled the Royal Free NHS Foundation Trust failed to comply with the Data Protection Act when it provided patient details to Google DeepMind. The Trust provided personal data of around 1.6 million patients as part of a trial to test an alert, diagnosis and detection system for acute kidney injury. But an ICO investigation found several shortcomings in how the data was handled, including that patients were not adequately informed that their data would be used as part of the trial.

The Trust has been asked to commit to changes ensuring it is acting in line with the law by signing an undertaking.

Elizabeth Denham, Information Commissioner, said:
Data Governance and Sharing

Current Governance Tapestry

- Meaningfully Engage
- Foster Trust
- Understand Expectations

Source Information
Direct collection from participants
- interviews
- surveys
- device data collection

Ethics Review

NHS
Public Health England

Other Data Safe Havens / Farr Collaborators

Trusted Platforms

Data Safe Haven
Identifiable / weakly pseudonymised Information Assets

- ISO 27001 certification
- Staff Contracts
- Data Sharing Contract
- Information Security Policies
- Staff Training and Education
- Data Export Module
- Logs details of what information is sent to whom
- De-identification Algorithms

Data Sharing Agreements

Secure Data Transfers

Logs details of what information is sent to whom

Other Collaborators / Publications

Secure Data Transfers

Independent, Periodic Audit

Current Governance Tapestry

Data Sharing Agreements

Credit: Dr Nathan Lea, UCL
UCL Data Safe Haven Platform

• Technical solution for the storage, handling, and analysis of identifiable data.
• Everything (data + applications) sit behind a firewall
• Certified to ISO/IEC 27001
• No communication with the outside world except via remote virtual screens.
• Dual factor authentication, 60 second tokens
• Vast list of services (e.g. MySQL, R, Stata, Office, etc.)
• Separate portal for data transfer in (TLS) and analysis

Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research.

Auditing the Platform

- Recently (May 2017) audited by NHS Digital across several areas:
  - Information transfer
  - Access control
  - Data use and benefits
  - Risk management
  - Operational management and control
  - Data destruction
- Achieved moderate and substantial assurance ratings across all areas.
Data Safe Haven
Denaxas Lab

Academic members

Spiros Denaxas
Senior Lecturer in Biomedical Informatics

Ghazaleh Fatemifar
Research Associate in Genetic Epidemiology

Arturo Gonzalez-Izquierdo
Research Associate in Electronic Health Records

Natalie Fitzpatrick
Research Data Coordinator

Kenan Direk
Research Associate in Electronic Health Records

Maria Pikoula
Clinical Data Scientist

Michalis Katsoulis
Medical Statistician

Vaclav Papez
Clinical Data Scientist

Hannah Evans
Biostatistician

Marcos Barreto
Royal Society Newton Fellow

Marina Daskalopoulou
Research Assistant

Colin Josephson
Assistant Professor of Neurology and Community Health Sciences (Univ of Calgary)

Marie Erwood
IMAGINE Data Manager
Data Lab

At the intersection of clinical research and computer science, the Data Lab seeks to develop data-driven computational methods and tools for exploiting electronic health records for biomedical research.

Find out more...

Projects

By providing a complete picture of patients' health care utilisation and deep longitudinal phenotyping, CALIBER has become a leading research platform for the understanding of health progression.

Find out more...

CALIBER Data

Frequently asked questions on who can use the resource, how to apply to use CALIBER data, fees, and resources and capacity building.

Read more here...

Materials

Resources including sample ISAC application forms, agreements and forms are available to guide researchers through the application process.

More information >

Contact and support

Click below for contact details regarding information about the CALIBER resource, fees and support, or to discuss potential collaboration for specific enquiries about existing CALIBER projects.

Read more here...

Publications

Please click here for a full list of publications using CALIBER data.

Read more here...