Data set minute madness

Whitehall II Study

Social determinants of health and interdisciplinary study of ageing

- 10,308 participants, aged 35-55, recruited from the London Civil Service offices in 1985
- 12 waves of data collection: self-completion questionnaires and clinical examinations
- Wide range of data on social inequalities, chronic disease, biomedical information, genetics

Data sharing

www.ucl.ac.uk/whitehallII/data-sharing

- Gated-access approach: application form and list of variables
- Medium-sized projects (1-2 papers), bona fide research
- Release of tailor-made datasets
- Enquiries: Aida Sanchez, a.sanchez@ucl.ac.uk

Data across the lifecourse in a rural South African community

~85,000 Individuals from ~23,000 Households over 16 Years

Population Platform

Family Environment

Sexual Behaviour

General Health

Bio-measures

Pregnancy Intention | Pregnancy Notification | Pregnancy Outcome | Child Health | Educational Attainment | Verbal Autopsy

Pre-conception | Antenatal | Delivery | Childhood | Adolescence | Adulthood ... Death

Clinical Platform

HIV & TB Care Electronic Systems

Hospital Information System

NHLS Laboratory Results

Patient Registration System (Tier, ETR)

Population Data Collection

Laboratory Data

Routine Health Information Systems

Clinical Data

Collaborations with AHRI-affiliated researchers strongly encouraged.
Instructions for data access at: http://www.africacentre.ac.za/index.php/data-rep

British Sign Language Corpus

- Collection of ~125 hours of videos & metadata from 249 deaf BSL signers from 8 regions across UK (2008-2011) in conversation, telling stories, being interviewed and giving their signs for 102 concepts
- Some annotations/translations via ELAN multimedia annotation tool (https://tla.mpi.nl/tools/tla-tools/elan)
- Important resource for sign language researchers, teachers, students, interpreters. Also interest from language technology sector.
- Ethics & access
  - Informed consent for unanonymised, open access video
  - Dual access corpus: one interface for researchers, one for deaf community. Some data restricted to researchers only, upon application.

http://www.bslcorpusproject.org ; Kearsy Cormier, Director

Census flow data

- Aggregate data about **movements** of people **between** (and within) places
  - Frequent events: daily commuting, journeys to school
  - Less frequent events: moving house
- An origin, a destination, and characteristics of the people moving between them
- Available from the 1981, 1991, 2001 and 2011 censuses
  - Data from 1981 to 2001 are all Open Data, free to all
  - Data from 2011 are a mixture of *open* data, *safeguarded* data and *secure* data

census.ukdataservice.ac.uk

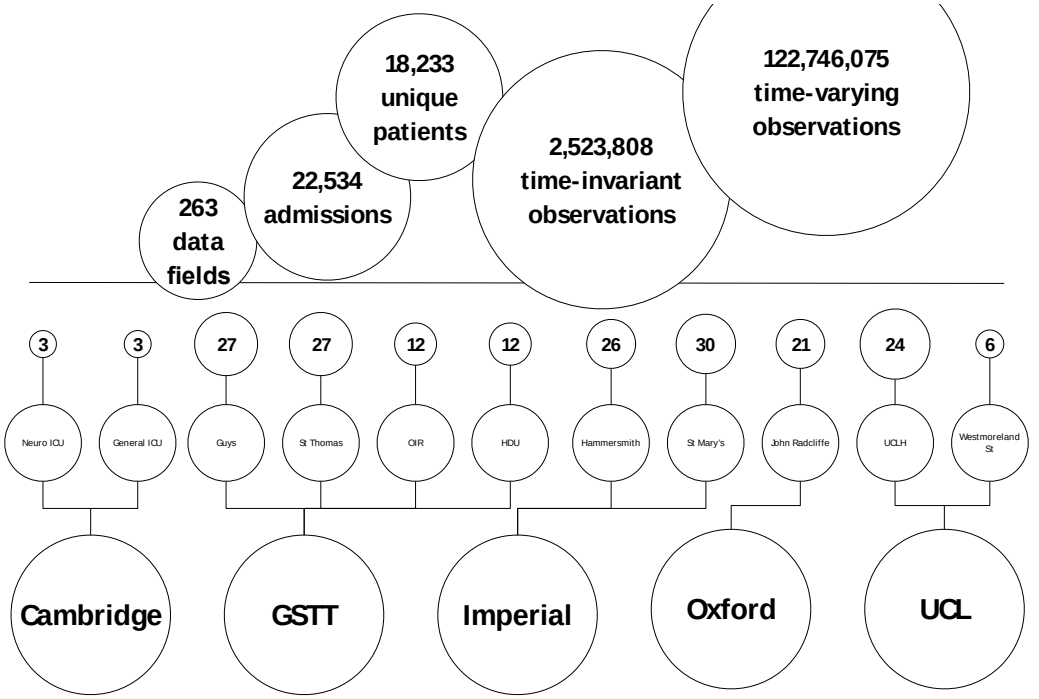
UK Data Service
Census Support

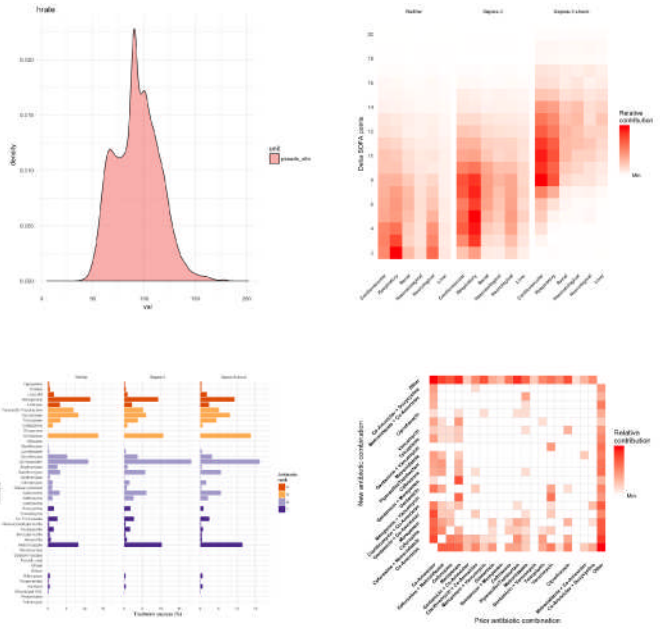
The Office for National Statistics Longitudinal Study (LS) *for studying .....*

- risk factors for inter-generational social mobility – or immobility
- health inequalities at small-area level
- second- or third-generation immigrants
- local labour markets – relationship to internal migration
- jobless or poor young people: what becomes of them?
- life expectancy by occupation, area, number of children
- timing of parenthood and career progression

www.ucl.ac.uk/celsius ***celsius@ucl.ac.uk***

Critical Care Data
Health Informatics Collaborative





Multicentre:
5 Biomedical Research Centres

Research ready:
Open source R package (cleanEHR)
provided for data manipulation

Linkable:
Patient identifiers retained
for linkage to other health and social care data

Rapid development:
Anonymised development data available

HIDDEN HISTORIES

- The **Hidden Histories of computing in the Humanities** project is compiling an **oral history** of the field of **Digital Humanities** (PI Julianne Nyhan, Senior Lecturer, Department of Information Studies)

- Nyhan has collected **c.40 hour**s of oral history recordings (and interviews are **ongoing**)

- The oral history recordings are available on the **project website**: http://hiddenhistories.omeka.net or **contact Nyhan directly** (j.nyhan@ucl.ac.uk)