Linking health and socioeconomic data: experience from UK

Chris Dibben

Director of the Scottish Longitudinal Study
and the Administrative Data Research
Centre - Scotland
University of Edinburgh

THE UNIVERSITY
of EDINBURGH

Overview

- Brief history – from a slightly personal perspective

- Developing institutions to support research uses of linked administrative data

- Governance, technologies and organisational structures being developed – focus on 2 new developments

Section of Epidemiology and Preventive Medicine

President Charles M Fletcher CBE MD

Meeting November 28 1963

Paper

The Oxford Record Linkage Study: A Review of the Method with some Preliminary Results

by E D Acheson DM MRCP and
J G Evans MB MRCP
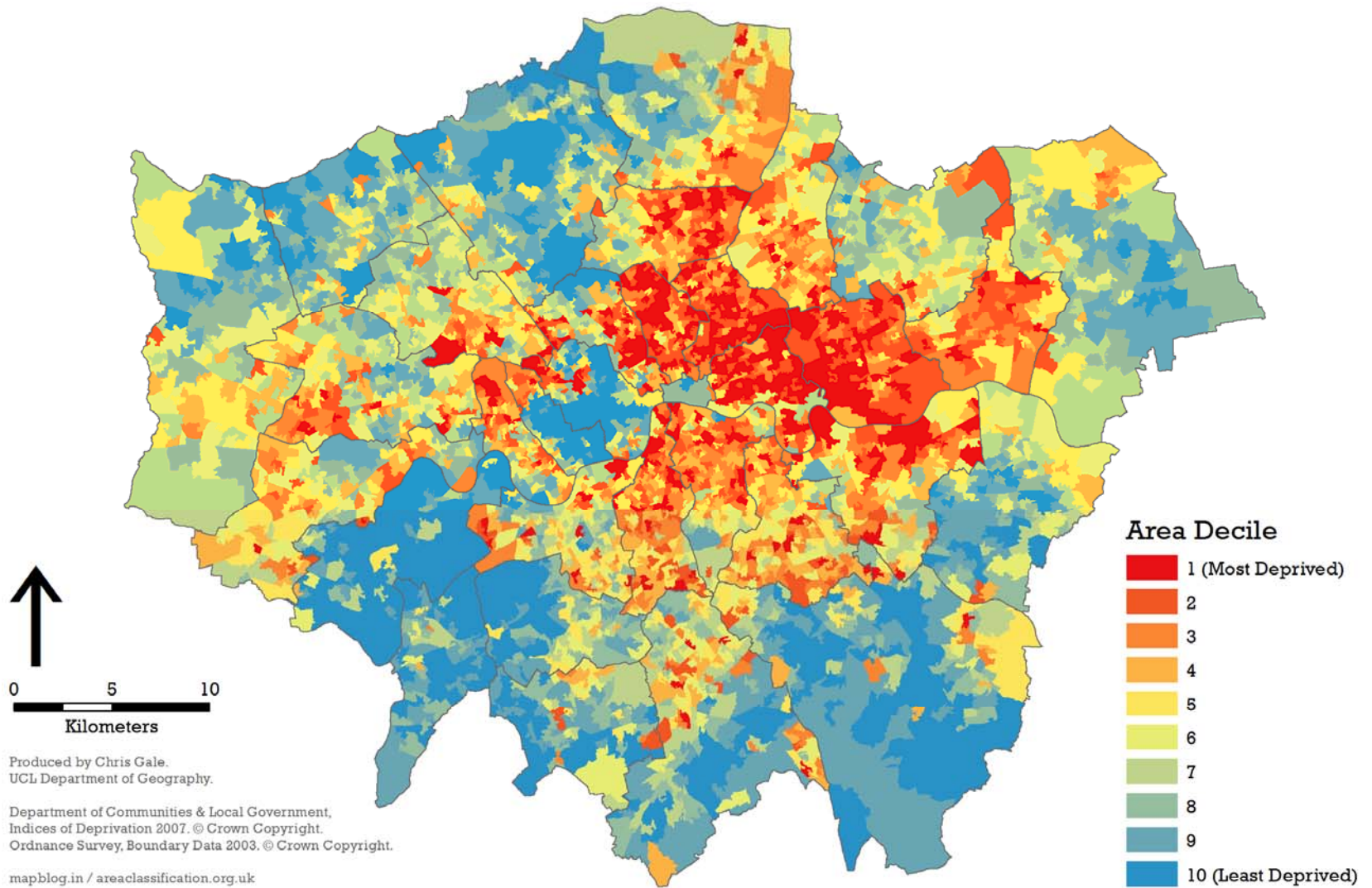(*Nuffield Department of Clinical Medicine,
Oxford University*)

Introduction
We have begun an experiment in Oxford in re-

Current Lack of Organization
In Britain many important data are recorded about health and disease, and we have certain types of health statistics available on a national scale. For example, for many years we have had the most sophisticated analysis of mortality in the world; valuable data concerning discharges from hospital are also provided by the Hospital In-patient Enquiry; an annual digest of certificates of incapacity for work is also available. Much information of epidemiological and administrative

The English Indices of
Deprivation 2010

Individual Domains and the IMD
Red = Most Deprived   Blue = Least Deprived

Income Deprivation
(22.5% of IMD)

Health Deprivation and Disability
(13.5% of IMD)

Employment Deprivation
(22.5% of IMD)

Education, Skills and Training
(13.5% of IMD)

Index of Multiple Deprivation
Source: Communities and Local Government, 2011

Barriers to Housing and Services
(9.3% of IMD)

Crime
(9.3% of IMD)

Living Environment Deprivation
(9.3% of IMD)

Health & Disability Index of Deprivation for London in 2007

Area Decile

- 1 (Most Deprived)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Least Deprived)

0   5   10
Kilometers

Produced by Chris Gale.
UCL Department of Geography.

Department of Communities & Local Government,
Indices of Deprivation 2007. © Crown Copyright.
Ordnance Survey, Boundary Data 2003. © Crown Copyright.

mapblog.in / areaclassification.org.uk

BBC NEWS

▶ LIVE   **BBC NEWS CHANNEL**

News services
Your news when you want it

✉ E-mail this to a friend   🖨 Printable version

Brown apologises for records loss

Prime Minister Gordon Brown has said he "profoundly regrets" the loss of 25 million child benefit records.

He apologised in the Commons for the "inconvenience and worries" caused and said the government was working to prevent the data being used for fraud.

But Conservative leader David Cameron said the government had "failed in its first duty to protect the public"

Mr Brown said the government was not "complacent" about the losses

▶ Watch   **Brown apologises**

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science & Environment
Technology
Entertainment
Also in the news

Video and Audio

VIDEO AND AUDIO NEWS
Police searching the offices where the records were lost
▶ Watch

BENEFIT DISCS LOST

KEY STORIES
▸ Discs loss 'entirely avoidable'
▸ MoD 'Facebook generation' fear
▸ New boss for HMRC after data slip
▸ Tougher data rules for Whitehall
▸ More firms 'admit disc failings'
▸ £20,000 reward offered for discs
▸ Fresh benefit data lapse admitted
▸ Discs 'worth £1.5bn' to criminals

Moving at a fast pace…

- **Dec 2012**: ADT Report & Chancellors Autumn Announcement
- **June 2013:** Government Response
- **October 2013**: Commissioned ADRCs and ADS
- **October 2013 – now:** Meetings to develop Network
- **April 2014:** First ADRN Board meeting
- **May 2014**: Public Dialogues Report & Website Launch & Beta Testing Phase
- **August 2014**: First Approvals Panel meeting
- **October 2014**: Gateway Review- Readiness for Service
- **November 2014:** Opening its doors to external researchers

Administrative Data Research Network

farr

ADRC-Scotland

▶ Co-located with Farr Institute, Scottish Government and NHS.

▶ Universities of Aberdeen, Dundee, Edinburgh, Glasgow, Herriot-Watt, St Andrews and Stirling.

▶ Expertise in admin. data and Public Engagement, linkage, Law and relevant computer science techniques.

Research the major Scottish administrative datasets

- Schools, colleges and universities
- The criminal and justice system
- Social work services
- Social welfare
- Housing system
- Transport system
- Health system
- Historical administrative data

Administrative Data Research Network

An ESRC Data
Investment

Administrative data   **Using the Network**   **Catalogue**   **Protecting privacy**   **News & events**   **About us**   **FAQs**

Home / Protecting privacy

▶ Approving projects

▶ De-identified data

▶ Trusted researchers

▶ Secure environment

▶ Safe results

▶ Legal framework

Protecting **privacy**

Share

We aim to have the highest standards of secure data sharing, which will be consistent across the Network. This means:

- safe projects - only projects approved by our Approvals Panel will have access to our services.
- safe people - only accredited researchers will have access to the Network's services.
- safe, de-identified data: researchers will not be able to see information which directly identify any individual.
- secure environments: state-of-the-art secure information technology and procedures will provide physical, hardware and software security across the whole Network.

Privacy-ethical advisory bodies

Data holder 1   Data holder 2   Data holder 3

Research centre

Indexing

Indexer

Trusted third party

Linking Database creation

Safe setting

→ Limited information flow - identifier only

--→ Limited information flow - anonymous data only

Remote Safe setting

Researcher   Researcher

Public Engagement

adls@st-andrews.ac.uk

Only then does Safepod unlock its door to the researcher.

When an unauthorised fingerprint is detected, access is denied.

SafePODs

Synthetic data

Real (input)

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR			
FEMALE	54	SECONDARY			
MALE	30	VOCATIONAL/GRAMMAR			
MALE	68	SECONDARY			
MALE	61	PRIMARY/NO EDUCATION			

Data that look (structurally) like original data but contain artificial units only

Synthetic (output)

Sex	Age	Education	Marital status	Income	Life satisfaction
MALE	81	PRIMARY/NO EDUCATION	MARRIED	2100	PLEASED
MALE	54	VOCATIONAL/GRAMMAR	MARRIED	1700	PLEASED
FEMALE	32	VOCATIONAL/GRAMMAR	DIVORCED	870	MIXED
FEMALE	98	PRIMARY/NO EDUCATION	MARRIED	800	MOSTLY DISSATISFIED
FEMALE	50	PRIMARY/NO EDUCATION	MARRIED	NA	MOSTLY SATISFIED
FEMALE	37	VOCATIONAL/GRAMMAR	MARRIED	158	PLEASED
MALE	28	VOCATIONAL/GRAMMAR	NA	1500	MOSTLY SATISFIED
FEMALE	62	PRIMARY/NO EDUCATION	MARRIED	830	MOSTLY SATISFIED
MALE	78	PRIMARY/NO EDUCATION	MARRIED	NA	PLEASED
FEMALE	29	SECONDARY	MARRIED	580	MOSTLY SATISFIED
MALE	59	PRIMARY/NO EDUCATION	MARRIED	1300	MOSTLY SATISFIED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
MALE	18	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	73	PRIMARY/NO EDUCATION	WIDOWED	1350	MOSTLY SATISFIED

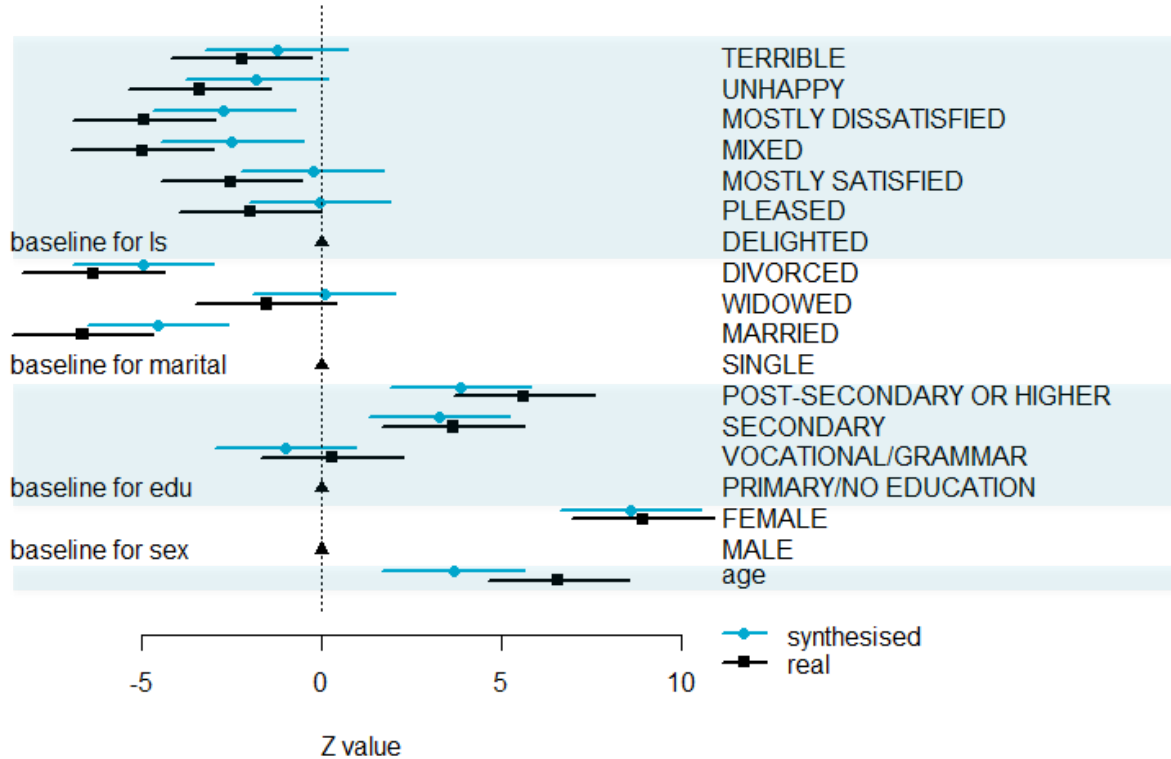# Generating synthetic data

Sequentially replacing original data values with synthetic values generated from conditional probability distributions

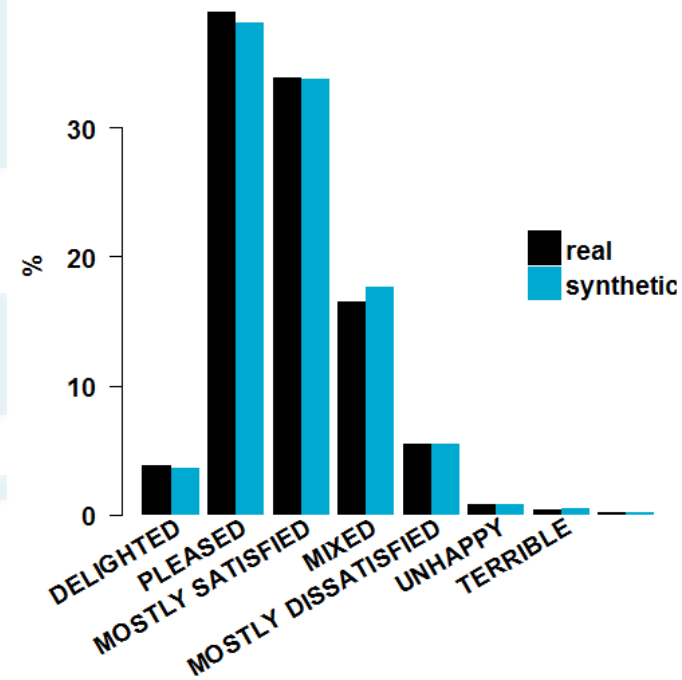$$Y_j \sim (Y_0, Y_1, \ldots, Y_{j-1})$$
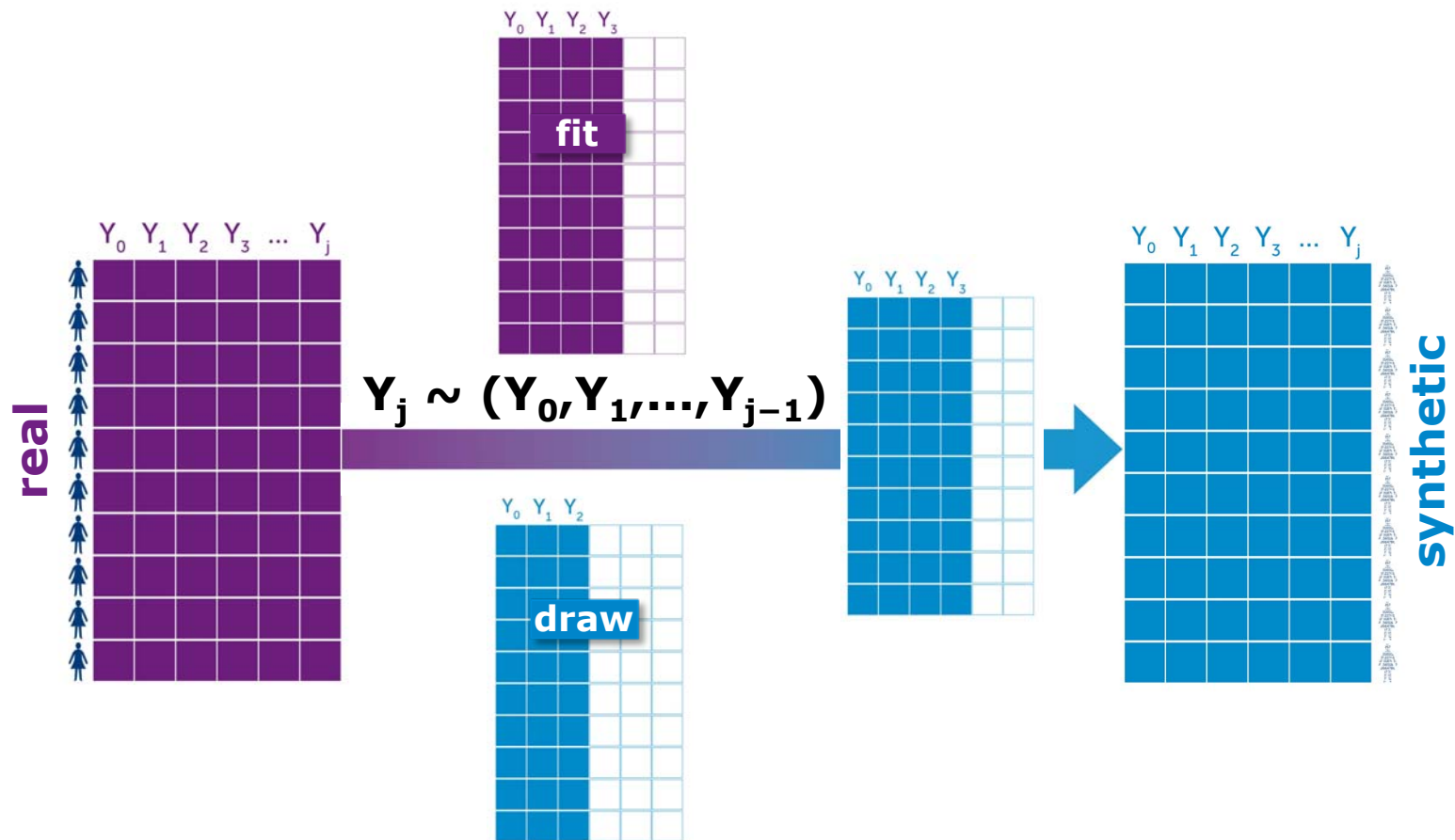
Data that behave (statistically) like original data
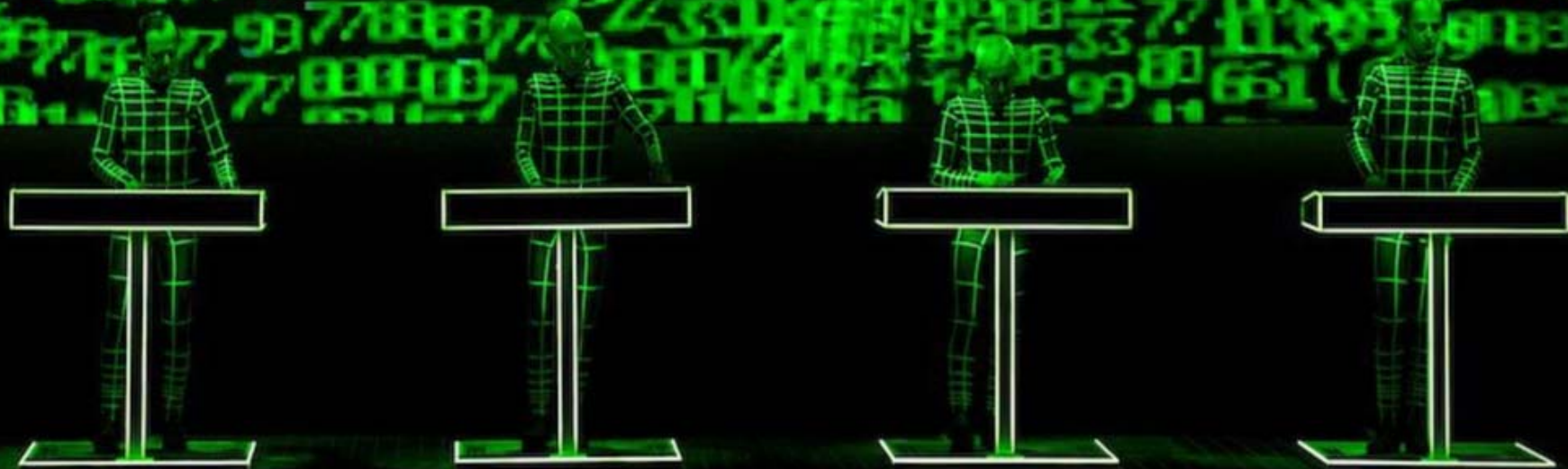
Final thoughts

- Established of a common approach
- Greater amounts of data sharing
- But risks to present system
  - Need for public engagement
  - No specific data sharing legislation
  - Changes to data protection legislation