HDRUK
Health Data Research UK

Improving our health
through data science

Potential of big data research for geriatric medicine

Professor Ronan Lyons, Swansea University

11th March 2019

r.a.lyons@Swansea.ac.uk
www.hdruk.ac.uk

Health Data Research UK is uniting the UK's health data to make discoveries that improve people's lives

Health Data Research UK

Our vision is that every health and care interaction and research endeavour will be enhanced by access to large scale data and advanced analytics

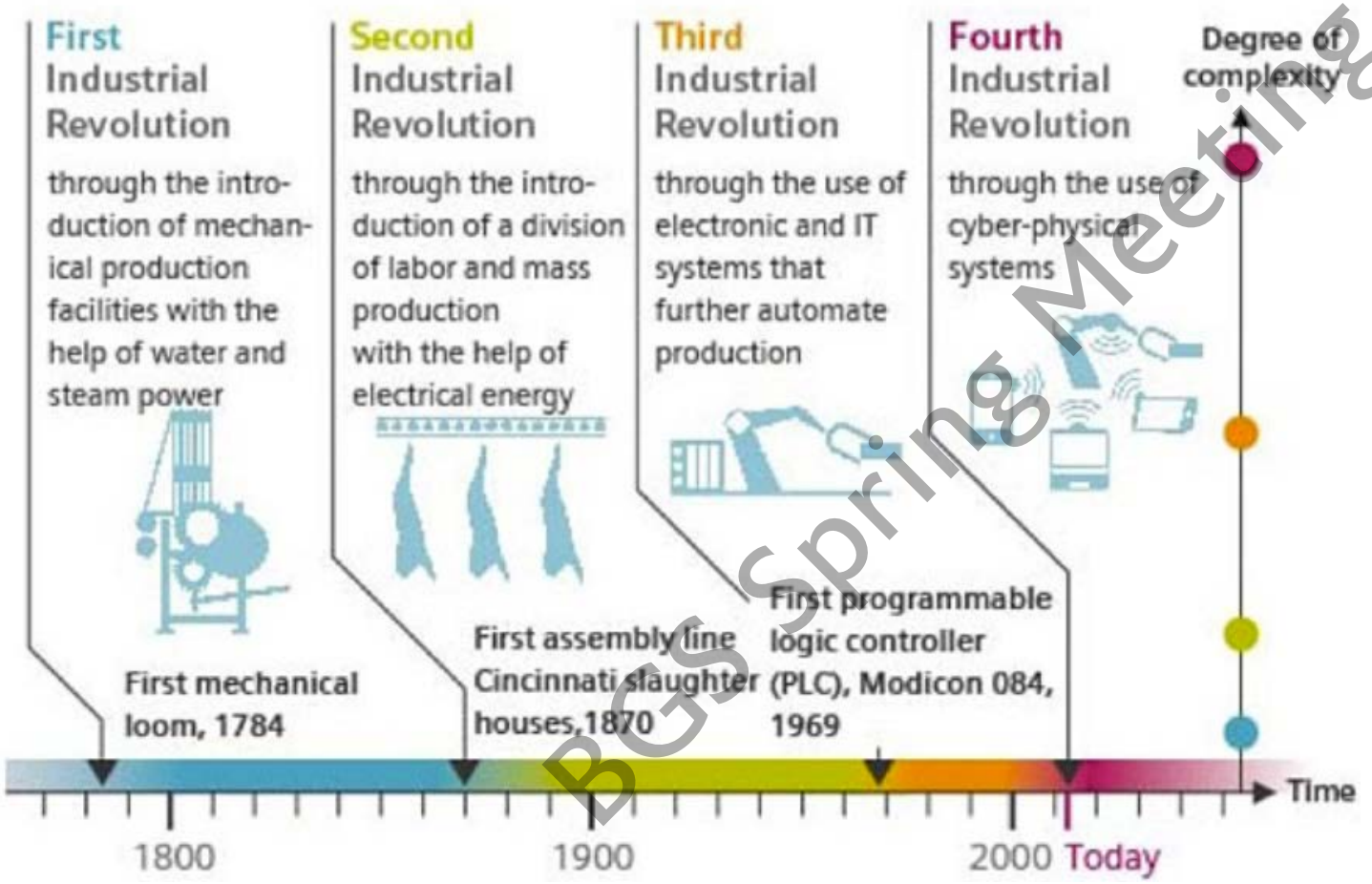
We are uniting health data assets across the UK to make health data research and innovation happen at scale

We work with people who share a common set of values:

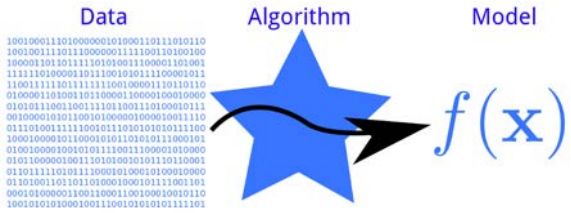
Transparency
Optimism
Respect
Courage
Humility

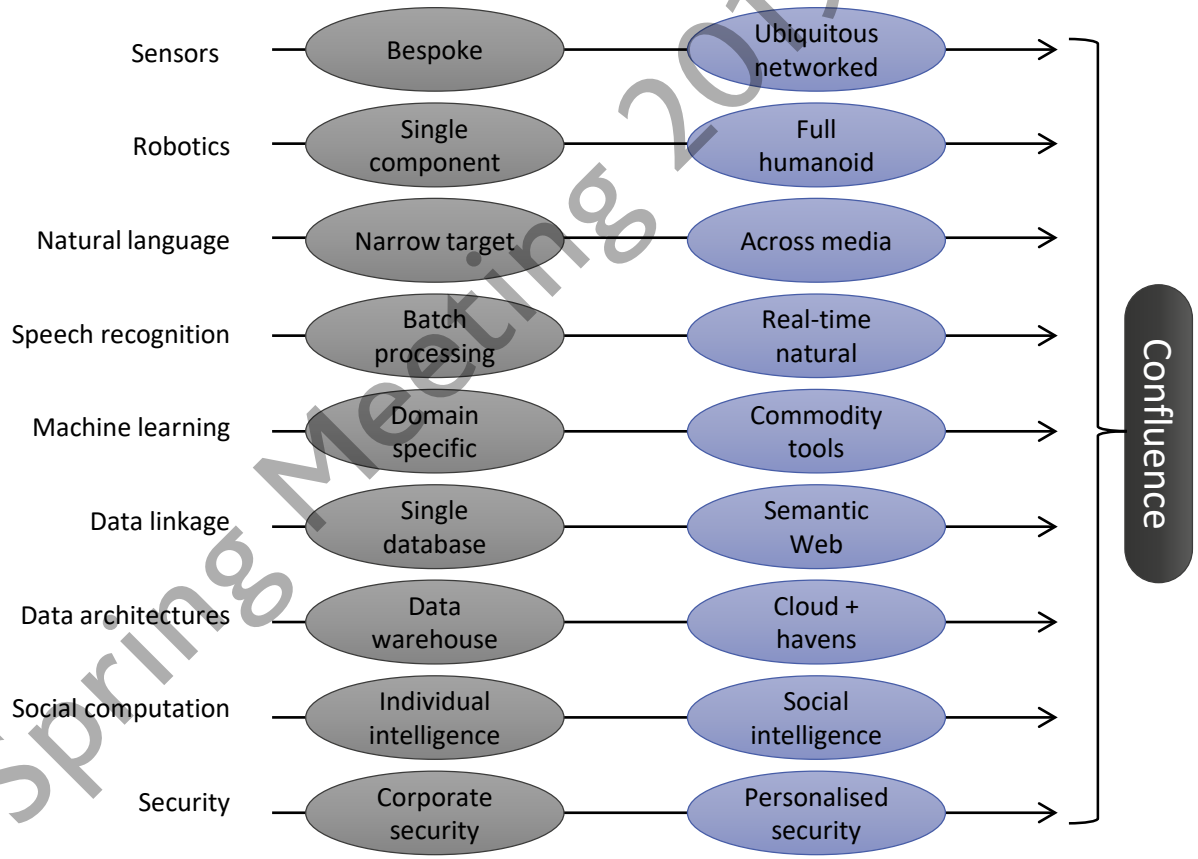
We will deliver our mission through: Great science, Great people, Great infrastructure

The 4th Industrial Revolution



Technologies





Why Now? "4P" Medicine

Predictive	***Customise diagnosis and treatment***
Pre-emptive	***Better than curative – earlier diagnosis***
Personalised	***Determine risk profiles, predict outcomes***
Participatory	***Involve patients***

Made Possible by

- **Genomics**
- **Phenotyping**
- **Informatics**
- **Analytics**
- **New social contract**

HDRUK
Health Data Research UK


The Translational Scientist

Disease Area | Research Field | Tools & Techniques | Outside the Lab | Issues | Product Profiles

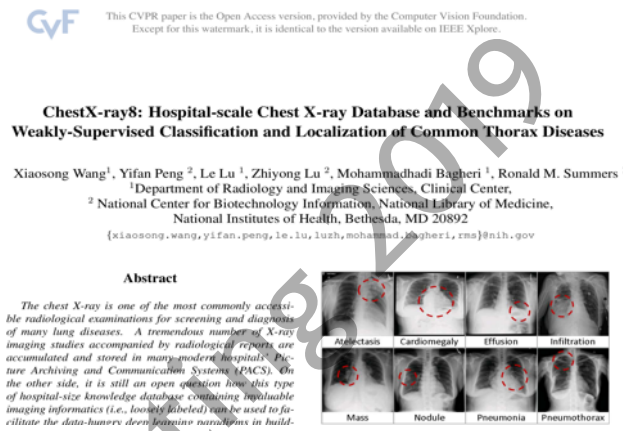
TOOLS & TECHNIQUES | Analytical science | Animal models | Clinical trials | Data mining | Diagnostics & prog

TOOLS & TECHNIQUES | Informatics, Omics

Three Gurus of Big Data

Big data. Everyone's talking about it, but what exactly is it? How can it be harnessed to advance translational science? And what perils lie within the oceans of data that now surround us? Three experts from different backgrounds go fishing for answers.

Dipak Kalra, Iain Buchan, and Norman Paton | 11/23/2016



ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases

Xiaosong Wang[1], Yifan Peng [2], Le Lu [1], Zhiyong Lu [2], Mohammadhadi Bagheri [1], Ronald M. Summers [1]
[1]Department of Radiology and Imaging Sciences, Clinical Center,
[2] National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health, Bethesda, MD 20892
{xiaosong.wang,yifan.peng,le.lu,luzh,mohammad.bagheri,rms}@nih.gov

Abstract

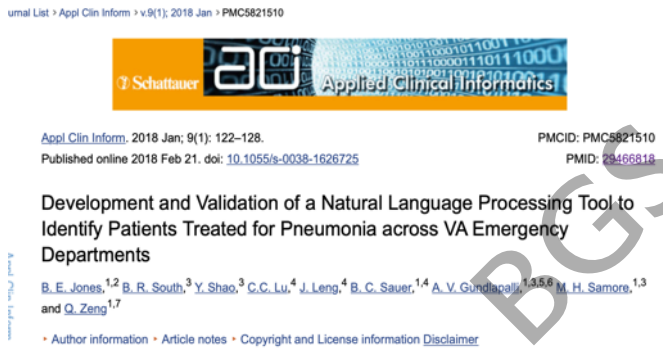
The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. A tremendous number of X-ray imaging studies accompanied by radiological reports are accumulated and stored in many modern hospitals' Picture Archiving and Communication Systems (PACS). On the other side, it is still an open question how this type of hospital-size knowledge database containing invaluable imaging informatics (i.e., loosely labeled) can be used to facilitate the data-hungry deep learning paradigms in build-

JAMA Network | Open

Invited Commentary | Critical Care Medicine

Can Big Data Deliver on Its Promises?—Leaps but Not Bounds

Ithan D. Peltan, MD, MSc; Sarah J. Beesley, MD, MSc; Samuel M. Brown, MD, MS

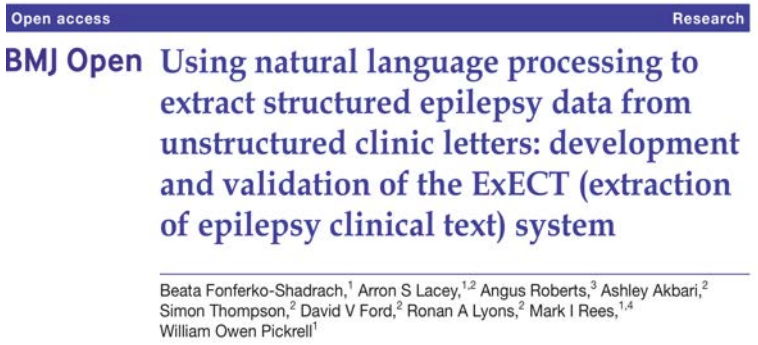

urnal List > Appl Clin Inform > v.9(1); 2018 Jan > PMC5821510

Schattauer | Applied Clinical Informatics

Appl Clin Inform. 2018 Jan; 9(1): 122–128.
Published online 2018 Feb 21. doi: 10.1055/s-0038-1626725
PMCID: PMC5821510
PMID: 29466818

Development and Validation of a Natural Language Processing Tool to Identify Patients Treated for Pneumonia across VA Emergency Departments

B. E. Jones,[1,2] B. R. South,[3] Y. Shao,[3] C.C. Lu,[4] J. Leng,[4] B. C. Sauer,[1,4] A. V. Gundlapalli,[1,3,5,6] M. H. Samore,[1,3] and Q. Zeng[1,7]


Open access Research

BMJ Open **Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system**

Beata Fonferko-Shadrach,[1] Arron S Lacey,[1,2] Angus Roberts,[3] Ashley Akbari,[2] Simon Thompson,[2] David V Ford,[2] Ronan A Lyons,[2] Mark I Rees,[1,4] William Owen Pickrell[1]

Distributed Team Science

HDRUK
Health Data Research UK

Essential, but often missing ingredient

Big data requires Big minds

Making sense of the vast amounts of complex multi-modal data requires large numbers of bright people, from many disciplines, working collaboratively

Breadth of perspectives needed are never found in one place

Collaboration can be difficult: system rewards ego-centric science

Solution is to support distributed team science

More big brains applied to the issues – join the club!

The biggest challenge of all: trust, trust, trust

HDRUK's niches in public health science

1: Develop privacy protecting, population based e-cohorts for the UK

- largely derived from real world routinely collected datasets
- To answer complex population health questions that involve 'a complex array of interlinking factors' in a timely manner requires linking data on exposures, interventions and outcomes longitudinally, and at scale, and:
- A reusable, population-based, data infrastructure which can support observational, interventional and post intervention real world impact evaluation
- Aspiration to cover UK **66M** population in 5 years

2: Embellish traditional cohorts with routine/other data and support access through multi-cohort platforms

- The UK's tremendous range of traditional cohorts are increasingly being brought together to answer public health research questions
  - MRC Dementias Research Platform
  - Mental Health Data Pathfinder initiative
  - MQ funded Adolescent Data Platform
  - CLOSER cohort initiative
- HDRUK should have a role in ensuring efficient routine data follow up, particularly through NHS Digital

3. Develop and apply innovative analytical methodologies that are fit for purpose for 21st century population health research

4. Develop the skilled workforce capable of utilising the above developments to answer bold and ambitious research questions

UK Secure eResearch (UKSeRP) technology platform

HDRUK
Health Data Research UK

Llywodraeth Cymru
Welsh Government

CHILDREN OF THE 90S

Dementias Platform UK
Medical Research Council

MS* REGISTER

CRIS NETWORK

MRC Cloud Infrastructure for Microbial Bioinformatics

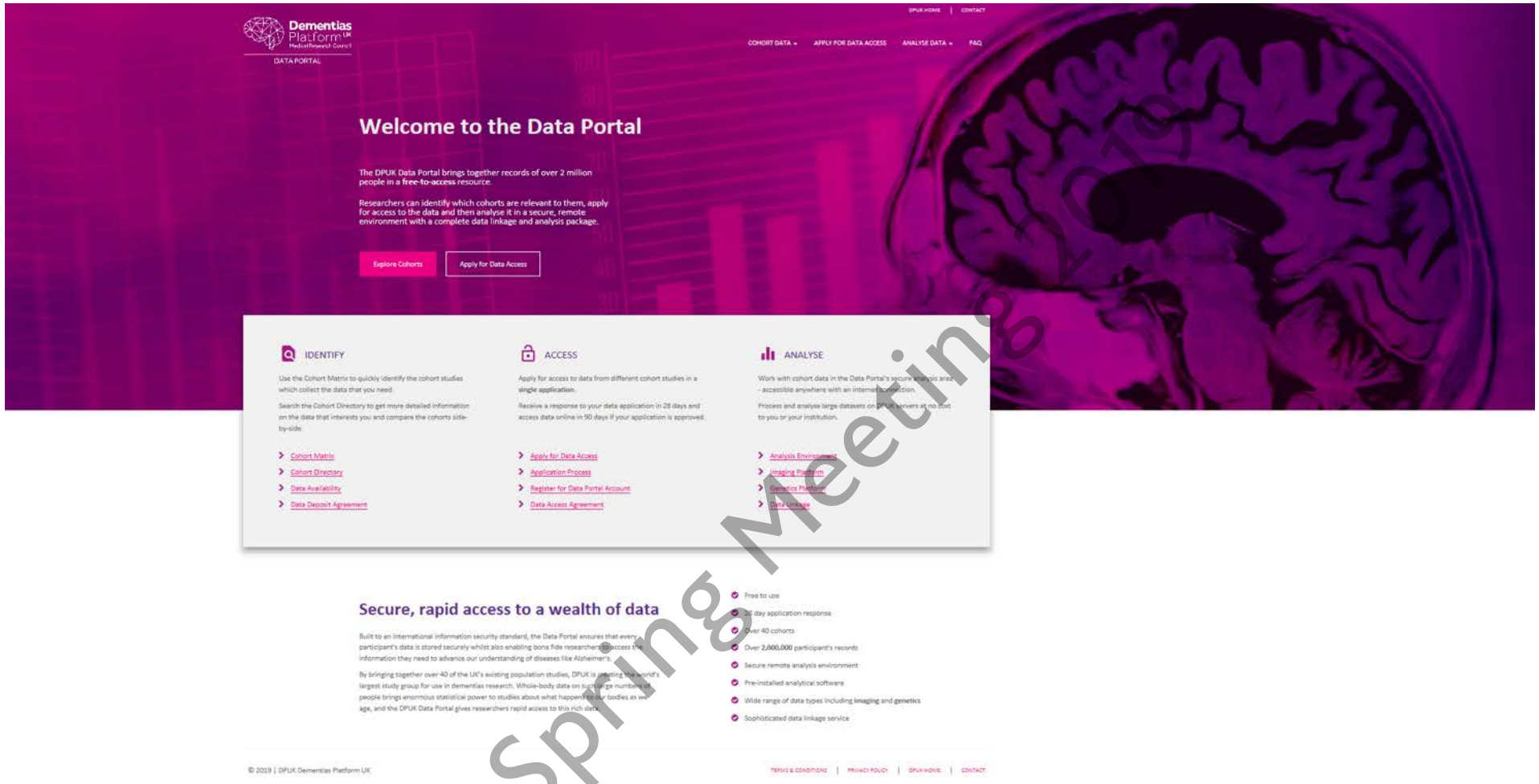
SAIL DATABANK

Administrative Data Research Network

East London Genes & Health

biobank uk
Imaging study

Remote Access	Security	Data Management	Data Storage	Administrative Control	File Storage
Data Analysis Tools	BioInformatics	MedGate (NLP)	Imaging	Genomics	Technical Support



Data Portal enhancement:

- Data Portal increase to 60 cohorts (n≈4m)
- Focus: mechanistically informative datasets
- Negotiating inward investment (in addition to MRC) of >£2m
- Reference paper submission to BioRXve, IJE
- Freely available

MUrMuR-UK - Measuring & Understanding Multimorbidity using Routine Data in the UK

Colin McCowan Bruce Guthrie
Jill Pell Corri Black
Steve Pavis Marion Bennie
Frank Sullivan Chris Dibben
Albert Tenesa Carole Morris

Anthony Brookes Krish Nirantharakumar
Keith Abrams Sophie Staniszewska
Phil Quinlan Umesh Kadam
Kamlesh Khunti

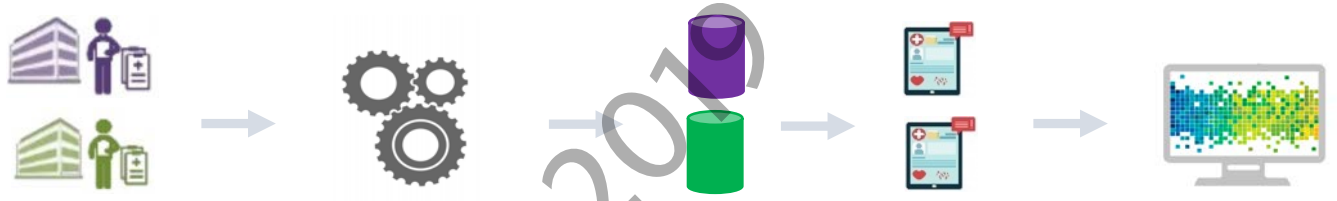
Ronan Lyons Dermot O'Reilly
David Ford, Alan Watkins,
Ashley Akbari

Silvia Richardson Emanuele Di Angelantonio

Jim Davies

Danny Alexander Ruth Gilbert
Spiros Denaxas John Robson
Carol Dezateux Waty Lilaonitkul
Aroon Hingorani Paul Elliot

Data Provider	Type	Base Population
Scotland: eDRIS	Hospital Prescribing Disease registries Population spine	5.4m
Northern Ireland	Hospital Prescribing Disease registries Population spine	1.2m
Wales: SAIL	Hospital Prescribing Disease registries Population spine GP	3.5m
CALIBER	Hospital Disease registries GP	10m
Leicester City, Leicestershire & Rutland	GP	1.1m
Discovery (London)	GP Hopital	1.2m



Data Extraction → Harmonisation & Standardisation → Storage in CDM → Morbidity Coding → Analysis

- Data extraction, harmonisation & standardisation
  - All people aged 25+ on 01/01/2000 followed to death or end of 2018
  - Consolidation of different ontologies
  - Extraction, Transform & Load to a Common Data Model
- Conceptualisation, coding & validation
  - Systematic reviews and international consensus to identify purposes of MM measures and relevant conditions
  - Methods for identifying morbidities feeding into phenotype library
  - Validation of identified code sets from existing or primary work
- Analysis
  - Descriptive analysis of prevalence of MM and differences by age, gender, SES, ethnicity, region etc.
  - ML and statistical approaches to identifying clusters of disease that cause the most burden and trajectories of disease
  - Distributed team science approach to analysis work
- PPI and Governance

SAIL DATABANK

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

‘Safe Haven’ for pseudonymised data (Welsh population)

World leading, privacy protecting research environment

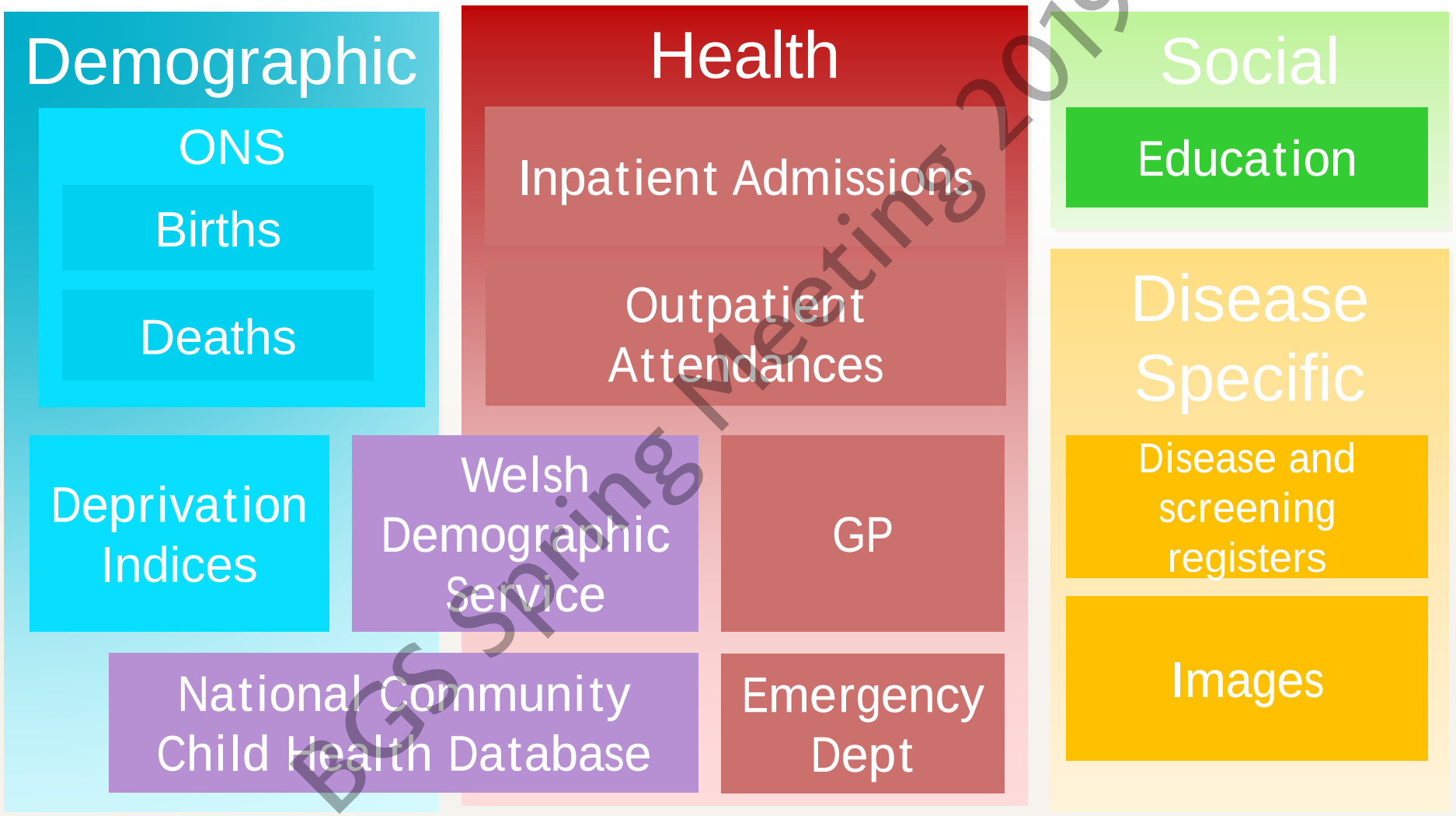
Remote access

Linked, longitudinal de-identified data on Welsh population
5 million people – 26+ billion rows of data
Over 20 core national datasets
200+ project specific datasets
Non health service, social and environmental data

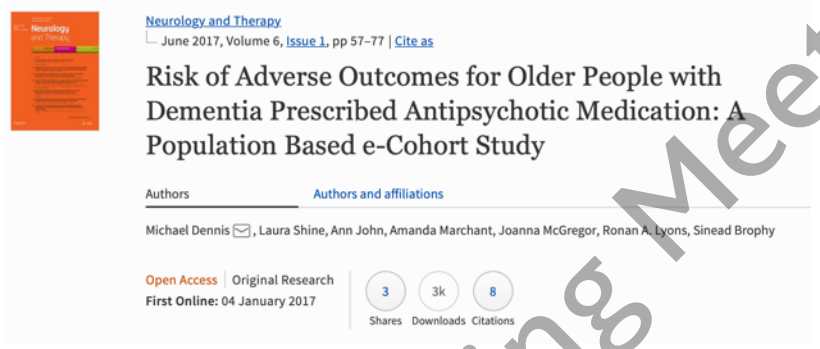
www.saildatabank.com

Some members of the Consumer Panel

SAIL Datasets



Clinical and population e-Cohorts



Neurology and Therapy
June 2017, Volume 6, Issue 1, pp 57–77 | Cite as

Risk of Adverse Outcomes for Older People with Dementia Prescribed Antipsychotic Medication: A Population Based e-Cohort Study

Authors | Authors and affiliations

Michael Dennis, Laura Shine, Ann John, Amanda Marchant, Joanna McGregor, Ronan A. Lyons, Sinead Brophy

Open Access | Original Research
First Online: 04 January 2017

3 Shares | 3k Downloads | 8 Citations

Open access Research

BMJ Open

Acute kidney injury in the UK: a replication cohort study of the variation across three regional populations

Simon Sawhney,[1,2] Heather A Robinson,[2,3] Sabine N van der Veer,[2,3] Hilda O Hounkpatin,[2,4] Timothy M Scale,[2,5] James A Chess,[2,5] Niels Peek,[2,3] Angharad Marks,[1,2] Gareth Ivor Davies,[2,5] Paolo Fraccaro,[2,3] Matthew J Johnson,[2,4] Ronan A Lyons,[2,5] Dorothea Nitsch,[2,6] Paul J Roderick,[2,7] Nynke Halbesma,[2,8] Eve Miller-Hodges,[2,9] Corrinda Black,[1,2] Simon Fraser[2,7]

Epilepsia
Official Journal of the International League Against Epilepsy

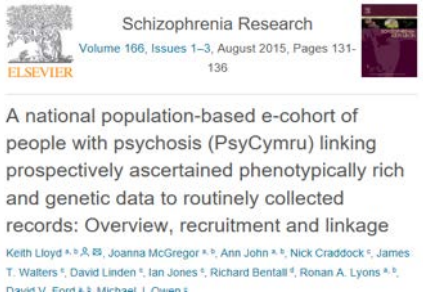
Full-Length Original Research | Free Access

Epilepsy and deprivation, a data linkage study

William O. Pickrell, Arron S. Lacey, Owen G. Bodger, Joanne C. Demmler, Rhys H. Thomas, Ronan A. Lyons, Phil E. M. Smith, Mark I. Rees, Mike P. Kerr

First published: 02 March 2015 | https://doi.org/10.1111/epi.12942 | Cited by

iGetIt@Swansea University



Schizophrenia Research
Volume 166, Issues 1–3, August 2015, Pages 131-136

A national population-based e-cohort of people with psychosis (PsyCymru) linking prospectively ascertained phenotypically rich and genetic data to routinely collected records: Overview, recruitment and linkage

Keith Lloyd, Joanna McGregor, Ann John, Nick Craddock, James T. Walters, David Linden, Ian Jones, Richard Bentall, Ronan A. Lyons, David V. Ford, Michael J. Owen

THE LANCET
Public Health
Volume 3, Issue 6, June 2018, Pages e279-e288
open access

Articles

Risk of emergency hospital admission in children associated with mental disorders and alcohol misuse in the household: an electronic birth cohort study

Prof Shantini Paranjothy PhD, Annette Evans MSc, Amrita Bandyopadhyay BTech, Prof David Fone MD, Behnaz Schofield PhD, Prof Ann John MD, Prof Mark A Bellis DSc, Prof Ronan A Lyons MD, Daniel Farewell PhD, Sara Jayne Long PhD

CHALICE CHange in ALcohol outlet density and alCohol-related harm to population hEalth

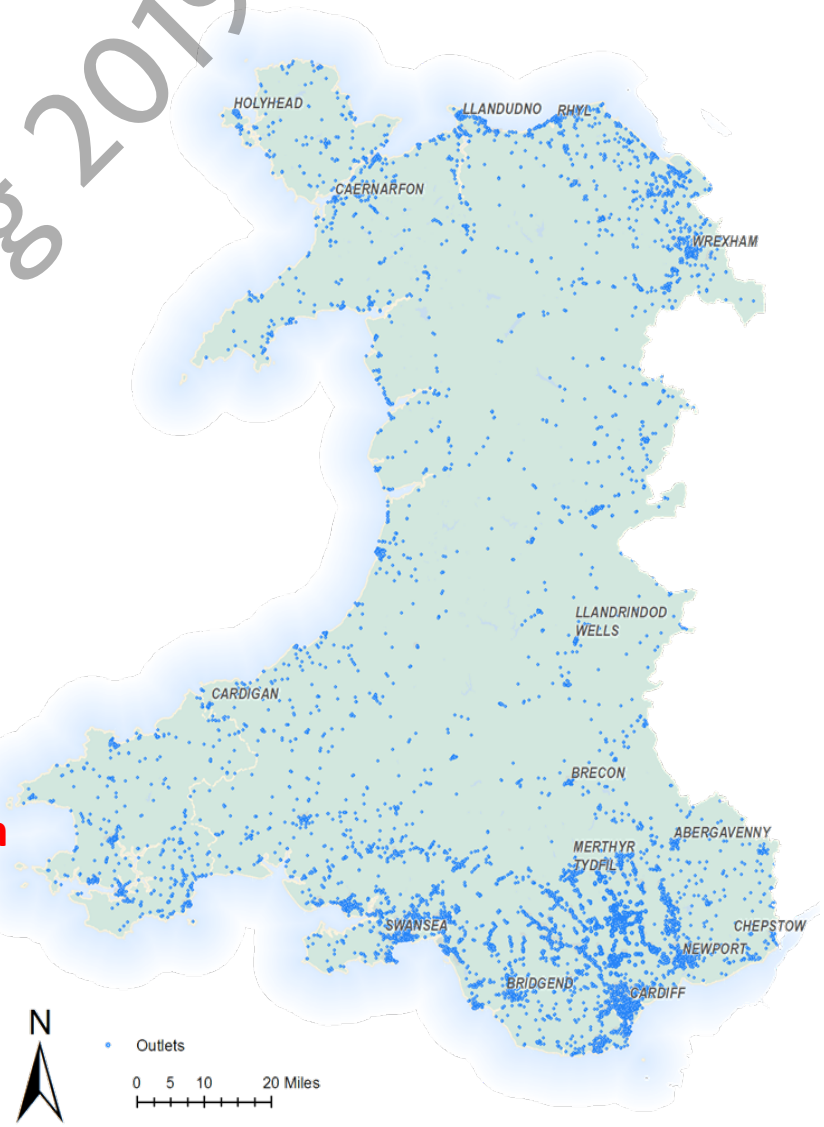
NIHR | National Institute for Health Research

GIS exemplar

- Evaluation of natural experiment over 6 years
- Impact of changing exposure on outcomes
- 2.5M people: 1.4M homes: 16K outlets
- GIS density calculations:
  - Modelled access by outlet type
    - 10 minute walk or drive time
  - Spatio-temporal interaction gravity model
    - 13 billion calculations

Changes in walking distance related to admissions from alcohol related conditions, injury and violent crime

Fone et al. Public Health Res 2016;4(3).



Many existing and developing opportunities

Great presentations to follow

Questions?

r.a.lyons@Swansea.ac.uk