The Farr Institute

Options and Opportunities for UK Health Data Science

29th November 2016

Andrew Morris

Director Farr Scotland
The Next 15 minutes

• Gearing an entire country for quality health care and research
• Data Science as the catalyst for change
• The challenge is the phenotype not the genotype!
• With big data goes big responsibilities
Roles of informatics in translational medicine

1st gap in translation

Basic research → Prototype discovery → Preclin. Devt. → Early trials → Late trials → HTA, HSR → Use in NHS

2nd gap in translation

Bioinformatics
Medicinal informatics
Text mining
Automated experimentation

Trial design & simulation tools
Trial recruitment & data man. tools
Guideline authoring tools

Analysis of linked anonymised clinical datasets

HTA: health technology assessment
HSR: health services research

Adapted from Cooksey report Chart 7.1, page 105
The Future? “4P” Medicine

• Predictive  
  *Customise diagnosis and treatment*

• Pre-emptive  
  *Better than curative*

• Personalised  
  *Determine risk profiles, predict outcomes*

• Participatory  
  *Involve patients*

Made Possible by:  
• Genomics  
• Phenotyping  
• Informatics  
• Analytics  
• New social contract
Our Thesis
Quality Health Care and Research: From Cell to Community

World Class Patient care

Translation Trials and Innovation

Excellence In Life Sciences

Data Science

Personalised Medicine

Better Quality at Reduced Cost
Layered access

Links to CHI / NHS records

Prescription records

Informatics to support patient care

£12.5B

Population 5M

Single health care provider

14 Territorial Boards

38 Hospitals, 1020 General Practices

High rates of morbidity of common complex disease

Collaboration – Aberdeen, Edinburgh, Dundee, Glasgow, St Andrews

Unique patient identifier
### The Scottish Health Service on a Slide

#### Key Trends
- **Population:** 5.3 million
- **% aged 75+:** 7.9%
- **GDP Per Head in 2011:** $42,124

#### Inputs
- **Acute Beds in 2011/12:** 16,500 (NHS)
- **Doctors in 2012:** 12,000 (NHS WTE)
- **Nurses / Midwives in 2012:** 56,600 (NHS WTE)

#### 2012/13

<table>
<thead>
<tr>
<th>Metric</th>
<th>2012/13</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated GP Patient Contacts</td>
<td>16,539,000</td>
<td>3.3%</td>
</tr>
<tr>
<td>Estimated Practice Nurse Patient Contacts</td>
<td>7,627,000</td>
<td>10.5%</td>
</tr>
<tr>
<td>New A&amp;E Attendances</td>
<td>1,561,529</td>
<td>6.8%</td>
</tr>
<tr>
<td>Total Outpatient Attendances</td>
<td>4,699,868</td>
<td>4.7%</td>
</tr>
<tr>
<td>Total Inpatient/Day Case Discharges</td>
<td>1,582,305</td>
<td>6.8%</td>
</tr>
<tr>
<td>Day Case Discharges</td>
<td>448,782</td>
<td>10.6%</td>
</tr>
<tr>
<td>Routine Inpatient Discharges</td>
<td>441,024</td>
<td>9.5%</td>
</tr>
<tr>
<td>Non-Routine (emergency) Inpatient Discharges</td>
<td>540,890</td>
<td>6.4%</td>
</tr>
</tbody>
</table>

### Urgent need to migrate from measurement of activity to REAL TIME MEASUREMENT of processes and outcomes meaningful for patients
Linking Data

the key to seamless care
National level data resources for 5M citizens

BIRTH
- Neonatal Record
- Maternity
- Child health surveillance
- Prescribing
- Dental

GP consultations
Diabetes
- Out patients
- Immunisation
- A&E

Substance misuse
Mental Health
- Hospital Admissions
- Suicide

Community care
Cancer registrations
- Laboratory

DEATH
- BIRTH
- Education
- Looked after children
- Marriage
- Community care
- Care homes

HMRC
DWP
Census (Scotland & UK)
Opportunity to Scale across the nation
The Farr Institute

Part of £200M investment
Our Six Key Activities

1. Cutting Edge Research
2. Harmonised eInfrastructure, methods, data curation
4. Governance (safe havens)
5. Capacity Building
6. Partnerships

To deliver impact nationally and internationally
Building the Infrastructure

“A Research Hotel”
Compute and Analytical Infrastructure
EPCC – national service provider

- Physical sciences have dominated HPC provision for 20 years
- Limited use by biosciences and medicine
- Technology is bringing HPC and Data Analytics together
- New users from medicine and genomics dominant
- 364 projects

EPCC is the UK’s national HPC provider
ARCHER and RDF - £96m UK Gov investment
3,500 users, 118,080 cores, 28Pb of
Farr Institute
Part of Federated Network in Dundee, Glasgow, Aberdeen
The Cohorts

Mature Population Cohorts

Familial Disease Cohorts

Prodromal Population Cohorts

Whitehall
Caerphilly
Lothian BC
Million Women
EPIC
ELSA
CFAS I & II
Aberdeen BC

NSHD
UK Biobank
Chariot
GenScot
Prevent
Interval
SABRE
CAMCAN

FAD
PICNIC
NIMROD
TRACK-HD
OPDC
PD : PD2
FTLD

CamPaIGN
Capacity Building

“Mentorship and Career Development for the next generation of leaders in the field of Data Science”

• Education and Training
  – Postgraduate level courses: MSc and PhD

• Doctoral Training Programme
  – Annual PhD Symposium and Summer School

• Researchers Exchange programme

• Future Leaders in Health Data Science
Innovative Governance

- SAIL and SHIP governance frameworks were endorsed in a report by Ireland’s Health Research Board and developed best practice contributed to the proposed DASSL Model (Data, Access, Sharing, Storage and Linkage) for safe access, governance, usage and linkage of data.
Public Outreach: ‘100 Ways’ Case Studies

100 Ways of Using Data to Make Lives Better

A series from The Farr Institute of Health Informatics Research showcasing the UK’s most significant examples of using data in research.

ClinTouch: An Innovative Mobile App to Support People Living with Psychosis

University of Manchester investigators at the Farr Institute of Health Informatics Research have developed a pioneering mobile phone app to improve the care and lives of people living with psychosis by monitoring real-time symptoms and reducing unscheduled and emergency hospital admissions.

Does Health at Birth Affect Educational Outcomes in Later Life?

Led by the University of Glasgow, investigators at The Farr Institute have discovered links between the delivery and health of newborn babies and performance in education later on in life.

Improving Care and Services for Patients with Mental Health and Substance Abuse Problems

Individuals experiencing both mental health and substance misuse problems have many complex needs that require high levels of care. This research project provided vital evidence to help improve and change how services are provided to patients.

100 Ways of Using Data to Make Lives Better

A series from The Farr Institute of Health Informatics Research showcasing the UK’s most significant examples of using data in research.

Did the Smoking Ban Work?

By analyzing data, researchers at The Farr Institute have shown how the introduction of smoke-free policies to a number of benefits to public health.

Improving Our Understanding of Multiple Sclerosis by Gathering and Analysing Data

The UK MS Register, funded by the MS Society, is the world’s first record, for any condition, to combine information provided by people with MS, with medical and NHS data.
Public and Patient Involvement & Engagement

- Social media: #dataspaveslives campaign
- Citizen Juries
- Science Festivals- Cheltenham, Manchester, Edinburgh, Swansea, London
- Including the public as co-researchers
Partnerships with Innovation Centres and Industry

- Development of new tools and methodologies
- Statistical and analytical consultancy in large and complex datasets
- Access to supercomputing infrastructure
- Randomised Control Trials & other data driven research
- Training and development

26th October 2016; part of Astrazeneca 2M genomes programme
“Collect Data Once, Use Often”

- Discovery Science
- Citizen-driven Health
- Data Analytics
- Precision Medicine
- Learning Health Systems
- Governance
- Public Health
- National Strategy & Leadership
- Partnerships

Interdisciplinary Skills & Capabilities

74 publications 2015/16
Recent Publications

The druggable genome and support for target identification and validation in drug development

Chris Finan, Anna Gaulton, Felix Kruger, Tom Lumbers, Tina Shah, Jorgen Engmann, Luana Galver, Ryan Kelly, Anneli Karlsson, Rita Santos, John Overington, Aroon Hingorani, Juan Pablo Casas

doi: http://dx.doi.org/10.1101/066027

Impact of statin related media coverage on use of statins: interrupted time series analysis with UK primary care data

Anthony Matthews,1 Emily Herrett,1 Antonio Gasparrini,2 Tjeerd Van Staa,2,4 Ben Goldacre,1 Liam Smeeth,1 Krishnan Bhaskaran1

Weekly variation in health-care quality by day and time of admission: a nationwide, registry-based, prospective cohort study of acute stroke care

Benjamin O'Reya, Geoffrey C Cloud, Martin A Jones, Harry Hemingway, Lizzie Polley, Kevin Stewart, Pippa J Tyrrell, Charles D A Wolfe, Anthony G Rudd, on behalf of the SSNAP collaboration

Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme

Mhairi Aitken1, Sarah Cunningham-Burley1 and Claudia Pagliari1

1Centre for Population Health Sciences, Medical School, University of Edinburgh, Teviot Place, Edinburgh, EH9 9AG, UK. Emails: Sarah.C.Burley@ed.ac.uk and Claudia.Pagliari@ed.ac.uk
Recent Publications

**Who Self-Weighs and What Do They Gain From It? A Retrospective Comparison Between Smart Scale Users and the General Population in England**

**Safer Prescribing — A Trial of Education, Informatics, and Financial Incentives**

**Prolonged dual antiplatelet therapy in stable coronary disease: comparative observational study of benefits and harms in unselected versus trial populations**

**Data Resource Profile: The Scottish National Prescribing Information System (PIS)**
Challenges and Opportunities
Towards a UK wide ecosystem
Big Issues
1: Complex environment

Interoperability: to work across systems with no additional effort
Big Issue 2. This is a tidal wave of data...
A gigabyte: 1000 megabytes
20 GB: Complete works of Beethoven

A terabyte: 1000 gigabytes
As of 2014, Wikipedia stored about 7 TB of information.

An exabyte: 1000 petabytes
Global Internet data: ~80 EB per month

A petabyte: 1000 terabytes
BBC iPlayer transfers 8 PB of programs every month.

A zettabyte: 1000 exabytes
World Wide Web: in 2015, holds 5 ZB of data...
Health care is becoming increasingly data intensive

- **Internet** and **cloud** provide connectivity to every corner of the globe
- **Smartphone**: 2 billion users; 80% of adult population by 2020
- **Socialome**: the digital data harvested for health and wellness
- **Quantified Self**: Non-invasive biometric sensing, Tricorder wearables. Apple Research Kit
- **Exposome**: Pervasive environmental sensing will bring new knowledge to public policy decisions about creating a healthier physical environment, and
- **$1000 genome** (genome, microbiome, transcriptome, lipidome, proteome, metabolome, multiome/panarome) Stem Cell and Genetic Tx (2000 + trials)
- **EHR data**: exponential growth of phenome from Electronic Health Records
- **Predictive Analytics** (Machine learning, A1, and Visualisation). Prediction: TenX more new knowledge from research in silico over RCT by 2020
- **Persuasive Technologies**: Behaviourial and motivational sciences
Big Issue 3 : Direct-to-patient

23andMe Turns Spit Into Dollars in Deal With Pfizer

Direct-to-patient trial recruitment

“After six hours, we have 7,406 people enrolled in our Parkinson’s study. Largest one ever before was 1,700 people” (Tweet from Parkinsons app developer, May 2015)

Stanford Uni cardiovascular trial attracted 11,000 volunteers in one day after releasing their ‘MyHeart Counts’ App in the App Store. It would normally take a year to enroll that many subjects (Bloomberg Business, March 2015)

Early days. Sampling challenges (noise, representativeness), but great potential....
Big Issue 4
Digital Maturity of Health Systems and Data

Level 8 - Personalized Medicine & Prescriptive Analytics
Level 7 - Clinical Risk Intervention & Predictive Analytics
Level 6 - Population Health Management and Suggestive Analytics
Level 5 - Waste & Care Variability Reduction
Level 4 - Automated External Reporting
Level 3 - Automated Internal Reporting
Level 2 - Standardized Vocabulary & Patient Registries
Level 1 - Enterprise Data Warehouse
Level 0 - Fragmented Point Solutions
EMR Adoption: World-Wide Benchmarks

HIMSS EMR Adoption Model Scores, Means per Segment, Q4/2014
(based on data from the last 24 months, no weighting applied)

All Acute Care Hospitals

<table>
<thead>
<tr>
<th>Country</th>
<th>EMRAM Score (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCLP (16)</td>
<td>2.3</td>
</tr>
<tr>
<td>Germany (220)</td>
<td>1.7</td>
</tr>
<tr>
<td>Spain (222)</td>
<td>3.5</td>
</tr>
<tr>
<td>France (14)</td>
<td>4.6</td>
</tr>
<tr>
<td>Italy (276)</td>
<td>2.4</td>
</tr>
<tr>
<td>United Kingdom (106)</td>
<td>3.4</td>
</tr>
<tr>
<td>Europe (1,166)</td>
<td>2.9</td>
</tr>
<tr>
<td>US (5,460)</td>
<td>4.4</td>
</tr>
</tbody>
</table>

*Europe includes: Austria (43), Denmark (24), France (14), Germany (219), Italy (271), Netherlands (63), Norway (3), Poland (3), Portugal (25), Spain (221), Switzerland (8), Turkey (143), United Kingdom (91)

The EMRAM algorithm differs from the European countries, United Arab Emirates, Saudi Arabia and Australia; data is based on Q3/2014
10 Recommendations

- National Engagement Strategy
- Capacity building
- Inter-operability
- Centres of Digital Excellence
Big Issue 5: Data Quality
Data maturity and standardisation
Pharmacogenetics of Metformin

1: 1 tablet twice daily
2: take one twice daily
3: 2 tablet twice daily
4: 1 tablet 3 times daily
5: 1 bd
6: take two twice daily
7: take one daily
8: 2 tablet 3 times daily
9: 1 tablet twice daily
10: take one twice a day
11: take one 3 times/day
12: 1 tablet daily
13: 2 tablet twice a day
14: 2 tablet bd
15: 1 tablet twice a day
16: 2 bd
17: 1 tablet bd
18: 1 tablet in the morning
19: take one 2 times/day
20: 1 tablet 3 times daily

5720 variations for Metformin alone!
Big Issue 6 Harnessing Inter-disciplinarity

Technologies

- Sensors
  - Bespoke
  - Ubiquitous networked
- Robotics
  - Single component
  - Full humanoid
- Natural language
  - Narrow target
  - Across media
- Speech recognition
  - Batch processing
  - Real-time natural
- Machine learning
  - Domain specific
  - Commodity tools
- Data linkage
  - Single database
  - Semantic Web
- Data architectures
  - Data warehouse
  - Cloud + havens
- Social computation
  - Individual intelligence
  - Social intelligence
- Security
  - Corporate security
  - Personalised security

Confluence
Big Issue 7 (The biggest of all) Governance of Trustworthy Use of Data

- Increasing transparency & reducing uncertainty
- Agreeing standards: Principles & Best Practices
- Clarifying Responsibilities: Data Flows & Data Controllers
- Seeking buy-in from stakeholders
Looking Ahead!

UK Institute for Health and Biomedical Informatics Research

Graham Spittle
Informatics Institute Strategic Advisor
Phase 1: £100m
MRC & Partner Investment:
Health & Biomedical Informatics Infrastructure (2012-2018)

MRC Medical Bioinformatics

Leeds

Oxford

Warwick-Swansea
- Cardiff
- PHE Wales
- Birmingham

UCL (eMedLab)
- EMBL-EBI
- Sanger
- KCL
- Crick
- LSHTM
- QMUL

Imperial
- EMBL-EBI
- Cambridge
- Nottingham
- Oxford
- Farr@Swansea

Uganda
- Sanger
- Cambridge
- Oxford

Farr@Scotland
- Aberdeen
- Dundee
- St Andrews
- Edinburgh
- Strathclyde
- Glasgow
- Leicester

Farr@HeRC
- Newcastle
- Lancaster
- York
- Bradford
- Manchester
- Liverpool
- Sheffield

Farr@CIPHER
- Swansea
- Cardiff
- Welsh Gov
- Bristol

Farr@London
- UCL
- LSHTM
- QMUL

HPA
- MRC CTU

Data Centre

Genomics England
PHASE 2
The UK Institute of Health and Biomedical Informatics Research

• A world leading interdisciplinary research institute

• A nationally coordinated programme of cutting edge data science to address the most pressing health challenges

• Distributed centres of excellence – brought together in a single, open and inclusive institute

• Partnership with health departments, charities and industry

• Led by internationally renowned director
Key Principles

• **Build on existing investments**
  - But not an exclusive club

• **Institutionally agnostic**
  - Open to broad partnerships and collaboration

• Create an informatics **ecosystem**

• **Predicated on team science** - valuing technical services

• Develop integrated multidisciplinary research groups
Core activities

• **Leadership** – deliver a co-ordinated national programme of research

• **Skills and capacity** - develop capability and expertise in “translational” health and biomedical informatics research

• **Secure research environments and data flows** - create secure, trusted and interoperable research environments

• **Analytics, tools and standards** - generate novel analytical tools for rapid translation into use.

• **Partnerships** – work with owners and controllers of data, NHS partners, academia and industry

• **Public Trust** – advocate for use of data and public engagement
Operational Model Principles

• **Model**
  - Separate legal entity
  - Distributed
  - Single set of Ts&Cs for collaborations
  - Full options appraisal to be conducted

• **Leadership**
  - Internationally-renowned, competitively appointed director
  - Small core team including a COO to support leadership and management
  - Director administered budget

• **Funding**
  - Long term, sustainable support
  - Core support of~ £45-50m over 5 years
  - Supported in partnership with other funders
Options and Opportunities for UK Health Data Science

- **Commitment** to communication, collaboration and public engagement in everything we do
- **Convergence** of care with innovation and research
- **Clinical data quality** – Bringing routinely collected data up to same standard as high quality research data
- **Collaboration** of health care providers/ academia
- **Commercial engagement** - encouraged but with transparent governance, collaborative and benefit sharing
- **Computer Science** key ingredients for change – educational implications
- **Clarity** about GOVERNANCE and data sharing

An Opportunity for the UK to Lead the Way?