Introducing the DSI

Professor Yike Guo
Imperial College Data Science Institute: A Focal Point

**FACULTIES**
- Faculty of Engineering
- Faculty of Medicine
- Faculty of Natural Science
- Imperial College Business School

**STRATEGIC APPLICATIONS**
- Health, Wellbeing & Personalised Medicine
- Discovery Science
- Sustainable Development
- Energy & Environment of Future Cities
Data Science Institute
Mission

1. Focal point for co-ordinating data science research
2. Supporting data driven research
3. Educating a new generation of data scientists
4. Advise College on data strategy and policy
5. Translation of data science innovation by close collaboration with industry
1. Focal point for co-ordinating data science research

**Medical**
- Systems oncology
- Realtime metabolic profiling
- Cardiovascular science

**Natural Sciences**
- Imaging
- Particle physics
- Infection/epidemiology
- System biology
- Neuroscience

**Engineering**

**Business**
- Social media & new data business
- Algorithmic trading
- Digital cities & urban life
- Public health management

**Complex systems & network data analytics**

**High-throughput screening**
Molecular profiles

Physiological measurements

Environmental monitoring

Data Driven Healthcare
Energy & Environment of Future Cities: Human Mobility

Understanding human urban mobility is key to better city planning.

London
- 270 tube stations
- 1 year of data
- 1.23 billion passengers per year
Health, Wellbeing & Personalised Medicine
Health and Development

Understanding the relationships between global health and economic development

187 countries
34 years
Diabetes, obesity, GDP per capita and diet
Health, Wellbeing & Personalised Medicine

Health and Development

Food balance
(Mediterranean vs Western diet)

Year
1980

Male obesity (% of population)

Male diabetes (% of population)

GDP per capita (current US dollars)
2. Supporting data driven research

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^{n} L \left( y_i, F_{m-1}(x_i) - \gamma \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial f(x_i)} \right)$$
eTRIKS: European translational informatics platform

**eTRIKS repository**
- Clinical and Molecular ('omics) data
- Scientific Evidence Repository
  - CDISC and ISA standards
- Study-centric

**eTRIKS analytics**
- Open Access Analytics
  - eTRIKS: European translational informatics platform
  - R-cloud

**eTRIKS research commons**
- KEGG
- Uniprot
- Atlas
- Entrez

**TM Knowledge Base**
- Research Results and Findings
- Science-centric

**Collaborative research access platform**

**eTRIKS user interface**
• Example of simulating full body temperatures and energy transformation
• Core body temperature & Safety
• Data assimilation with body sensor data for building individual models

- Age: 27, Weight: 80kg
- 1-60 min running at a speed of 6mph with air temperature 30°C
- 60-120 min sit still with air temperature 42°C
- 120-180 min sit still with air temperature 10°C
- 180-240 min running at a speed of 6mph with air temperature 10°C
eTRIKS and Wiki-Health: Big data platform for personalized medicine

Smoothed images → Single subject GLM → Contrast maps → MVPA → Activation map

Hemodynamic Response Function

eTRIKS/tranSMART

 Ontology

 Data Model

 Analytics

 Phenotypic profiles

Molecule perturbations

Effects

EHR

Laboratory test

Scan image

Diagnosis code

Discharge letter

Medical notes

Data collection app

Clinical data

09/03/2015

Yike Guo (y.guo@imperial.ac.uk)
Research in 2015

Transdisciplinary Research
- Medicine: eTRIKS apply to 4 disease fields
- Engineering: Wikisensing for smart city
- Health: Wikihealth for bodysensor
- Business School: data economy

Data Science Methodology
- R machine
- Deep learning
- Statistical HPC

Visualization
- GDO
- Imperial Visual APP

Infrastructure: ICBig
Research Collaboration 2015

Europe
- EU H2020

China
- ZJU Collaboration Consolidation (PhD scholarships and IC/ZJU joint projects)
- Apply for Newton Funding
- One more joint Lab with Chinese organization

JAPAN
- Collaboration with Research Organization of Information and Systems (MoU in Feb)
- Joint projects via ROIS

USA
- Co-design education course
- Focus on industry (IBM)
3. Educating a new generation of data scientists

Data Science Institute:

The introduction of data science modules, with a balance derived from Computing, Mathematics, EEE, and the Business School, as additional electives within existing MSc programmes across Imperial’s four faculties, to teach core topics in data science within domain specific specialisms.

To launch in the academic year 2016–2017
Education 2015
Continuing Professional Development

AIM: Up to five modules will be offered by the DSI as CPD courses for industry professionals and PhD students.

1. Foundations of Data Science
2. Analytics and Machine Learning
3. Programming Big Data
4. Very large Data Management and Visualization
5. Data Economy

2015 Target: By May have advertised one CPD course to run in late 2015.
Education 2015/16
MSc X (Data Science)

The same modules for the CPD will plug into existing MSc courses around College in place of existing elective modules or streams to create a Data Science Stream.

1. Foundations of Data Science
2. Analytics and Machine Learning
3. Programming Big Data
4. Very Large Data Management and Visualization
5. Data Economy

Choose 4

MSc Materials with Data Science

During 2015 DSI will be designing these modules, enlisting teaching staff, liaising with Departments and gaining internal approvals.

2015 Target: By Dec 2015 have advertised the Data Science stream in at least 2 MScs for Sept 2016 uptake
4. Advise College on data policy and strategy

Impact and Influence

- DSI Book – Springer
- Distinguished Speakers Series
- Data Insight Speaker Series
- Host major international conference in Big Data for Brain Health, 8/15
- Leading the “International Data Science Academy “
- Research Data Management at Imperial College
5. Translate innovation with industry partners

- Industry funded labs
- Joint research labs for academic collaboration
- Joint research programme
- Industrial funded PhD studentships
- Visiting researcher scheme
Collaborations with Industry

Data Science Institute

Huawei Data Science Innovation Lab

KPMG
Collaborations with Industry 2015

Data Science Institute

Huawei Data Science Innovation Lab
- Establish 5 projects across the main themes
  - high performance deep learning
  - scalable models of information propagation in networks
  - scalable algorithms for mining high-dimensional / high-frequency data
  - body sensor networks and informatics
  - big data for big science
- Hire Lab Manager

Add at least 1 additional major UK partner
- Explore relationships with IBM, Shell, BT, Bank of England and Vodafone

KPMG
- Build GDO and establish as a core facility
Creating a Hub

Physical Location
- Open DSI Hub on SK site in May 2015
- High profile launch event in October 2015, inviting global partners
- Develop hub and spoke model for researchers, allowing integration across all campuses

Visualisation
- Develop and build GDO
- Advertising space – content, relationships, finance model

Finance model
- Research o’heads, visualisation content, CPD course, MSc Stream, PhDs