Data Science and Analytics - Skills Gaps and Innovation Opportunities

Roger M Whitaker

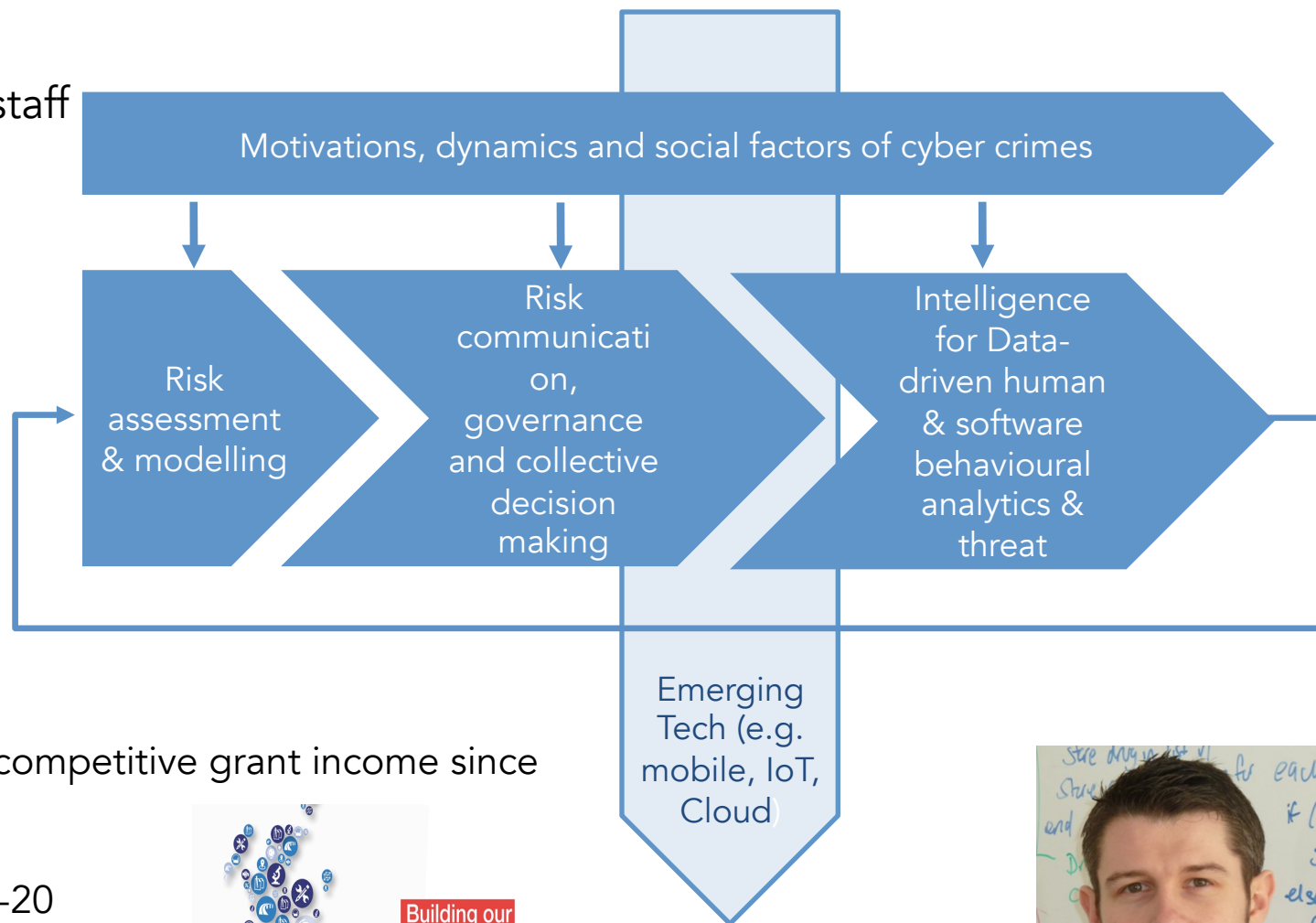SOCIAL SCIENCE PARKS — SOCIETY'S NEW SUPER-LABS

- A practical innovation in facilitating **interdisciplinary** social science research and its application.

- SPARKs are purpose-built facilities housing applied social science research groups alongside researchers from other disciplines, **external research stakeholders and collaborators from the private, public and third sectors**

- Facilities and physical spaces that **encourage creative interaction** and promote the adoption of collaborative approaches to research and impact.

Social Science Research Park

Data Innovation Lab

During 2017-2020 Supercomputing Wales will:

Deliver upgraded supercomputing hub facilities

Create a network of highly skilled Research Software Engineers

Develop a national specialist technical team

Facilitate innovative scientific collaborations with industry and other partners

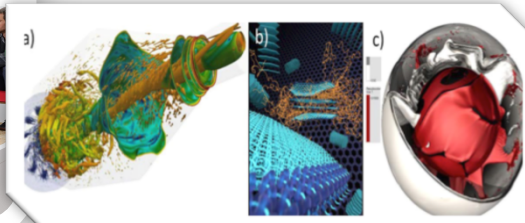Capture more competitively awarded research funding

Provide a step-change in research within Science, Technology, Engineering and Mathematics

The national supercomputing research facility for Wales

SUPERCOMPUTING WALES
UWCHGYFRIFIADURA CYMRU

12,000 cores in total:3 "lots" at two locations
Lower TRL level research collaboration (typically)
Simulation and modelling focused (2 lots)
NEW - **HPDA: Big data meets HPC** (1 lot)

UNDEB EWROPEAIDD
EUROPEAN UNION

Llywodraeth Cymru
Welsh Government

Cronfa Datblygu
Rhanbarthol Ewrop
European Regional
Development Fund

PRIFYSGOL
BANGOR
UNIVERSITY

PRIFYSGOL
ABERYSTWYTH
UNIVERSITY

Prifysgol Abertawe
Swansea University

CARDIFF
UNIVERSITY
PRIFYSGOL
CAERDYDD

National Data Science Academy

To provide high quality employment-ready graduates in **data science** and **artificial intelligence**;

To work **across disciplines** to provide a high quality and distinctive educational experience, primarily focusing on Masters provision and professional courses;

To provide a **University-wide unified point of contact** to sustain relationships with employers and industry concerning data science and artificial intelligence;

To further **create critical mass** and capacity around the data science and artificial intelligence.

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

Courtesy of MarketingDistillery.com

Data Innovation Research Institute

Computational social science

As humans increase their use of technology, so does so does the size of their digital footprint providing us with a unique perspective on social and human behaviour.

Biological and life sciences

The medical and life sciences provide huge data sets which are ideal for researching for a variety of uses and applications.

Computational science and engineering

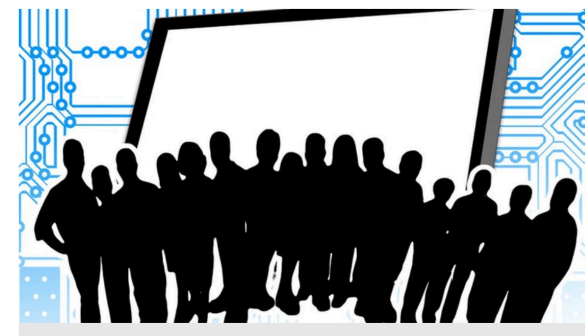Our research can be applied to numerous applications in the fields of physics, astronomy and engineering.

New centre to train next generation of data scientists
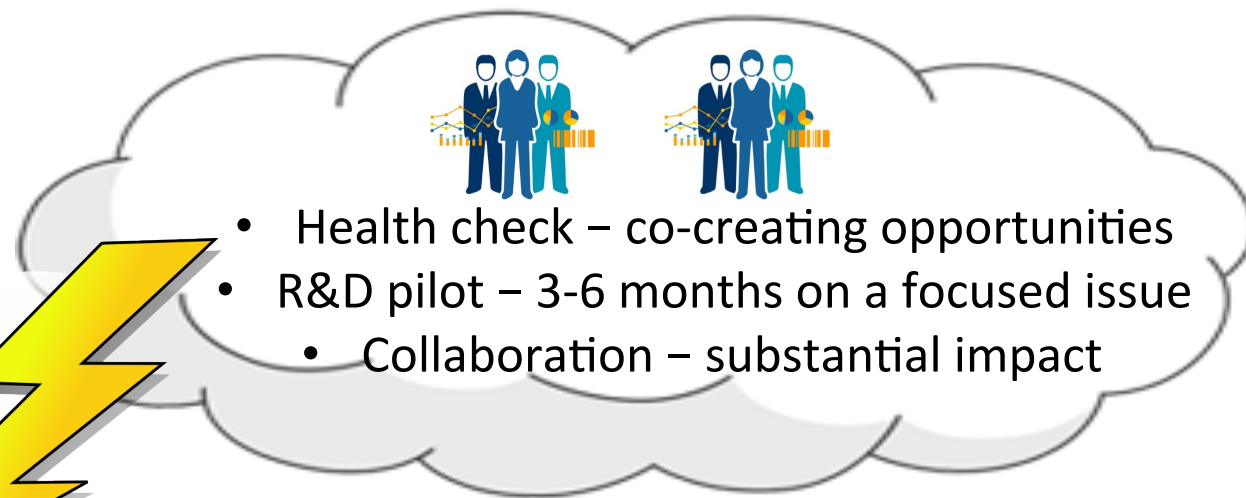
27 October 2017

Cardiff Research Software Engineer Conference

27 March 2018

An invitation to attend Cardiff's first Research Software Engineer Conference 30th April 2018

Data Innovation Lab @ SPARK

DATA INNOVATION ACCELERATOR

- Health check – co-creating opportunities
- R&D pilot – 3-6 months on a focused issue
- Collaboration – substantial impact

SMEs

Data Innovation Research Institute

CARDIFF UNIVERSITY / PRIFYSGOL CAERDYDD

UNDEB EWROPEAIDD EUROPEAN UNION / Llywodraeth Cymru Welsh Government
Cronfa Datblygu Rhanbarthol Ewrop
European Regional Development Fund

Steve Fairhurst – Director
FairhurstS@cardiff.ac.uk

Data Innovation Research Institute

Owain Huw – Programme Manager
HuwO1@cardiff.ac.uk

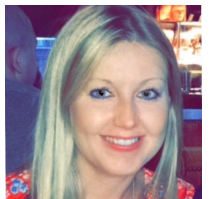SUPERCOMPUTING WALES
UWCHGYFRIFIADURA CYMRU

Pete Burnap – Director
BurnapP@cardiff.ac.uk

AIRBUS Centre of Excellence for Cyber Security Analytics

Rick Delbridge – Dean of R, I & E
DelbridgeR@cardiff.ac.uk

Social Science Research Park

Sarah Woods – Project Officer, Innovation System Team
WoodsS1@cardiff.ac.uk

National Data Science Academy

Roger M Whitaker – Co-director (with Pete Burnap)
WhitakerRM@cardiff.ac.uk

DATA
INNOVATION
ACCELERATOR