

## SCHOOL OF INFORMATICS – investment proposals

### Introduction

The digitalization and transformation of society through the effective use of data is likely to be the defining change in the early 21st century. The School of Informatics is internationally leading in a range of areas related to this phenomenon. In particular databases, machine learning, internet of things, and security and privacy. To position the school to play a leading role in the research and innovation underpinning this transformation we see the need to develop three additional areas related to data and digitalisation. These areas will help stimulate new research and innovation challenges, provide a clearer route to impact for our fundamental research and will provide stronger linkage to EPCC and to the City Deal. All of these posts strongly support the development of our proposed CDTs in AI, Security, Parallelism and Biomedical AI. In addition we see a strong opportunity and need to build partnership with the Usher Institute in CMVM to focus on the development of data science in the medical domain. We seek to appoint at reader level in the following areas:

- **Data Infrastructure:** As volumes of data increase and the embedding of data into processes advances apace, there is an increasing need for work on the infrastructure that manages and facilitates the deployment of data science into real world contexts. This will involve some understanding of the physical infrastructure supporting the management and control of data but will focus on the software necessary to manage data and provide systems with the performance crucial to support modern data science algorithms. This post lies at the intersection of our interests in Databases, Machine Learning and Systems (e.g. Internet of Things systems). The research and innovation supervised by this post will have direct relevance to the construction of data infrastructures in enterprises and in the support of data-driven innovation.
- **Data Governance:** As we are seeing at the moment with the Facebook controversy, the broad issue of data governance will be a dominant concern in the coming decade or more. This is a technical area that embraces topics such as differential privacy, explainability and transparency while linking to a broad legal and policy community in University of Edinburgh. We envisage this interaction will be focused (but not exclusively) on EFI. This post lies at the intersection of work on Security and Privacy, Machine Learning and Databases with strong links to law and policy. The work will be to develop new approaches to governance, exploiting technologies such as distributed ledgers and articulating their relevance to legal and policy issues as well as data governance in enterprises and the public sector. Depending on the development in this area there is the potential for strong linkage to work on Health Data.
- **Medical Data Science:** With the award of a Centre in the UK Health Data Research organisation there is considerable opportunity for the development of work that synergizes routinely collected medical information with deeper analysis of disease mechanisms that arise

from our growing understanding of disease at the level of the genome and mechanisms underlying phenotypic variation. This has the potential to impact service delivery when these mechanisms are understood at a population level and in the development of new treatments for individuals. This post will investigate the data sources and work to develop new techniques and algorithms that utilize the data resources available in Edinburgh as effectively as possible. For example, this post would bolster the interdisciplinary potential of the proposed CDT in Artificial Intelligence, Applications and Implications. We see this post as a nucleus around which we can further develop our collaboration with CMVM around data in Medicine.

### Business Case Elements

- **Data Infrastructure:**
  - **Research:** This is a highly active area of research as we struggle to design infrastructures that will scale to the demands of ever growing data volumes. This issue poses important research challenges and their practical solution has considerable innovation potential.
  - **Teaching:** This post will contribute to our teaching in data science. This will help support the development of our Data Science Graduate Apprenticeship degree programme and will cover critical areas in applied data science and in the infrastructure to support secure operations. We envisage this post will make significant contributions to the Data Science and Cyber-Security Masters programmes.
  - **Commercialisation:** The ubiquity of data in the commercial world provides myriad opportunities for industry engagement, with links to recent and planned appointments in Business Applications of Informatics.
- **Data Governance:**
  - **Research:** This is a rapidly growing area of interest and we are perfectly situated to build a significant research portfolio in this area: <https://royalsociety.org/topics-policy/projects/data-governance/>
  - **Teaching:** This post would make significant contributions to the EFI teaching programmes and would strengthen our Cyber Security offering significantly by making links between the technical mechanisms and legal and policy matters. This has the potential to generate significant revenue both from our MSc in Sec/Priv and from EFI programmes.
  - **Commercialisation:** Companies are particularly concerned about this issue, particularly in the wake of the Facebook, Cambridge Analytica scandal. There are significant opportunities in Executive Education and partnering with companies on the development of governance systems. Again, there are links to our developing area of Business Informatics.
- **Medical Data Science:** We see this post predominantly building capacity to deliver on the UKHDR in Edinburgh
  - **Research:** In research the goal is to develop a portfolio of work

that furthers the UKHDR development of cross-cutting analytical methods, tools and standards by developing new algorithms and techniques that are relevant to the medical landscape.

- **Teaching:** We see this individual acting as a focus for the Informatics contribution to UKHDR interdisciplinary skills and capacity through the development of new courses and programmes in collaboration with the Usher Institute:  
<https://www.mrc.ac.uk/mrc/assets/Image/proposed-career-framework.jpg>
- **Commercialisation:** Applications are clearly in the public health care sector in the UK but also opportunities in the private sector which serves UK health care and in international health care organisations.

### Summary

We see linkages between all three of these posts which we believe are complementary and which also build and extend existing areas of research and teaching within the School. All have significant translational potential. We therefore see strong academic and business cases for investment in these posts.