THE UNIVERSITY of EDINBURGH informatics

REF Impact Case Study Development

The table below outlines our 'long list' of potential REF Impact Case Studies (not in any priority order at this stage). The table includes provisional titles, areas of impact, envisaged nature of metrics and person responsible for supplying them, indication of which case studies were submitted to the previous REF exercise, and which cases are multidisciplinary.

A brief narrative summary for each potential case study follows below the table.

No	Title	Area of impact	Metrics to be supplied (type/ person responsible)	Submitted before?	Multi-disciplinary
1	AppGuarden resilient application stores	Commercial	No. of users/downloads. David Aspinall		
2	Dermofit, a skin cancer detection app for trainee medical professionals	Health	Scale of deployment/no. of patients. Bob Fisher		
3	RemIX rural broadband network	Public services / economy / quality of life	No. of users, installations and imitators. Peter Buneman.	Y	
4	Text to speech synthesis	Health	No. of patients. Simon King	Y	Y
5	DexFuzz: Application of Domain-aware Binary Fuzzing to Aid Android Virtual Machine Testing	Commercial	Scale of deployment. Björn Franke		
6	Tackling inequality and deprivation through digital connectivity	Policy	Policy uptake. Mike Fourman		Y?
7	Edinburgh Living Lab	Society/environment / economy	No. of users affected. Ewan Klein		Y?

8	Next generation prosthesis: enhancing control and sensation in the i-Limb robotic hand	Health	No. of patients. Sethu Vijayakumar		
9	EnCore Microprocessor and the ArcSim Simulator	Commercial	Volume of chips shipped/ energy saved/revenue. Nigel Topham	Y	
10	Optimising big data	Commercial	No of users affected/ energy saved/ applications enabled. Wenfei Fan		
11	RESpeck: Remote Monitoring of COPD Patients using Speckled Computing	Health	No of patients. DK Arvind		
12	Streamlining Qualcomm Hexagon DSP firmware	Commercial / economy	No of Hexagon units, scale of deployment. Björn Franke		
13	WorkflowFM	Public services / health	No of patients, resources saved. Jacques Fleuriot		
14	FiveAI self-driving cars	Commercial / society	No. of cars. Ram Ramamoorthy		
15	Neural translation: MOSES, KenLM, Nematus, AmunMT	Commercial	No. of users, including high profile. Kenneth Heafield		
16	Speech synthesis for [multilingual] spoken content production	Public services/ Society	Audience figs/no. of languages / volume of content. Simon King.	Y	
17	Modelling city-centre public transport to minimise inefficiency, cost and carbon emissions	Commercial / public services / environment	Resource & efficiency savings. Jane Hillston		
18	Neuropolitics Research Lab (new, one to watch)	Policy?	TBC. Robin Hill.		Y

19	Public engagement	Public engagement	Audience figures, pupils affected by curriculum change. Kasia Kokowska		
20	Quantum verification	?	TBC. Elham Kashefi		?
21	Blockchain Technology Laboratory	Commercial/other TBC	TBC. Aggelos Kiayias		Y
22.	Automatic speech recognition: Quorate Technology/ Olly by Emotech	Commercial	(1) Volume of subtitles, no of viewers. (2) units shipped, no. of users. Steve Renals		
23	Combinatory Categorial Grammar (CCG)	Commercial	Uptake. Mark Steedman		
24	Speech Graphics	Commercial	Market value/uptake Steve Renals / Hiroshi Shimodaira	Y	
25.	eXtensible Markup Language (XML)	Commercial	Scale of use. Henry Thompson, Philip Wadler, Peter Buneman	Y	
26.	Ultrax speech therapy	Health	No. of patients. Steve Renals.		Y
27.	Intelligent warehousing	Commercial	Robots deployed. Savings made. Sethu Vijayakumar		
28	Wikisimple	Commercial	Uptake. Mirella Lapata		
29	Object detection – annotating images in large image banks	Commercial	Uptake. Vittorio Ferrari		
30	Bonseyes: improving efficiency and security in low power IoT devices	Commercial	Uptake. Amos Storkey / Mike O'Boyle		
31	Formal Semantics of Graph Query Languages	Commercial	Uptake. Leonid Libkin		

Summaries

1. AppGuarden resilient application stores
Application stores are very successful eg. July 2008-September 2016, 140 billion apps were downloaded from Apple's App Store and in May 2016 Google stated that 65 billion apps had been downloaded from Google Play. However, app stores and the phones, tablets and other devices connected to them are not immune to damage or cyber attack. AppGuarden provides algorithms that will automatically analyse apps to ensure they are safe, adds 'digital evidence' to increase safety and employs simple, user-friendly security policies and permissions to ensure the user's data only goes where it needs to go. One example is EviCheck, a tool for the verification, certification and generation of lightweight fine-grained security policies for Android.

2. Dermofit, a skin cancer detection app for trainee medical professionals
There are approximately 13 million GP consultations for skin conditions a year and 716,830 new referrals. GPs receive little dermatological training and there are only 650 consultants to advise and provide specialist care. Many skin conditions can be managed cheaply and effectively in primary care. Dermofit is a digital tool which helps trainee and non-specialist healthcare professionals to detect skin cancer more effectively. It uses algorithms that automatically group over 1300 library photos of skin lesions based on their colour and texture. By allowing users to compare different kinds of lesions, Dermofit can reduce the number of unnecessary referrals by GPs to specialist consultants. Dermofit was a Wellcome Foundation funded research project, commercialised by digital healthcare specialist Simedics Ltd (www.simedics.org).

3. RemIX rural broadband network
RemIX is a distributed internet exchange which provides a broadband service across a 2000 square km area of the west coast of Scotland, spanning sea and mountainous mainland. It allows small networks to access good quality and affordable backhaul - the portion of the network which links the core backbone of a hierarchical telecommunications network and the small subnetworks at the 'edge' - to provide broadband services to remote, sparsely populated areas. RemIX was engineered to be easily reproducible, so it is hoped that this model can be replicated in other rural areas struggling with poor connectivity. The project won an European Broadband Award 2016 in the 'Future-proof and quality of service' category.

4. **Text to speech synthesis**
Motor Neurone Disease (MND) is a rapidly progressing terminal illness which affects up to 5,000 adults in the UK at any one time. For around 25% of people diagnosed with MND, their first symptom is a problem with their speech. A person's voice is not only fundamental to the way they communicate, it is an essential component of his or her identity. Edinburgh's Centre for Speech Technology Research has developed the personalised speech synthesis technology behind Speak:Unique (http://www.speakunique.org/), a project by the Euan MacDonald Centre for Motor Neurone Disease Research and The Anne Rowling Regenerative Neurology Clinic which aims to equip people with MND with communication aids that speak for them in their own voice.

5. DexFuzz: Application of Domain-aware Binary Fuzzing to Aid Android Virtual Machine Testing
The development of a new application virtual machine (VM), like the creation of any complex piece of software, is a bug-prone process. DexFuzz is a tool, developed with the Android team at ARM, that combines domain-aware binary fuzzing and differential testing, exploiting the presence of multiple modes of execution within a VM to test for defects.

6. Tackling inequality and deprivation through digital connectivity
Digital exclusion compounds other forms of deprivation by denying access to opportunities in education, employment and health. The Gini index has been used as an indicator of the digital divide. However, increased uptake among the better off can reduce the Gini index while widening the gap between rich and poor. Fourman has shown that Wagstaff's variation on the Gini index (developed at the WHO in 2005, in the context of health inequality) provides a direct measure of the effects of digital inequalities on existing deprivation. This index is applied to identify communities where existing deprivation (as measured by the Scottish Index of Multiple Deprivation) is exacerbated by digital exclusion, and help both the Scottish Government and third-sector organisation to target interventions intended to reduce these effects.

7. Edinburgh Living Lab
With a population of half a million, Edinburgh is one of the biggest cities in the UK. Edinburgh Living Lab was founded by the City of Edinburgh Council and the University of Edinburgh's School of Informatics to bring together academia, public sector, industry, third sector and local people to design, test and implement new services, processes and products that generate social, environmental and economic value. Research focuses on issues around mobility, energy and learning by developing.

8. Next generation prosthesis: enhancing control and sensation in the i-Limb robotic hand
First invented by Edinburgh Engineering alumnus David Gow in 2007, the i-Limb prosthetic robotic hand, developed by Livingston-based company Touch Bionics, is unique in that it has highly advanced feedback sensors that allow wearers to control grasp, position fingers and feel/ sense fingertip sensation. Our Statistical Machine Learning and Motor Control Group worked in collaboration with Touch Bionics, fitting subjects with the i-limb and an array of vibrating motors to communicate feedback from force and position sensors to the wearer. [By the end of 2015 over 5,000 patients had been fitted with Touch Bionics i-limb and i-digits products and i-Limb had become accepted as a reimbursable health insurance cost in USA, Canadian and French healthcare markets.]

9. EnCore Microprocessor and the ArcSim Simulator
Created in 2009, the EnCore microprocessor and associated ArcSim simulation software achieved global impact after being brought to market by Synopsys Inc., a world leader in Electronic Design Automation and Semiconductor Intellectual Property. It is the second largest supplier of processor IP, by unit volumes shipped by its customers. The processors in its current portfolio of products are derived from Encore technology, and since 2013 the worldwide uptake of this technology has grown significantly. Full details of Synopsys' customers and their products based on Encore derivatives are commercially confidential, but: Intel announced that they use an ARC EM processor core "at the heart" of their sensor subsystem on their latest IoT edge chips; The ARC EM processor was the first Synopsys product based on the Encore processor technology.

10. Optimising big data
Big data analysis offers big benefits in terms of faster, better decision making for organisations and customers but also brings big challenges in terms of managing volume, variety, velocity (change) and veracity (quality). Research which improves data quality and the speed, security and accuracy of analysis impacts on millions of people worldwide, when it impacts on companies like Microsoft, IBM, Facebook or Google. One global information and communications technology (ICT) provider, Huawei, has seen some of its systems improved by 10^5, following projects such as: Query optimisation techniques for massive data volume in analytical databases.

11. RESpeck: Remote Monitoring of COPD Patients using Speckled Computing
The British Lung Foundation estimates that 1.2 million people in the UK live with diagnosed Chronic Obstructive Pulmonary Disease (COPD) and that it is the second largest cause of emergency admission in the UK. This is costly. In 2011, the National Institute for Clinical Excellence (NICE) estimated that an inpatient admission for COPD cost £1960. The RESpeck respiratory monitor is a wireless patch that measures breathing and activity, sending data via broadband or mobile phone to a care manager. This allows patients to be monitored at home, preventing unnecessary hospital visits and irreversible damage to health. Partners: Moray Council, NHS Grampian, The Maryhill GP Practice, Elgin

12. Streamlining Qualcomm Hexagon DSP firmware
Efficient processing is vital to delivering advanced multimedia on mobile devices. It also affects manufacturing costs and power consumption. Qualcomm Technologies developed the Hexagon Digital Signal Processor (DSP) to support the processing needs of developers working on the mobile platform. Working with Qualcomm, our researchers reduced the code size of a business-critical piece of the Hexagon DSP firmware by 10%, substantially lowering unit costs. Between 200 million and 1bn units have since been shipped. The novel LLVM compiler pass included in Qualcomm's Hexagon SDK is now used by developers world-wide, and was released as open-source, enabling others to benefit from the work.

13. WorkflowFM
Researchers have been working closely with clinical teams such as the sexual health team at the Chalmers Sexual Health Centre, NHS Lothian and the Care Of Burns In Scotland (COBIS) group at the Glasgow Royal Infirmary, NHSGGC to help them clarify and record their daily processes, optimize their schedules and organise and improve patient care. In addition, they use the WorkflowFM tool developed as part of their research, to build models of care, helping to remove inconsistencies and reduce errors, streamline health services to avoid omissions and reduce drop out rates, record outputs and perform quantitative analysis (such as completion rates, duration and cost of care, etc), identify potential problems and unnecessary delays in daily practices as well as key points to improve care integration.

14. FiveAI self-driving cars
Building on research into artificial intelligence to ensure driverless vehicles can interact cooperatively and safely with the world around them. [Speaking in Five AI press release, Jan 2017, Ram Ramamoorthy says: "We now know how to use context-sensitive observations and learnt behaviours to predict what actors will do, update those predictions at high frequency and so build systems that predict and co-operate with other actors just like safe human drivers. Developing, integrating and commercialising this novel science into systems will help differentiate the Five AI solution."]

15. Neural translation: MOSES, KenLM, Nematus, AmunMT – update closer to submission
The open source statistical machine translation (MT) toolkit Moses is widely used commercially and by international organisations such as the United Nations, European Commission and World Intellectual Property Organization. The economic impact of this technology is estimated to be about $45 million per year according to a market report but usage has not changed significantly, 2013-2017 and the field of MT has changed to focus on neural methods. Open source products in this area (Nematus, AmunMT and Marian), partially developed at Edinburgh, are starting to gain traction outside of academia. For instance, WIPO (the World Patent Organisation) has a deployed system using Nematus/AmunMT and the EC's translation office, and Samsung Poland are experimenting with Nematus. [*WIPO, probably the UN soon, EC trials, Intel wants it to be a benchmark.]* KenLM has a slightly broader user base than Moses (since it's not restricted to MT). SRI is using it, despite having written the previously-dominant SRILM.

16. Modelling city-centre public transport to minimise inefficiency, cost and carbon emissions.
A well functioning public transportation network can help a modern city minimise carbon emissions and reduce operating costs. The task of formulating a sensible bus service is non-trivial, and is complicated by the fact that the service is subject to random behaviour. In this project, powerful stochastic models were used to assess whether current service requirements can reasonably be met and to gauge the impact of proposed changes to the network. The methodology was used to investigate the effect of the introduction of trams to the City of Edinburgh.

17. Speech synthesis for spoken content production (multilingual broadcasting) – new, one to watch
The government has set the BBC a target of reaching a global audience of 500 million people by 2022, compared with today's 308 million. The only way to reach such a huge audience is through new language services and efficient production techniques. Technology which automatically produces speech from text offers an attractive solution. Synthetic voices have the potential to create news broadcasts in languages such as Swahili, Hausa, Somali and Amharic, enabling the BBC to reach a larger audience in more countries around the world.

18. Neuropolitics Research Lab – new, one to watch
The Neuropolitics Research Lab produces transdisciplinary research, utilising developments in the cognitive neurosciences, to shed new light on political attitudes, identities and decision-behaviours. The aim is to test the utility of methods more typically associated with neuroscience, informatics and cognitive psychology in helping us to understand more about political attitudes and behaviours.

19. Public engagement
Informatics harnesses all channels in breaking down the barriers to public understanding of IT.

- Harmonium: researchers from Design Informatics used cognitive data capture to transform the eye movements, pulse rates and facial expressions of the Edinburgh International Choir into abstract digital animations, producing a spectacular and highly publicised opening to Edinburgh International Festival 2015, attracting over 20,000 people on the night and positive media coverage around the world.
- Professor Alan Bundy is founder/adviser for a Royal Society of Edinburgh (RSE)/BCS Chartered Institute of IT project to develop materials for the teaching of computing that exemplify the Curriculum for Excellence. He is a key member of the Scottish Professional Learning and Networking for Computing (PLAN C) project and has represented BCS on the Learned Societies' Group on Scottish Science Education, which advises on the teaching of STEM subjects in Scottish schools.
- As the researcher responsible for bringing the first NASA Valkyrie humanoid robot to Europe, as an expert speaker in the media and as a judge on the BBC 2 television show show Robot Wars, Professor Sethu Vijayakumar has inspired an interest in AI and robotics among millions.
- At grassroots level, Professor Ewan Klein established PreWired, Edinburgh's programming club for under 19s, which has been running regular weekly meetings since 2013.
- College Outreach Champion Dr Areti Manataki introduces people of all ages to elementary programming through fun introductory workshops while her online course, 'Code Yourself! An online introduction to programming,' which has attracted over 120,000 participants worldwide. Available in Spanish as well as English, the MOOC is run in collaboration with Universidad ORT Uruguay.
- Hugh Leather set up and runs Compucast, a public podcast about computer science. c15k listeners.

20. Quantum verification – one to watch
The ability to compute with encrypted data, while hiding the underlying function, has opened a new approach toward verification, through the detection of a cheating server. In every aspects of quantum information processing, from complexity theory to cryptography, from simulation to sampling, from tomography to implementation, from foundation to interpretation, over and over we are hitting the wall of efficient verification. A milestone for an efficient verification scheme was proposed and demonstrated experimentally in 2013 and is now a focus of interest for: UK national Program (NQIT quantum technology hub), US research agency (Airfroce), France industry R&D (ATOS under negotiation), Europe Quantum Falgship (QSoft Platform under development), Singapore Research Agency (NRF fellowship). A company has just been set up.

21. Blockchain Technology Laboratory – one to watch
The Blockchain Technology Laboratory was established within Informatics in February 2017 by IOHK, a leading blockchain research and development company, and will serve as the headquarters for IOHK's growing network of global university partnerships. Research collaborations will be interdisciplinary and will include, beyond cryptography and computer science, economics, game theory, regulation and compliance, business, and law. The lab will provide a direct connection between developers and researchers, helping to get projects live faster and aims to pursue outreach projects with entrepreneurs in Edinburgh's vibrant local technology community. With a focus on industry inspired problems, the potential impact is significant – one to watch.

22. Automatic speech recognition
Automatic speech recognition technology developed at Edinburgh is now in deployment in multiple domains: (1) Broadcast subtitling. Speech recognition and alignment technology, developed by UoE spinout Quorate Technology is used in the process for subtitle generateion by Ericsson who supply subtitles for BBC, C4, Sky. (2) Emotech, a UK startup, have produced Olly a personal tabletop robot which won multiple awards at this year's CES in Las Vegas. The speech recognition component of Olly was developed by a recent UoE PhD student using approaches developed and published at UoE. Metrics: number of units shipped; number of users

23. CCG
Combinatory Categorial Grammar (CCG) is an efficiently parsable, yet linguistically expressive grammar formalism, which was developed computationally by Mark Steedman and his students and postdocs. The CCG has enabled large scale semantic parsing, amenable to machine learning techniques, able to cover the broad range of text encountered on the internet. CCG is underpinning semantic parsing technology further developed at Facebook, Google, etc. as well as well-funded startups such as Semantic Machines.

24. Speech Graphics
Speech Graphics, a UoE startup, has a unique technology (based on UoE research and award-winning publications) for audio-driven animation that may be used to reanimate head animations to include lip synchronisation. This technology has been used in a number of high profile games and music videos.

25. eXtensible Markup Language (XML)
The success of the eXtensible Markup Language (XML) has been due in large part to the technologies built around it for constraining, querying, styling and otherwise processing XML documents. Research carried out at Edinburgh has been instrumental in the creation and/or design of many of these core XML technologies, including XSLT, XML Schema, XInclude, XQuery and XProc. Edinburgh staff played key roles in bringing these technologies into widespread use in both the private and public sectors through participation in standards development work.

26. Ultrax speech therapy
The use of ultrasound imaging for the treatment of persistent speech sound disorders is a rapidly growing area of research. However, it remains difficult for clinicians to interpret raw ultrasound images during the therapy process. Moreover, the diagnostic power of ultrasound has been underexploited, despite technical advances which allow recording of synchronous ultrasound and audio at fast frame rates. Current work at UoE, in collaboration with 3 NHS Trusts and SME Articulate Instruments, is developing and trialling an approach which transforms ultrasound into a diagnostic and visual biofeedback tool for speech therapy. There is a reasonable possibility that this technology will be in clinical use by 2020.

27. Intelligent warehousing – one to watch
In March 2016, Edinburgh Centre for Robotics signed a contract with Hitachi's Central Research Labs (CRL-CER) in Tokyo, with support from UKTI and the Foreign and Commonwealth Office, to collaborate in the domain of intelligent warehousing and autonomous manipulation robots. It is estimated that by 2025 such advanced robotic and autonomous systems (RAS) could have a worldwide economic impact of $1.7 trillion to $4.5 trillion annually, with an emerging market value €15.5 billion.

28. Wikisimple – one to watch
Software that simplifies the text in a given document/webpage and simplifies it to match the level of the reader. Using machine learning, the software can shorten the sentences or shorten the content. The software has successfully been used to simplify Wikipedia articles. The software is currently being licensed to a Czech company - e-artnow s.r.o.

29. Object detection – annotating images in large image banks
Google is currently using a technique we have developed in Edinburgh to efficiently annotate object locations on the very large OpenImages dataset (9M images). The plan is to publicly release (at least part) of this sometime in 2017. It already now has impact on Google, as the data produced by this method contributes to training the model that then runs on Google Photos. When the data will be released to the public, it is expected to become the 'next ImageNet', i.e. the standard large-scale dataset on which people develop and measure object detection algorithms.

30. Bonseyes: improving efficiency and security in low power Internet of Things devices – one to watch
Bonseyes is a a three-year, €7.4m project which aims to develop a platform consisting of a data marketplace, deep learning toolbox, and developer reference platforms for organisations wanting to adopt artificial intelligence (AI) in low power Internet of Things (IoT) devices (edge computing), embedded computing systems, or data centre servers (cloud computing'). It is expected to deliver significant improvements in efficiency, performance, reliability, security, and productivity in the design and programming of systems of artificial intelligence that incorporate smart cyber physical systems. Partners include emotion-analytics company nViso, which will co-ordinate the work.

31. Formal Semantics of Graph Query Languages – one to watch
We are working with Neo Technology, a large graph database vendor (based in London and Malmo) on formalizing their query language Cypher. The language currently has 2 independent implementations, by Neo and by SAP, and Orace has a very closely related language as part of their graph product, but none has formal semantics. Neo is leading the effort to standardize graph query languages (much like W3C did for XML; the organization doing it for graph data is called LDBC), and they want to use Cypher as the basis of the future standard. Their model is already based on our work with Domagoj Vrgoc (PhD from here in 2014), and they are now supporting a postdoc working on the definition of the language, with a grant worth about £55K.