# \\scotland.gov.uk\dc1\FS5_Home\U206737\SSAC Drafts\SSAC Logo.JPG

# Big Data and Data Science in Scotland: An SSAC Discussion Document

Lead Author: Jon Oberlander, j.oberlander@ed.ac.uk

20 January 2014

# Summary

Big data and data science are two emerging areas that have developed quickly are having an increasing impact in Scotland.

Big data is characterised by its increasing volume, velocity, and variety. Data science is the emerging area of extracting knowledge from big data, based on establishing the principles of the underlying algorithms, statistics, methods, software, and systems.

The purpose of this paper is to provide an overview of the Scottish context, activities and initiatives in big data and data science, and to indicate questions and issues for discussion.

The paper consists of four sections. First, some context for big data and data science is given, including the relationship between big data, open data and open government. Second, an overview of recent Scottish government policy developments in big data and data science is given. Third there is an overview of the major initiatives and activities in Scotland. These are focussed on four areas: scientific research and research infrastructure, health and medical research, public sector information, and innovation centres and training. Finally, a number of questions and issues for discussion are raised, for example concerning overlaps and gaps, and relations to UK, EU and international initiatives and activities. Further details of all the activities are given in the Appendices.

*Caveat*

The Scottish landscape is changing rapidly, this document outlines only those initiatives that have been announced or approved to date.

# 1. Setting the Scene

## Big Data and Data Science

Commerce, government, academia and society now daily produce vast volumes of data, too fast and too complex to be understood by people without the help of powerful informatics tools. A 2001 report by Doug Laney drew attention to the growing volume, velocity, and variety of “big data”. Both commerce and academia struggle to deal with these already, and new categories of commercial products (like the Internet of Things) and scientific projects (like the Square Kilometre Array) promise to add to the burden. Genomics, personalised medicine, smart meters, e-commerce, mobile applications and cultoromics: even small teams can now generate big data, by interacting with millions of users. Much of this data–85%, according to TechAmerica–does not occur in a standard relational form but as unstructured data: images, text, video, recorded speech, and so on.

So, how can we convert big, complex data into human-usable knowledge? We need more than faster computers with bigger memories. We have to draw together ideas from machine learning, statistics, algorithms, and databases, and test them safely and at scale on streams of messy data. Data science is the emerging area that focuses on the principles underlying methods, software, and systems for extracting actionable knowledge from data. McKinsey Global Institute predicts a shortage of up to 190,000 data scientists in the US. Support for data science matters, because of the predicted skills gap, and because data science increasingly supports diverse sectors, including:

* Healthcare: translational biomedicine, information fusion for personalized medicine
* Digital commerce: algorithmic marketing and personalization from big data
* Science: image analysis in astronomy and neuroinformatics; systems and synthetic biology
* Energy: increasing end-to-end system efficiency through analytics, feedback and control
* Computational social science: Studying social networks using rich data sources
* Sensors: making inferences from streaming data from heterogenous sensor networks
* Archives and metadata: Searching and structuring massive multimodal archives
* Open data: enabling citizen engagement, smart cities, and government transparency

## Big Data versus Open Data versus Open Government

The following diagram from Open Data Now paints a reasonably clear picture of the relationships between these three different ideas:

opendatanow.tiff

http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/

Note: ESG = environmental, social, governance. SEC = (US) Securities and Exchange Commission.

From this, it should for instance be clear that: not all big data is open; not all open data is big; and only a subset of big, open data is relevant to open government. One definition of open data (offered by the UK’s Open Data Institute) is: “Open data is information that is available for anyone to use, for any purpose, at no cost. Open data has to have a licence that says it is open data. Without a licence, the data can’t be reused. The licence might also say: (i) that people who use the data must credit whoever is publishing it (this is called attribution); and (ii) that people who mix the data with other data have to also release the results as open data (this is called share-alike)”.

## How Big is Big?

Scientific researchers deal with vast amounts of data; but it is important to see this in perspective. In 2012, the Large Hadron Collider (LHC) was generating around 20 petabytes (PB) of data for further analysis per annum; however, by 2009, Google was already dealing with at least 20PB per *day,* and so by 2012, significantly more. On the one hand, LHC researchers were automatically discarding the vast majority of data theoretically collectable; had they captured it, their throughput would have been about 300 times Google’s. On the other hand, much (but certainly not all) scientific data is highly structured, but Google and its competitors have always dealt with unstructured data, so the commercial sector has been engaging in a particularly challenging task.

Looking forward, when it comes online, the volume and velocity of data generated by the Square Kilometre Array will dwarf that produced by previous scientific activities. But at the same time, the commercially deployed Internet of Things will also be delivering much more (and more varied) data than is available now, generated by networked sensors and actuators, distributed en masse throughout the natural and built environment. We are moving beyond exabytes to zettabytes.

# 2. Immediate Context

The Digital Directorate of the Scottish Government is lead by Mike Neilson. Its Digital Strategy has four priorities: connectivity, digital public services, digital economy, and participation. Of these, public services and the economy are most relevant to big data. The first focuses on transforming “public services to ensure they can be provided online whenever possible and are shaped around peoples’ needs”; the second on encouraging “a vibrant and thriving digital economy where our research base and indigenous companies are recognised internationally and are supported and encouraged to grow.” Two bodies are relevant: the Data Management Board, and the Data Linkage Framework Board; there is also supporting documentation in “Scotland’s Digital Future”.

*Data Management Board*

“The [Scotland’s Digital Future: Delivery of Public Services](http://www.scotland.gov.uk/Publications/2012/09/6272) strategy sets a number of objectives in relation to effective use and management of public sector data, both to improve service delivery and to promote economic growth. To ensure a cohesive approach is taken across Scotland, a Data Management Board has been established. The Data Management Board met for the first time on 7 June 2013.  The group will provide strategic direction and overview across the various data workstreams: [linkage](http://www.scotland.gov.uk/Topics/Statistics/datalinkageframework), innovation, spatial information and sharing.” The group is chaired by Paul Gray, Director General Health and Social Care and Chief Executive of NHS Scotland (Scottish Government), members include the Chief Scientist for Health and the Chief Scientific Adviser.

*Data Linkage Framework Board*

Until a planned Data Linkage Centre and associated Privacy Advisory Committee are in place, this directs “delivery of the Data Linkage Framework which … aims to: (1) build on existing successful programmes collaboratively to create a culture where legal, ethical, and secure data-linkage is accepted and expected; (2) minimise the risks to privacy and enhance transparency, by driving up standards in data sharing and linkage procedures; (3) encourage and facilitate full realisation of the benefits that can be achieved through data-linkage to maximise the value of administrative and survey data.” Chaired by Andrew Morris, Chief Scientist for Health (Scottish Government).

*Scotland's Digital Future - Supporting the Transition to a World-leading Digital Economy – Emerging Findings* Published April 2013 (See Appendix A to the current document)

<http://www.scotland.gov.uk/Publications/2013/05/2347>

“Through the Technology Advisory Board (TAG), Scottish Enterprise and its partners have identified a series of near and medium-term market opportunities for Scotland in the technology and engineering sector. These are:

* *Digital Health & Care* – the ability to improve patients’ and clients’ care and outcomes (and providers’ productivity) through the use of appropriately structured service delivery, integrated ICT data systems and digital devices.
* *Big Data* – deriving value from the huge amounts of unstructured data. Big data opportunities exist in most sectors, but in particular energy, retail, financial services, life sciences, engineering, manufacturing and the public sector.
* *Smart Mobility* – the ability to access applications, or reach customers on the move.
* *Smart Sensor and Sensor Systems* – the combination of a sensing element with processing capabilities (embedded intelligence) provided by a microprocessor.
* *Smart Built Environment* – the ability to improve public and private civic services through the use of appropriately structured service delivery, integrated ICT data systems and digital devices. “ (p15).

# 3. Scottish Landscape

We identify four main areas of activity in Scotland already engaging with big data:

* scientific research and research infrastructure
* health and medical research
* public sector information
* innovation centres and training

A more detailed view of this landscape is given in Appendix B.

*Scientific Research and Research Infrastructure*

Physicists (especially in particle physics and astrophysics) have long been involved in data intensive work, as evidenced by their close involvement in past activities such as the National E-Science Centre (based in Edinburgh and Glasgow). They will continue to require advanced compute facilities. SUPA can provide an overview of data intensive work in Scotland, but two areas of activity are worth noting. In Glasgow, gravity wave research, and the Max Planck International Partnership focussing on quantum phenomena involve big scientific data. In Edinburgh, EPCC hosts the UK’s advanced compute facility, key HPC support for big science; HECTOR users are now being introduced to ARCHER. The recently-announced Higgs Centre for Innovation focuses on space science and big data. As well as physics, other areas of science are increasingly data intensive. Life sciences (in bioinformatics from genomics, to increasingly, “phenomics”), geosciences (including environmental monitoring), agriculture, social sciences, statistics, and of course, informatics itself. Relevant institutes in Scotland include BioSS, James Hutton Institute, and the Roslin Institute. A number of JISC units provide further supporting infrastructure, as does the CREATe centre.

Contacts: Dave Britton, John Chapman, Rod Murray-Smith (Glasgow); Richard Kenway, Arthur Trew, Lesley Yellowlees, David Robertson (Edinburgh).

Further Contacts: David Elston (Director, BioSS), Iain Gordon (CEO, Hutton), David Hume (Director, Roslin), Julie Simpson (CAMERAS Programme coordinator, Scottish Government), Peter Burnhill (Director of EDINA and Head of Edinburgh University Data Library), Kevin Ashley (Director of DCC, Edinburgh), Ralph Weedon (Director of JISC Legal, Strathclyde), Martin Kretschmer (Director of Create, Glasgow).

*Health and Medical Research*

There is significant focussed activity in this area, via the Health Directorate of Scottish Government, working with NHS Scotland. The UK’s four eHealth Informatics Research Centres include one lead by Dundee, which laid the foundations for the Farr Institute (lead by Dundee, with a significant presence in Bioquarter Edinburgh). The Farr hosts a ‘data safe haven’, to facilitate medical research depending on controlled access to patient records. A related activity involving research based on electronic patient records, is the Scottish Health Informatics Partnership (SHIP). Public confidence in the way its data is treated is a critical consideration, in the health sector, and beyond.

Contacts: Andrew Morris (Chief Scientist, Scotland), John Savill (CEO MRC)

*Public Sector Information*

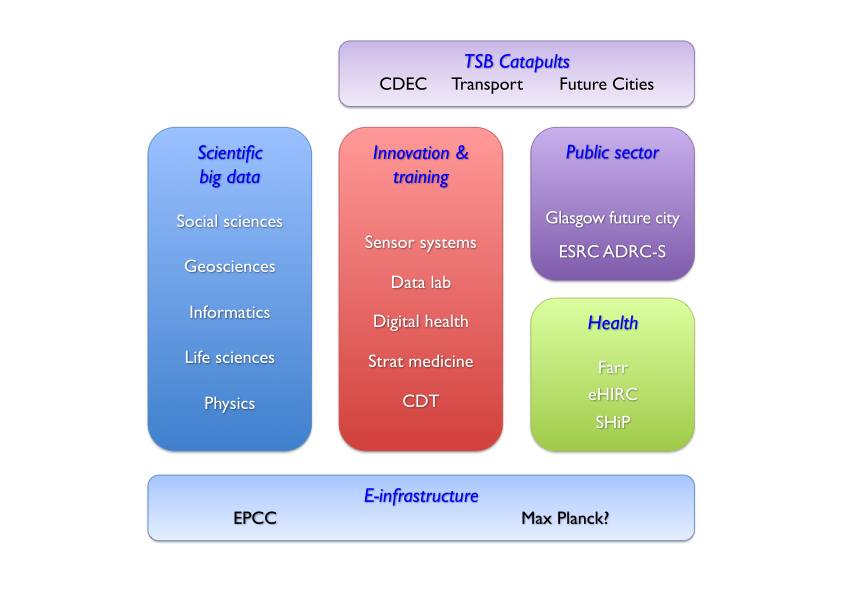
ESRC recently announced four Administrative Data Research Centres. One of these is in Scotland (based in Edinburgh), and there are connections to Scotland’s Data Management Board. Phase 2 of the Big Data Network was recently announced, which comprises Business and Local Government Data Research Centres that will deliver the infrastructure to support access to business or local government data or both, for the mutual benefit of researchers and data owners. Glasgow University will lead the one of the centres, the Urban Big Data Research Centre (UBDRC), Furthermore, activity around the TSB Future Cities Demonstrator in Glasgow includes the development of a Big Data Store; this and related platforms may find wider benefits via the Scottish Cities Alliance. Given that much future cities work involves mobility and transport, there are connections to three of TSB’s Catapult investments outside Scotland: Future Cities, Transport Systems, and Connected Digital Economy. A key mover in UK public sector information is the Open Data Institute (ODI), but there is currently no Scottish node at either city or regional level. There are also some questions as to how Scottish agencies relate to the UK’s National Information Infrastructure, and to UK Open Government Partnership commitments, such as revisions to the Local Authorities Data Transparency Code requiring local authorities to publish key information and data.

Contacts: Chris Dibben (Geosciences), Vonu Thakuriah (Urban Studies, Glasgow University), Peter Triantifillou (Computing Science, Glasgow University) Maria Sigala, Paul Meller ESRC, Scott Sherwood (Glasgow Future Cities Demonstrator), Andrew Unsworth (Smart Cities & Communities Programme Manager, Scottish Government); Jackie McAllister (Digital, Scottish Government), Nigel Shadbolt (Chair, ODI), Paul Gray (Chair, DMB, and Health and Social Care and Chief Executive of NHS Scotland), Mike Neilson (Director of Digital, Scottish Government)

*Innovation Centres and Training*

Through the Scottish Funding Council and Scottish Enterprise and Highlands and Islands Enterprise, Scottish Government are establishing a series of Innovation Centres. Of these, four are related more or less closely to Big Data: DHI - Digital Health Institute (based in Edinburgh and Glasgow); SMS-IC - Stratified Medicine (based in Glasgow); CENSIS – Sensors and Imaging Systems (based in Glasgow); and DataLab – Data Science (based in Edinburgh, Glasgow, and RGU). It is notable that the DHI is expected to apply insights from big data-based health and medical research, and this applications focus should complement the research orientation of bodies such as the Farr Institute. Regarding training in data science skills, the Data Lab is likely to expedite new masters level training, while EPSRC have just announced a new Centre for Doctoral Training (CDT) in Data Science (based in Edinburgh), which includes partners such as Amazon, Google, IBM and Oracle; the CDT is one of just two serving the UK’s growing needs for data scientists qualified to PhD level.

Contacts: Dave Clark (Interim CEO, CENSIS), Anna Dominiczak (Glasgow VP, SMS-IC), Justene Ewing (CEO, DHI), Neil Logan (Lockheed Martin/DataLab), Chris Williams (Director, CDT in Data Science).



# 4. Discussion Points

It will be apparent from the map that Scotland has a number of existing and new big data activities, which can be somewhat roughly labelled: science, medicine, public sector, and innovation. There is an opportunity to draw the threads together, both to reduce the risk of fragmentation, and to capitalise on the scale of existing activity. Some questions will need to be addressed.

1. Question: Have any significant activities been missed?
2. Question: Where would they fit on this map?

It will be clear (especially in the light of Appendix A) that there are points of overlap between some of the activities – for instance, between the Farr Institute and the Digital Health Institute. But equally, support for big data in construction (see Appendix B) or agriculture is not currently prominent.

1. Question: How can overlaps be developed into strengths rather than weaknesses?
2. Question: Are there obvious gaps between Scotland’s big data offerings?

UK policy-making has been developing rapidly (see Appendix C), and in particular has developed an information economy strategy. There are also significant opportunities for big data-oriented research in the EU’s Horizon 2020 programme (see Appendix D).

1. Question: What specific strengths can Scotland draw on in contributing to UK activity?
2. Question: What specific strengths can Scotland draw on in contributing to EU activity?

UK policy-making has also framed requirements for a national information infrastructure and actions to open government data, in the context of the information economy strategy. Scotland’s Digital Directorate is currently forming an Open Data Policy for the public sector.

1. Question: How do Scottish bodies relate to UK NII activities?
2. Question: How do Scottish bodies relate to UK open government activities?

The Open Data Institute (see Appendix C) expedites the opening up of datasets via data.gov.uk, and has recently announced a series of open data nodes (cities and regions within and beyond the UK)

1. Question: Is there value in having an ODI node in Scotland?
2. Question: Could data.gov.uk be used for opened Scottish public sector data?

At every level – Scotland, UK, EU, USA – the skills gap is identified as a threat. At a more general level, big data and data science will only succeed if the public can trust the relevant bodies.

1. Question: Is the new EPSRC centre enough, at the doctoral level?
2. Question: How do we ensure that the public’s privacy and confidentiality is secured?

# Appendix A: Details on Activities

## A.1 Scientific Research and Research Infrastructure

www.epcc.ed.ac.uk

[www.supa.ac.uk](http://www.supa.ac.uk)

www.stfc.ac.uk

www.sicsa.ac.uk

www.bioss.ac.uk

www.hutton.ac.uk

www.roslin.ed.ac.uk

www.camerascotland.org

As noted, Scotland already hosts major facilities for UK High Performance Computing (HPC) and associated storage at EPCC, and there is significant engagement in scientific computing in Glasgow and beyond. Scotland’s physicists involved in a number of large-scale projects, while fields such as life sciences and environmental sciences make increasing use of big data techniques.

EPCC, the Edinburgh Parallel Computing Centre, “hosts and administers a number of national-level facilities: [HECToR](http://www.epcc.ed.ac.uk/facilities/hector) - UK national supercomputing facility provided by RCUK; [DiRAC](http://www.epcc.ed.ac.uk/facilities/dirac) - UK national supercomputing facility provided by STFC; UK Research Data Facility … provided by RCUK.” EPCC also hosts Supercomputing Scotland, promoting the use of HPC in business. Physicists in the Scottish Universities Physics Alliance (SUPA) are major users of HPC e-Infrastructure; furthermore, SUPA’s CEO Jim Hough also leads an International Max Planck Partnership , which links five SUPA departments with five Max Planck Institutes in Germany, focussing on observation and measurement at the quantum limit. The Higgs Centre for Innovation was announced on 5 December 2013; STFC quote David Willetts saying it will focus on “big data and space technologies, two of the Eight Great Technologies of the future.” Opening in 2016, with STFC capital investment of £10.7M, the Centre will be sited at the UK Astronomy Technology Centre in Edinburgh, supporting entrepreneurship through access to specialist instruments and big data capabilities.

As noted earlier, current data science demands techniques from algorithms, databases, and

machine learning. Dealing with unstructured and semistructured data demands deep expertise in capturing, organising, extracting and communicating information from streams of text, speech and images; Scotland is therefore fortunate to host world-class groups in (for example) information retrieval, natural language understanding, and many-core processing. Furthermore, many machine learning methods are rooted in traditional statistical techniques, and so Scotland’s strength in computing is supplemented by further academic expertise in statistics and applied mathematics.

Contacts: Dave Britton, John Chapman, Rod Murray-Smith (Glasgow); Richard Kenway, Arthur Trew, Lesley Yellowlees, David Robertson (Edinburgh).

As an example on the life sciences side, Edinburgh Genomics is distinctive in being funded by NERC, BBSRC, and MRC. It is part of the “NERC Biomolecular Analysis Facility, a collaboration of nodes across the UK that … [provide] advanced technologies in sequencing, genotyping, metabolomics and bioinformatics”. It also has core funding “as part of the BBSRC commitment to National Capability in genomics, delivered through the Roslin Institute”, and it is “one of MRC’s two currently funded High Throughput Sequencing Hubs”, with particular experience in human genomics and transcriptomics.

Turning to environmental and related sciences, BioSS focuses on mathematics and statistics as applied to agriculture, the environment, food and health. BioSS is “one of the Main Research Providers (MRPs) for strategic research in environmental, agricultural and biological science funded by the Scottish Government’s Rural and Environment Science and Analytical Services Division (RESAS). We have a distributed staff structure … with headquarters on the King’s Buildings science campus of the University of Edinburgh and offices in Edinburgh, Aberdeen, Dundee and Ayr.” BioSS is formally part of the James Hutton Institute, and has five scientific themes: (i) Cell and Molecular Sciences; (ii) Ecological Sciences; (iii) Environmental and Biochemical Sciences; (iv) Information and Computational Sciences; and (v) Social, Economic and Geographical Sciences. The Roslin Institute “undertakes research within the framework of BBSRC Institute Strategic Programmes focussed on the health and welfare of animals, and applications of basic animal sciences in human and veterinary medicine, the livestock industry and food security.”

The Coordinated Agenda for Marine, Environment and Rural Affairs Science (CAMERAS) “is a partnership initiative between: Scottish Government – Marine Scotland Science, Rural and Environment Science and Analytical Services (RESAS), and Science Advice for Scottish Agriculture (SASA); Scottish Environment Protection Agency; Scottish Natural Heritage; Forestry Commission Scotland; Food Standards Agency Scotland; Quality Meat Scotland; and Scottish Water. Its purpose is to align and coordinate the scientific activity of the partner organisations to ensure best use of existing resource and enhanced support to Scottish Government policy development and delivery”.

Contacts: Mark Blaxter (Director, Edinburgh Genomics), David Elston (Director, BioSS), Iain Gordon (CEO, Hutton), David Hume (Director, Roslin), Julie Simpson (CAMERAS, Scottish Government).

*Other Academic Units/Services relevant to Big Data and/or Digital Economy more broadly*

*Joint Informatics Systems Committee*

www.edina.ac.uk

www.dcc.ac.uk

www.jisclegal.ac.uk

JISC has three UK-level units in Scotland: EDINA; DCC, the Digital Curation Centre; and JISC Legal. EDINA provides access to an online library of information resources to enhance the work of UK academics. The DCC (Edinburgh, Glasgow and Bath) provides expert advice and resources to those who store, manage, protect and share digital research data in UK HEIs. Jisc Legal raises awareness of, and provides guidance on, legal issues surrounding the use of technology in academia.

Contacts: Peter Burnhill (Director of EDINA and Head of Edinburgh University Data Library), Kevin Ashley (Director of DCC, Edinburgh), Ralph Weedon (Director of JISC Legal, Strathclyde).

*Creativity, Regulation, Enterprise and Technology*

http://www.create.ac.uk/

CREATe (based in Glasgow) is the RCUK centre for copyright and new business models in the creative economy. With an ambitious programme of 40 projects delivered by an interdisciplinary team of academics (law, economics, management, computer science, sociology, psychology, ethnography and critical studies), CREATe is a pioneering academic initiative designed to help the UK cultural and creative industries thrive within the global digital economy.

Contact: Martin Kretschmer (Director of Create, Glasgow).

## A.2 Health and Medical Research

*1. E-Health Informatics Research Centre*

http://www.mrc.ac.uk/Newspublications/News/MRC008799%20

A consortium of 10 UK government and charity funders, led by the Medical Research Council (MRC), has made a historic £19 million investment to establish four e-health research Centres of Excellence in London, Manchester, Dundee and Swansea. … The Centres will open in late 2012 and will harness the wealth of UK electronic health records to improve patient care and public health.

The four Centres will investigate a wide range of conditions that place a huge burden on the UK population, including diabetes and obesity, cardiovascular disease, cancer and child and maternal health. Maximising the unique value of the NHS, the Centres will undertake cutting edge research that links e-health records with other forms of research and routinely collected data, which will lead to patient and public benefit and ensuring the UK remains at the forefront of global medical research. … The four Centres will make use of patient data sets available through the Clinical Practice Research Datalink, a £60 million service recently announced by the Medicines and Healthcare Products Regulatory Agency and the National Institute for Health Research. The public and charitable funding for these Centres builds on this important commitment from the Government and on similar bodies that link patient records in Scotland and Wales.

Professor Andrew Morris, Dean of the School of Medicine at the University of Dundee and Chief Scientist at the Scottish Government Health Department said “Colleagues in Scotland are thrilled to be awarded Centre of Excellence status. This builds upon over 40 years’ experience of using electronic patient records not only to drive improvements in the quality of health care in Scotland, but also to innovate in the way we deliver clinical trials and discover the best treatment options for patients and communities. The spirit of collaboration between NHS Scotland and the Universities of Aberdeen, Dundee, Edinburgh, Glasgow, St Andrews and Strathclyde has been tremendous. There is a great opportunity to make the United Kingdom the destination of choice of eHealth research, and in doing so help deliver the best quality health care to the people of Scotland."

The members of the E-Health Research Initiative who have jointly-funded the four Centres are: Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Chief Scientist Office (Scottish Government Health Directorates), the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the National Institute for Health Research, the National Institute for Social Care and Health Research (Welsh Government) and the Wellcome Trust.

Contact: Andrew Morris (Chief Scientist for Health, Scottish Government)

2. *The Farr Institute*

http://www.mrc.ac.uk/Newspublications/News/MRC009207

http://www.dundee.ac.uk/pressreleases/2013/july13/institute.htm

In July, it was announced that the Medical Research Council (MRC) will invest £20m capital funding in the establishment of a UK health informatics research institute, to be known as the Farr Institute. This investment will support the safe use of patient and research data for medical research across all diseases. The Institute’s independent research will support innovation in the public sector and industry leading to advances in preventative medicine, improvements in NHS care and better development of commercial drugs and diagnostics. It will also provide new insights into the understanding of causes of ill health which in turn will guide new biomedical research discovery. In addition to health benefits for patients and UK citizens, the Institute will help to cement the UK’s reputation as a world leader in research using large electronic health data.

The Farr Institute will have major centres in London, Dundee, Manchester and Swansea and will link research in 19 universities across the UK. It builds on the four e-health informatics research centres (eHIRCs) recently funded by a consortium of three Research Councils, three health departments and four leading medical research charities. …

Scotland is represented in the new Farr Institute by the Universities of Dundee, Edinburgh, Aberdeen, Glasgow, St Andrews and Strathclyde and NHS Scotland. Their combined expertise will be coordinated from Dundee to support the safe use of patient and research data for medical research across all diseases. … High-quality space at Dundee’s School of Medicine will be refurbished to create an inter-disciplinary health informatics research environment, while new facilities will also be established at Edinburgh’s BioQuarter. Together these locations will act as the hub of the Scottish Institute. Additional networking and compute capability will be provisioned at the Universities of Aberdeen and Glasgow, working closely with their NHS partners. …

A total of £39 million has been invested in the Farr Institute. The concentration of funding in developing UK health informatics research base will provide a focus for collaborations with IT and pharmaceutical companies, attracting inward investment into the UK economy.

Contact: Andrew Morris (Chief Scientist for Health, Scottish Government)

*3. ScottisH Informatics Programme*

http://www.scot-ship.ac.uk

SHIP is an ambitious, Scotland-wide research platform for the collation, management, dissemination and analysis of Electronic Patient Records (EPRs). The programme brings together the Universities of Dundee, Edinburgh, Glasgow and St Andrews with the Information Services Division (ISD) of NHS Scotland.

SHIP is funded by the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council and aims to:

* Provide access to an exciting new national research facility, firmly embedded within and supported by NHS Scotland, providing the basis for numerous future studies using EPRs.
* Create a research portal for EPRs already held by NHS Scotland that will provide rapid, secure, access to the type of data that clinical scientists require.
* Develop and evaluate systems that work across institutional boundaries to allow linkage between large, federated, third party research datasets and the NHS research portal.

Contact: Andrew Morris (Chief Scientist for Health, Scottish Government, and Chair of SHIP Scientific Management Group)

Finally, it is worth noting that medical imaging–such as neuroimaging–routinely generates significant amounts of unstructured image data, which may not yet count as big data, but which demands similar data science analytic skills as other branches of science and medicine.

## A.3 Public Sector Information

*1. Administrative Data Research Centre – Scotland*

http://www.esrc.ac.uk/news-and-events/press-releases/28673/The\_Big\_Data\_Family\_is\_born\_\_David\_Willetts\_MP\_announces\_the\_ESRC\_Big\_Data\_Network.aspx

Four new innovative administrative data research centres and a data service will strengthen the UK's competitive advantage in Big Data. The centres and service together form the Administrative Data Research Network (ADRN) which will enable research based on linked data between government departments and be overseen by a single governance structure.

At the core of the ESRC's Big Data Network are the centres to be led by the Universities of Southampton and Edinburgh, Swansea University and Queens University Belfast, with the administrative data service to be based at the University of Essex. Collectively, they will benefit from a grants package totalling approximately £34 million. The centres will make routinely collected administrative data accessible for research in ways that prevent the identification of individuals, while providing a sound evidence base to inform research, and policy development, implementation and evaluation.

The ADRN has been informed by the work of the Administrative Data Taskforce, which was formed in December 2011 by the Economic and Social Research Council (ESRC), the Medical Research Council (MRC) and Wellcome Trust, and chaired by Sir Alan Langlands. The ESRC's Big Data Network has been divided into three phases. In Phase 1 of the Big Data Network the ESRC has invested in the development of the Administrative Data Research Network. Phase 2 will focus primarily on business data and local government data and Phase 3 will focus on Third Sector data and social media data.

Contact: Chris Dibben (Director of ADRCS, Geosciences Edinburgh)

*2. Big Data Network Phase 2: Business and Local Government Data Research Centres*

<http://www.esrc.ac.uk/funding-and-guidance/funding-opportunities/27813/business-and-local-government-data-research-centreshomepage-promo.aspx>

Phase 2 of the ESRC Big Data Network centres will deliver the infrastructure to support access to business or local government data or both, for the mutual benefit of researchers and data owners. All the commissioned Centres will be expected to provide safe and secure access to a range of extremely rich datasets.

The University of Glasgow was awarded a £7 million Phase 2 centre in January 2014, the Urban Big Data Research Centre (UBDRC), which aims to establish a world leading facility to create an multi-sectoral urban linked data resource from local government authorities and business owners in Glasgow; provide outstanding training and research support services to ensure wide exploitation of the data; and deliver a strategic approach to knowledge transfer and training to build capacity and engage with policy, business, and the wider public. The UBDRC will provide a unique facility for researching cross-cutting urban issues and complex urban challenges by enabling access to multi-sectoral linked data from local government, business and other sources.

UBDRC will focus on bringing together myriad of datasets, from multiple urban sectors, to create a linked urban data resource that allows comprehensive and crosssectoral research. The centre will provide data curation services and the necessary metadata and provide a range of data  access services to users, including, where necessary, secure access to confidential data UBDRC will develop, test and evaluate a wide range of methodological approaches including urban and regional modelling, agent-based models, machine learning and other methods and will support research leading to new cross-cutting theoretical insights, hypotheses and understanding of urban systems, thereby stimulating foundational research on new models of urban behaviour, processes and service provision. The data resource will be used to develop spatially-indexed(and perhaps temporally-indexed) urban indicators on myriad aspects describing the quality and character of urban spaces, and the spatial distribution of the urban processes, e.g, on environmental risks, mobility and accessibility patterns, housing and educational aspects, and other aspects that describe the socio-demographic, economic, environmental, built environment, physical and other aspects of urban areas. The data will further allow policy research on a wide range of urban sectors and the derivation of a multitude of approaches for urban governance and business development. Research projects on substantive urban issues such as transport, housing, migration and education will demonstrate to data owners and policy makers the value of large-scale, cross-sectoral data linkage and lead to policy insights for public, private and non-profit decision-makers.

Contact: Vonu Thakuriah (Director, Glasgow University), Peter Triantifillou (Computing Science, Glasgow University) Maria Sigala, Paul Meller ESRC

*3. Future Cities Demonstrator Glasgow*

http://futurecity.glasgow.gov.uk

A £24M UK-level investment by TSB. This is “a collaboration between public and private sector agencies providing a range of services to the city. They include Glasgow City Council, Police Scotland, housing providers, NHS Greater Glasgow & Clyde, universities, energy providers and Scottish Enterprise”. Big data is an important part of the picture: according to the UK Information Economy Strategy: “The programme encompasses several projects:

* The creation of an Integrated Operations Centre managing public space CCTV network and TRAFFCOM roads management systems
* Sustainable Glasgow – addressing issues such as energy conservation and generation, green technology air pollution and the integration of transport routes
* The creation of a Big Data Store collecting and analysing information from previously unconnected databases to influence services and make them more accessible
* City Dashboard giving real time information via smartphone apps on subjects like traffic flow, rail and bus services, weather, accident and emergency waiting times.”

Contact: Scott Sherwood (Open Data Platform lead); Andrew Unsworth (Smart Cities & Communities Programme Manager, Scottish Government); Jackie McAllister (Digital, Scottish Government)

## A.4 Innovation Centres and Training

*1. Data Lab: Data Science Innovation Centre*

An innovation centre with £20M of investment from Government, Industry and Academia over 5 years, aimed at generating jobs and GVA to the economy. Data science presents opportunities for business efficiency, business innovation and business creation. The Centre for Economics and Business Research (CEBR) estimates that the Big Data marketplace could create 58,000 net new jobs within the UK alone and the benefits to the economy are estimated at £216bn by 2017. The focus of the Data Lab’s work will be on key industry sectors with high growth potential and the capacity to boost productivity:

* *Financial Services* (Potential Value to UK ~ £16.3bn) There is potentially huge value in the data attached to the billions of transactions processed by this sector. Data science can unlock this value in areas such as fraud detection, risk modelling, pricing optimization and data quality and integrity.
* *Online & Digital* (Potential Value to UK > £30bn) household names like Amazon, eBay, Facebook and Google apply sophisticated analytics to customer usage and transaction histories to help these businesses better understand their customers, leading to improved customer acquisition and retention. However there are many other, less well- known companies generating revenue from continued innovation and exploitation of data through online and digital services.
* *Energy and Utilities* (Potential Value to UK ~ £5.5bn) Scotland's oil and gas strategy is focused on achieving higher long-term recovery rates and greater exports. To date the industry has concentrated mainly on volume, with the need to manage large volumes of sensor and production data for their control systems and compliance. One new field is estimated to produce 250,000 data events every 15 seconds. The industry is now exploring how it can use analytics to extract maximum value from this data.
* *Public Sector & Healthcare* (Potential Value to UK ~ £20.4bn and ~ £14.4bn, respectively) Delivery of public services will be transformed by leveraging both public and private data sources. The opening up of existing government data sources will facilitate the creation of new products and services. Combining the knowledge contained in genomic and proteomic analysis with data from healthcare records or patient monitoring will deliver new ways of preventing, managing and treating illness and disease.

Hubs will be located at Edinburgh University, Glasgow University and RGU.

Private sector partners include: Amazon Development Centre, Amor Group, Aridhia, Avaloq, Baker Hughes, Barrhead Travel, BP, Brightsolid, CGI, Clear Returns, Energistics, Mallzee, Nexen, OutPlay Entertainment, Petrotechnics, RBS, RockStar North, Skyscanner, Simul8, Standard Life, Sumerian, and Wood Group. Public sector partners include: Scottish Government, City of Edinburgh Council, City of Glasgow Council, and NHS Scotland.

Contact: Neil Logan (Lockheed Martin, SE TAG and Data Lab industry lead)

*2. Digital Health Institute*

http://dhi-scotland.com

http://www.sdi.co.uk/news/2013/10/pioneering-healthcare-institute-launched.aspx

Representatives from Samsung Electronics, IBM, Philips, Deutsch Telekom Celesio, Continua Health Alliance and Lockheed Martin recently attended the launch of Scotland’s first Digital Healthcare Institute. Based in Edinburgh, it will be the focus for collaboration between leading health and care operators and technology businesses across Europe, the United States and Asia.

The Institute’s role is to speed up research and development to produce innovative new technologies that will transform the quality of people's lives. All with an eye on the potential of both the local and European-wide markets. The former is estimated to grow to between £0.5 and 1 billion over the next 5 to 10 years, and the Europe-wide market to as much as £70 billion over the same timeline. The Institute and its partners will offer facilities to develop and test new ideas, plenty of opportunities for collaboration, and support to bring products to market.

Academic partners: Edinburgh, Glasgow School of Art.

Contact: Justene Ewing (CEO, DHI), George Crooks (Chair, DHI and Medical Director, NHS 24)

*3. Sensors and Imaging Systems*

http://sensorsystems.org.uk/censis/

http://www.gla.ac.uk/research/news/headline\_275905\_en.html

The University of Glasgow has received funding to create a world-leading sensor and imaging systems centre. The Scottish Funding Council has pledged £10m over the next five years to support the Innovation Centre – Sensor and Imaging Systems (IC-SIS), which will engage in industrially collaborative projects to develop new technologies and form links with industry to bring innovative products to market.

Areas of Interest: Offshore, renewables and energy; Built environment; Defence; Intelligent transport; Smart grid and energy distribution; Environmental and agriculture; Ocean science; Manufacturing and process control; Food and drink; Healthcare; Life sciences, pharmaceuticals and drug discovery.

Industry Partners. IC-SIS has received industrial support from large multinationals including Freescale, Texas Instruments, IBM, SELEX ES, ST Microelectronics, Thales Optronics, BAE Systems, BP, and FMC Technologies. Other confirmed industry partners include Scottish and Southern Energy, and Scottish Water, as well as globally leading companies Optos and Toshiba Medical, and high-technology Scottish SMEs Gas Sensing Solutions Ltd. Academic Partners. … Collaborating on the IC-SIS project with the University of Glasgow are the Universities of Aberdeen, Dundee, Edinburgh, Highlands and Islands, Stirling, Strathclyde, West of Scotland, Glasgow School of Art, Glasgow Caledonian, Heriot Watt, and Robert Gordon Universities.

Contacts: Bob Downes (Chair), Dave Clark (Interim CEO from Nov 2013, late of Thales, Strathclyde)

4. *Stratified Medicine Scotland*

http://www.aridhia.com/our-collaborations/stratified-medicine-scotland-innovation-centre

The Stratified Medicine Scotland Innovation Centre (SMS-IC) brings together experts from academia, industry and the NHS to implement a biomedical informatics service to aid clinical and translational research and enable stratified medicine. The immediate aim is to enable precision targeting of population subsets in order to demonstrate the benefit of stratification in clinical trials, with a series of projects commencing throughout 2013.

With the impact of chronic disease increasing in prevalence globally and the projected costs of delivering care unsustainable, there is a recognised need to move towards integrated models of care and the use of safe, cost-effective treatments tailored to the likelihood of individual response. The SMS-IC is a response to the need for close multidisciplinary collaboration to ensure that innovation extends beyond the medical aspects of stratified medicine to encompass disciplines such as IT, genomics, public health and data science.

Scotland’s significant past investment in electronic health records and translational medicine research, coupled with a vibrant healthcare technology industry, means that the country is well positioned to act as a incubator for innovation in stratified medicine and chronic disease management to inform the delivery of quality healthcare not only in Scotland, but across the world.

The Scottish Funding Council is providing [£8m] funding over five years to back the creation of the £20m SMS-IC at the new South Glasgow Hospitals Campus. The location … will provide unique access to sequenced human genomes combined with clinical data … with the centre fully operational by September 2015. In the interim, an informatics platform based on Aridhia’s Research Analytics service and a modular SmartLabTM Genetic Analysis laboratory … based on the Life Technologies Ion technology will be hosted … in Paisley, Scotland. The SME partners involved in the SMS-IS are Arrayjet, Axis Shield, Biopta, DestiNA Genomics, Fios Genomics and Sistemic Ltd.

Academic partners: Glasgow, with Aberdeen, Dundee, and Edinburgh.

Contact: Anna Dominiczak (Glasgow VP, SMS-IC).

5. *EPSRC Centre for Doctoral Training in Data Science*

http://datascience.inf.ed.ac.uk

21 November 2013: EPSRC funds two CDTs in the general area of data science, one at Nottingham, the other in Edinburgh. “Data science is the study of the computational principles, methods, and systems for extracting knowledge from data. Large data sets are now generated by almost every activity in science, society, and commerce – ranging from molecular biology to social media, from sustainable energy to health care.

Data science asks: How can we efficiently find patterns in these vast streams of data? Many research areas have tackled parts of this problem: machine learning focuses on finding patterns and making predictions from data; ideas from algorithms and databases are required to build systems that scale to big data streams; and separate research areas have grown around different types of unstructured data such as text, images, sensor data, video, and speech. Recently, these distinct disciplines have begun to converge into a single field called data science.

The Centre for Doctoral Training in Data Science will train a new generation of data scientists, comprising 50 PhDs over five intake years, with the technical skills and interdisciplinary awareness necessary to become R&D leaders in this emerging area. The first cohort will start work in September 2014.”

Private sector partners include: Agilent Technologies, AlertMe.Com, Amazon Development Centre, Amor Group, Apple Inc., Brightsolid, Carnego Systems, Cloudsoft Corporation Ltd., Freescale Semiconductor, Google, HSBC Bank, IBM, Logicblox Inc., Microsoft Research, Oracle, Pharmatics, Psymetrix, Quorate Technology, RangeSpan, RBS, Scottish Power, Selex ES, Skyscanner, Time Out Group, UCB Celltech, Xerox Research Centre Europe, and Yahoo! Labs Barcelona. Public sector partners include: BBC, BioSS, CDEC, City of Edinburgh Council, Digital Health Institute, DCC, James Hutton Institute, ODI, Roslin Institute, and SICSA.

Contact: Chris Williams (Director, CDT in Data Science).

# Appendix B: Recommendations from ‘Scotland’s Digital Future’

*Scotland’s Digital Future - Supporting the Transition to a World-leading Digital Economy*

May 2013

http://www.scotland.gov.uk/Publications/2013/05/2347/7

1. Strengthen strategic partnerships: There should be greater strategic co-ordination between our enterprise and business support agencies to promote agile and integrated responses to emerging opportunities and deliver fully integrated and effectively deployed programmes of intervention. This should be ensured by establishing a Digital Excellence Partnership to take forward the recommendations of this review and to create a 'digital buzz' within Scotland. The appointment of a Digital Excellence Champion or co-ordinator would help to spearhead this Partnership.
2. Strengthen the breadth and depth of support offered to Scottish companies of all sizes: A seamless programme of support and advice should be offered that integrates and builds upon the range of current support that is available. This should include the development of a Digital Excellence Programme, a highly intensive and bespoke package of support targeted at growth companies on a one-to-one basis. Serious consideration should also be given to the merits of a 'digital voucher' scheme [for] new digital technologies.
3. Strengthen the capability (supply) sector: The emergence of the Innovation Centres in Scotland will do much to advance the digital economy in Scotland and we should explore the potential for a Big Data Innovation Centre to help realise opportunities in this area.
4. Strengthen our professional ICT skills base: We should agree, publish and implement a skills investment plan in conjunction with key industry partners and stakeholders. As part of this process, Education Scotland and Skills Development Scotland (SDS) should work in collaboration to make recommendations to enhance awareness of, and help our current and future workforce make the most of career opportunities within the digital economy.
5. Strengthen our digital literacy skills: User skills and digital leadership skills are important across the workforce in all sectors of the economy and a broader digital literacy plan would help organisations to realise the potential benefits that can come from digitisation. A Digital Leadership Programme should be developed to target Managing Directors (MDs) and senior staff within our SME base, whilst Education Scotland, in collaboration with SDS, should examine how digital literacy skills are promoted in our schools, benchmark international practice and make recommendations on how to take this agenda forward. The Digital Participation Charter can play a significant role in ensuring partnership working around this agenda.
6. Strengthen knowledge exchange opportunities: Partners should work together to explore the potential for Digital Excellence & Demonstration Centres. These would be a shared resource between industry partners and the public and academic sectors and would demonstrate the practical application of existing and innovative digital technologies.
7. Strengthen ways of working in the public sector to promote commercial opportunities: The Scottish public sector should build upon its existing commitment to open public data, e procurement and the delivery of digital public services and take these agendas forward in a way that will maximize their positive impact on the digital economy. Specific opportunities could come from opening up government system APIs …, the development of an agile procurement framework for digital projects and using the public procurement platform to build the capability of Scottish suppliers to interact electronically.

## Capabilities and initiatives listed in ‘Scotland’s Digital Future’

*Digital Health and Care:* Using ICT and informatics to support and exploiting reform of service delivery in health and care. Includes initiatives in remote monitoring, call-centre technology and interfacing, medical devices.Context/Drivers: Demographics: global economic imperative to treat more patients in the community rather than in hospital.Internal efficiencies for global health services.

* Digital Health Institute. Collaboration between NHS Scotland, University of Edinburgh (UOE) and Glasgow School of Art to provide a platform to support service delivery reform and for exploration, experience-testing and realisation for products and services.
* Dallas in Scotland. Small Business Research Initiative (SBRI)-based initiative to provide digital solutions. Seven companies, NHS24, SE and HIE.
* Digital Health and Care (DHC) Network Integrator. SME networking and intelligence gathering resource
* DHealth. A HIE initiative for a business networking and opportunity matching organisation

*Big Data:* Requirements for increased velocity, volume, veracity and variability of data. Includes data analytics, data industrialisation, cyber-security and data personalisation. Context/ Drivers: Supporting and exploiting opportunities arising from paradigm shift which considers 'data' to be the primary asset, not the software.

* SuperComputing Scotland. Enhance the competitiveness and productivity of Scottish Companies through the use of advanced Modelling and Simulation, by supporting collaboration between Scottish Companies and EPCC: Edinburgh University's world-class supercomputing centre.
* Data Lab Innovation Centre. Proposed Scottish centre for innovation and collaboration around data analytics etc. *Status?*
* Connected Digital Economy Catapult (TSB/UK). Help businesses and researchers to collaborate and to address some of the large and complex challenges facing the UK. It will provide access to testing facilities and demonstrators, and it will co-ordinate and link expertise across the country. (Potential connections through UOE Informatics.)

*Smart Mobility:* Using digital capabilities to support and exploit opportunities arising at the junction of ICT, transport and energy. Context/Drivers: delivering an efficient and sustainable transport system as part of the transition to a low carbon economy.

* Future Cities Demonstrator, Glasgow. TSB funded initiative to provide a flagship project to develop and showcase 'smart city' capabilities
* Grand Challenge. To stimulate a new flow of large-scale smart mobility demonstrator projects to meet societal challenges and accelerate relevant innovation by industry*. Status?*
* Transport Systems Catapult (TSB/UK). West Midlands hub with Scotland's Smart Mobility programme as a virtual spoke. Strong likelihood that Scottish demonstrator projects become a Living Lab for Catapult-led innovations.
* Smart Cities Inter-operation Standards. Scotland has the opportunity to define and demonstrate a unique digital platform (as a public good) which can unlock export markets through increased global usability. *Status?*
* Future Cities Catapult (TSB/UK). Hub expected in London, looking to Glasgow as a exemplar project.

*Sensors and Imaging Systems:* Using ICT capabilities to convert analogue data such as temperature, movement and electromagnetic radiation into forms useful for the digital economy. Context/Drivers: Excellent Scottish capability with good penetration into global markets.

* CENSIS Sensors and Imaging Innovation Centre. Collaboration between Scottish Universities and companies to support innovation around Scottish strengths and global market needs
* Subsea Engineering Programme. Possible programme to exploit global requirements based on Scottish supply-side strengths in rugged engineering and sensors and imaging. *Status?*

*Smart Built Environment:* Includes the use of digital capabilities to better control energy use in our existing and new building stock. Context/Drivers: Support better living and the transition to a low carbon economy.

* Built Environment Innovation Park. The project provides a Scottish innovation pathway via a clearly articulated support system from concept through to market demonstration within Scotland.
* Built Environment Supply Chain Innovation Support (Ending this year). Aims to increase the level of innovation within the construction industry by connecting market opportunities with the relevant company-base and to encourage, stimulate and support innovative products which are aligned with real and current market needs
* Construction Scotland Innovation Centre. Proposed Scottish initiative to create a dynamic innovation support body to facilitate new construction products and processes from inception through to market delivery to benefit Scotland PLC. *Status?*

# Appendix C: UK Policy Context

There have been rapid developments over the last year at the UK level, a number of which form part of the UK Government’s response to the Shakespeare Review of Public Sector Information (May 2013). The review recommended the development of a national data strategy, with clear leadership, and a pragmatic policy to protect citizen privacy while maximising the value of citizen-generated data. It also recommended early and imperfect releases of public sector data, in parallel with open data of the highest quality, and urged a revision to the “trading funds” model currently used by e.g. the Met Office. New bodies especially relevant to big data include the Information Economy Council, the E-infrastructure Leadership Council, and the Open Data Institute. Here, we highlight some policy elements particularly from BIS, and the Cabinet Office: (i) RCUK Strategic Framework for Capital Investment (Nov 2012), (ii) David Willett’s Eight Great Technologies (Jan 2013), (iii) Information Economy Strategy (June 2013), (iv) Open Government Partnership UK National Action Plan 2013 to 2015 (Oct 2013), (v) National Information Infrastructure (Oct 2013), and (vi) Seizing the Data Opportunity: A Strategy for UK Data Capability (Oct 2013).

*Information Economy Council*

http://www.techuk.org/about/information-economy-council

Co-chaired David Willetts and Victor Chavez (Thales/TechUK). Representatives from Government and Industry (Amazon, Cisco, IBM, HP, Samsung, Tata, TalkTalk, etc.), with CDEC and TSB, TechCity, and small number of academic representatives (Imperial, ODI/Southampton).

“The IEC will set the agenda for implementation of and future work against the Information Economy Strategy and monitor progress. The purpose is to provide a vehicle for government and industry to work in partnership to develop and deliver a long-term strategy to support the growth of the Information Economy (IE) in the UK.”

*E-infrastructure Leadership Council*

https://www.gov.uk/government/policy-advisory-groups/e-infrastructure-leadership-council

Co-chaired David Willetts and Dominic Tildesley (EPFL). Representatives from Academia (Cambridge, UCL, Southampton, Edinburgh [Kenway]), Industry (GSK, Microsoft, Mozilla), RCUK (STFC, EPSRC, BBSRC), Government (TSB, Met Office), and the charitable sector.

“The ELC advises government on all aspects of e-infrastructure including networks, data stores, computers, software and skills. [It is] [r]esponsible for developing a strategy to provide a world class e-infrastructure and High Performance Computing (HPC) capability for the UK. It works in partnership with stakeholders across the academic community, industry, government and society.”

*Open Data Institute*

http://www.theodi.org

Chaired by Nigel Shadbolt. CEO Gavin Shanks.

“Open data is information that is available for anyone to use, for any purpose, at no cost. …

The Open Data Institute [ODI] will catalyse the evolution of an open data culture to create economic, environmental, and social value. It will unlock supply, generate demand, create and disseminate knowledge to address local and global issues … We will enable anyone to learn and engage with open data, and empower our teams to help others through professional coaching and mentoring. … The ODI has secured £10 million over five years from the UK Government (via the UK innovation agency, the [Technology Strategy Board](http://www.innovateuk.org/)), and $750,000 from [Omidyar Network](http://www.omidyar.com/), and is working towards long-term sustainability through match funding and direct revenue.”C.1 RCUK Strategic Framework for Capital Investment

Published November 2012

http://www.rcuk.ac.uk/documents/publications/RCUKFrameworkforCapitalInvestment2012.pdf

The RCUK document first notes the need to support the relevant e-Infrastructure, such as data centres and energy-efficient High Performance Computing (HPC). The E-Infrastructure Leadership Council now plays an important role in guiding activity in this area. The document then goes on to discuss:

“A National Framework for Biomedical Informatics Infrastructure: The UK has some of the world’s best and most complete biomedical, healthcare and social data. Investment is required to build a co-ordinated, open-access infrastructure to drive knowledge generation from genotype (simple ‘big-data’) to phenotype (massively complex ‘big data’) to ensure that health research has access to a suite of informatics technologies

Administrative data centres: To drive forward responsible data sharing it will be imperative to establish a national network of Administrative Data Centres. Overseen by a series of data access advisory boards, the centres will facilitate the anonymised linkage of personal administrative records in safe and secure settings. Using public data drawn from right across government the centres will they to provide critical new insights for combating poverty, poor mental and physical health and the drivers of criminal behaviours.

Business data: The formation of new private sector organisations will be critical to fuelling economic recovery and driving forward future UK innovation and growth. Yet our understanding of how successful new organisations are created, operate and succeed in the face of fierce global competition is hampered by fragmented and underutilised information about their structure, activities and performance. There is a pressing need to bring together and provide new software tools to support enhanced access to core organisational survey data such as the Workplace Employment Relations Survey, the Annual Survey of Hours and Earnings UK Innovation Survey (UKIS) and the National Employer Skills Survey. This must be accompanied by the creation of new data resources, including facilities that will support access to customer databases and held by major retailers, utilities and other companies to understand customer choice and behaviours and to more effectively target goods and services.

Longitudinal studies: The UK has a unique, world-leading collection of longitudinal studies spanning 65 years which follow the life trajectories of families and individuals. New investments such as the Life Study will strengthen this series bringing together social, economic and biological measures to deepen our understanding of early childhood health and development. The enhancing of other studies, including the Millennium Cohort and Understanding Society, through the for example large scale addition of biological samples and responsible linkage to administrative data sources with radically extend our capacity to understand the key influences that affect people’s life chances and shape policy intervention to improve life outcomes.

Materials collections and data: Making research data and materials available to users is essential to deliver faster progress in research, better value for money and higher quality research. RCUK has developed data policy principles and is committed to promoting greater access to and use of data in ways that are equitable, ethical and efficient. Specific capital investment is required to catalyse progress.”

## C.2 Eight Great Technologies

David Willetts, Published January 2013

http://www.policyexchange.org.uk/images/publications/eight%20great%20technologies.pdf

Big Data is the first of the “eight great technologies” discussed by David Willetts, and underpins several of the others. This document highlights areas where the UK has distinctive strengths – expertise, and access to extensive data sets – these are also true of the Scottish scene more particularly.

“We have a comparative advantage in IT because of two distinctive strengths. First, we are good at the algorithms needed to handle diverse large data sets, with strengths in mathematical and computer sciences. Secondly, we have some of the world’s best and most complete data-sets in healthcare, demographics, agriculture and the environment.” (pp12-14)

In addition, while noting the general drive towards big data in the physical, medical, and social sciences, the document notes that the UK’s urban identity could make it particularly well-placed to take more of a lead on the smart cities agenda:

“IT is thriving in these urban environments and it is worth understanding why as it gives us another insight into our comparative advantages. The sheer density of data about us living closely together in a city is of enormous value. So the challenge is to link up the data about everything from traffic conditions to energy use to enable cities to work better. Humans are middle men who can be cut out as the internet of things links them without us.” (p16).

It is perhaps also worth noting that having trust-worthy agencies and systems underpins successful citizen engagement with government and commercial online services; hence, cybersecurity is another area of interest in fostering a successful digital economy; this factor is made explicit in the Information Economy Strategy.

## C.3 Information Economy Strategy

Published June 2013

https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/206944/13-901-information-economy-strategy.pdf

The strategy notes that “95 per cent … of the 120,000 enterprises in the UK information economy employ fewer than ten people.” (p12). In discussing the daya revolution, it observes that

“Information is an unusual good: the more widely it is shared, the more people benefit from it. The taxpayer currently funds the production of an array of information such as legislation; academic research; maps and postcode frameworks; school and hospital performance metrics; and a broad swathe of national statistics. The Government believes that publicly funded information should be freely available; and be provided in formats that computers as well as people can easily read.

The UK is leading the world on open data, through its data.gov.uk portal which brings together over 9,000 data sets into one searchable website, and the world’s first Open Data Institute (ODI) which aims to foster the creation of value from open data.” (p13).

Discussing the importance of clustering, and against a background in which much resource (such as that controlled by the TSB) has been devoted to nurturing Tech City in East London, the strategy acknowledges the importance of regional clusters: “It is widely recognised that industrial clustering can bring benefits for the businesses and organisations involved. Clusters create an environment where companies can collaborate and innovate. Successful clusters help companies attract the best people and investment. There are number of established and emerging digital technology clusters across the UK including Cambridge, East London, Manchester, Glasgow, Bristol, Malvern and Sunderland. These, and other, smaller, clusters feature highly on the priorities of the relevant local enterprise partnerships (LEPs) …

In April 2013, the Tech City Investment Organisation in East London brought together over twenty leaders and influencers from technology clusters across the UK and Northern Ireland for the first ever Technology and Business Cluster Summit. This led to the creation of the UK Tech Cluster Alliance. One of the main objectives of the Alliance is to help gain further insight into the needs of technology clusters. The Alliance set out a number of themes that will shape its future work, namely: the need to change cultural perception around technology and entrepreneurship; nurturing talent and skills; supporting high growth businesses; commercialising great ideas; and preventing intellectual property from leaving the country.” (p18).

It is positive that one Scottish cluster is mentioned here, because it is plausible to assume that the UK Tech Cluster Alliance may come to have some say in setting policy for cluster support.

The strategy identifies the need “to build the capability and the capacity in the UK to be at the forefront of extracting knowledge and value from data for the benefit of citizens, business, academia, and Government.” (p28).

“This will require continued investment in and development of the physical and virtual infrastructure, such as high performance computing and data centres. We also need to boost our workforce skills and the research base of data scientists and analysts across disciplines. There is a real opportunity to develop that capability and capacity, and at the same time create new business opportunities, through solving real world problems and challenges. … Government holds and analyses a vast amount of data, and is leading the world in making that data available. Deloitte’s market analysis, published alongside Stephan Shakespeare’s review of public sector information in May 2013, gave a value of £1.8 billion per annum as the direct economic benefit from re-use of public sector information, and £6.8 billion per annum when wider economic and social impacts were considered. The Government response to the Shakespeare Review broadly accepts the recommendations, which include the recommendation for a growth-focused “national data strategy” to provide business with clarity about the Government policy on open public data, and provide the certainty needed to encourage innovation and investment in new data-driven opportunities.” (p28).

The strategy in fact goes further, in stating “Companies and organisations across business sectors have a vast supply of data to be analysed. They could therefore benefit from advances in data science and analytics. We should explore how industry and academia can collaborate to stimulate development of data analytics – and if there are barriers to doing so, how they can be resolved.” (p29). The strategy points forward to the Strategy for UK Data Capability, discussed later.

Within the Information Economy Strategy, attention is also drawn to the midata programme, which aims to enable citizens inspect the data held on them by bodies including commercial groups; in this connect, the document notes: “the midata Innovation Laboratory has been set up as a voluntary accelerator project in partnership between Government, the Open Data Institute, the Information Commissioner’s Office, business and consumer organisations. The laboratory will stimulate innovation in data services and applications and demonstrate the growth potential of midata.” (p37).

Finally, regarding smart cities, the TSB’s investments in this area are noted: “£24 million to fund a large scale Future Cities demonstrator in Glasgow, £3 million each to Bristol, London and Peterborough, a further £50 million over five years to create a Future Cities Catapult in London and £5 million in an SBRI competition to support innovative companies to create new solutions to challenges identified by UK cities.” (p38). “The Future Cities Demonstrator in Glasgow will enable businesses in Glasgow to test, in practice and at scale, new solutions for connecting and integrating their city systems, to deliver practical benefits for visitors and residents, attracting hi-tech jobs.” (p39).

Several of the projects are then listed – such as the Big Data Store – and these are extracted in Appendix A.3 of the current document. In addition, notable commitments to action are as follows:

“Action: The Government will establish a Smart Cities Forum, comprising representatives from Departments, cities, business and the research community. The Forum will bring together those with an interest in smart systems to develop and coordinate policy more effectively. It will provide advice to Ministers and local government leadership, ensuring that policy makers and city leaders are informed by a global perspective of best practice.

Action: Through our data capability strategy the Government will examine the feasibility of a (randomised control) trial through which public data is made available to pilot cities.

Action: BSI will work with stakeholders to identify where standards can help address barriers to implementing smart city concepts, including the interoperability of systems and data sharing between agencies, promoting the uptake of smart cities.

Action: UKTI will work with the Smart Cities Forum to ensure that UK firms are supported in their efforts to export their expertise in world markets.” (p40).

## C.4 Open Government Partnership UK National Action Plan 2013 to 2015

Cabinet Office, Published October 2013

http://data.gov.uk/sites/default/files/library/20131031\_ogp\_uknationalactionplan.pdf

The document promises that “UK government will continue to develop and list an inventory of all the datasets it owns, whether published or unpublished, in order to identify the National Information Infrastructure (NII) – the datasets which are likely to have the broadest and most significant economic and social impact if made available. The identification of the NII will facilitate discussions to prioritise the release of these datasets.”

The document makes a number of commitments. Two representatives ones are as follows:

“NHS England will work with governments and civil society organisations internationally to create an online space to share experiences of embedding high quality standards into information, with a view to building an accreditation scheme to enable citizens and organisations to assess their progress.”

“The UK government will issue a revised Local Authorities Data Transparency Code requiring local authorities to publish key information and data. This will place more power into citizens’ hands and make it easier for local people to contribute to the local decision making process and help shape public services.”

Both of these commitments might be read as being specific to England (a later commitment is Scotland-specific). This being so, it would be good to seek clarity on the intended reach of the commitments made in this document, so that relevant agencies in Scotland are identified.

## C.5 National Information Infrastructure

Published 31 October 2013

<https://www.gov.uk/government/publications/national-information-infrastructure/national-information-infrastructure-narrative>

Aligned with the Open Government Partnership Action Plan, the National Information Infrastructure document sets out which government-associated datasets should be prioritised, in efforts to stimulate the economy via the release of data: “G8 members identified 14 high-value areas, jointly regarded as data that will help unlock the economic potential of open data, support and encourage innovation, and provide greater accountability to improve our democracies. The UK has aligned these categories to inform the creation of its NII. … Datasets listed against Transport and Infrastructure include datasets owned and held by government agencies, ALBs [arms length bodies] and the wider transport industry, reflecting the organisation of information in the sector.

Overlaying these data themes, we have analysed user feedback, ODUG [Open Data User Group] benefits cases, applications and services which successfully use government data, and expert feedback to develop 4 primary uses of data. These are:

1. Location: Geospatial data which can inform mapping and planning.
2. Performance and Delivery: Data which shows how effectively public bodies and services are fulfilling their public tasks and the delivery of policy.
3. Fiscal: Government spend, procurement and contractual data as well as data about the financial management of public sector activities. This also includes data that government holds about companies which may be of value to users.
4. Operational: Data about the operational structure, placement of public service delivery points and the nature of the resources available within each of them.

We will encourage departments to put all of the datasets currently available under the Open Government License through the ODI’s open data certification process, prioritising those included in the NII, and make the outcome available through data.gov.uk. There will be a strong expectation that departments will adhere to the best practice embodied in the ODI open data certificate for new dataset releases.” (p7).

*nii.tiff*

Source: <https://www.gov.uk/government/publications/national-information-infrastructure/national-information-infrastructure-narrative>

## C.6 Seizing the Data Opportunity: A Strategy for UK Data Capability

Published October 2013

https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

This document contains more detailed discussion thana the Information Economy Strategy, and especially identifies needs to build UK expertise in data science, at multiple levels. Annex A (p47) summarises the list of actions:

1. The government will work with employers, e-skills UK, Nesta, Universities UK and the Open Data Institute to explore the skills shortages in data analytics and set out clear areas for government and industry collaboration.
2. The government will hold a workshop in November 2013 bringing together representatives from universities, businesses and other relevant bodies to discuss computer science graduates and how to get the right skills to meet current and future needs.
3. Universities UK will review how data analytics skills are taught across different disciplines and assess whether more work is required to further embed these skills across disciplines.
4. The government will work with the Information Economy Council, e-skills UK and Intellect to develop a plan to bolster the image of the discipline by spring 2014. As part of this, the government will work with the Information Economy Council, Research Councils and other the relevant professional bodies, including BCS and IET, to collate career profiles of people working in data analytics and the different career pathways.
5. The government will work with the Data Centre Alliance, Intellect and UKTI on options to attract overseas investment and customers to the UK data storage market.
6. The E-infrastructure Leadership Council will monitor a programme of activity to drive awareness, support, and access to e-infrastructure for businesses across six key sectors, as well as a separate campaign specifically aimed at SMEs.
7. The EPSRC is developing a proposal for a national network of centres in big data analytics to be considered as part of the Research Councils’ UK Strategic Framework for Capital Investment and dependent on their future delivery plan funding.
8. At the Open Science Data Forum in early 2014, the research community will work together to develop proposals to support the access to, and use of, research data.
9. The government will convene a working group on widening the midata programme, including stakeholders from the private sector, CDEC and consumer organisations.
10. Following the technical review of the published draft legislation on copyright exceptions, the government will bring into force secondary legislation to enable text and data mining for non-commercial purposes in 2014.
11. Working with the Information Economy Council, the government will look at options to promote guidance and advice on the rights and responsibilities of data users.

# Appendix D: International Context

This appendix points first to a couple of initiatives and a report, indicating some of the activities under way in the USA, and then summarises big data research and infrastructure opportunities imminent under the EU Horizon 2020 Programme.

## D.1 The United States

*1.* [*National Big Data Research and Development Initiative*](http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal)

http://www.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data

In 2012, “the Obama Administration announced the [National Big Data Research and Development Initiative](http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal)—a major step toward addressing the challenge and opportunity of “Big Data.”  … At its launch, the Big Data Initiative featured more than [$200 million in new commitments](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf) from six Federal departments and agencies aiming to make the most of the explosion of Big Data and the tools needed to analyze it.” The departments and agencies were:

* National Science Foundation
* Department of Defense – Data to Decisions
* National Institutes of Health – 1000 Genomes Project Data Available on Cloud
* Department of Energy – Scientific Discovery Through Advanced Computing
* US Geological Survey – Big Data for Earth System Science
* NSF & NIH - Core Techniques & Technologies for Advancing Big Data Science & Engineering

*2. Demystifying Big Data: A Practical Guide to Transforming the Business of Government*

TechAmerica, October 2012

http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf

Later in 2012, TechAmerica, a US ICT industry association, argued that big data required specific measures from government: “From a policy perspective, the federal government should examine existing organizational and technical structures to find and remove barriers to greater Big Data uptake and, where needed, take action to accelerate its use. Specifically, the government should:

1. Expand the talent pool by creating a formal career track for line of business and IT managers and establish a leadership academy to provide Big Data and related training and certification.
2. Leverage the data science talent by establishing and expanding “college-to-government service” internship programs focused specifically on analytics and the use of Big Data.
3. Establish a broader and more long-lasting coalition between industry, academic centers, and professional societies to articulate and maintain professional and competency standards for the field of Big Data.
4. Expand the Office of Science and Technology Policy (OSTP) national research and development strategy for Big Data to encourage further research into new techniques and tools, and explore the application of those tools to important problems across varied research domains.
5. Provide further guidance and greater collaboration with industry and stakeholders on applying the privacy and data protection practices already in place to current technology and cultural realities.” (p8).

The report draws attention to a number of lessons learned in big data initiatives so far. From these, two points especially stand out:

* “Successful Big Data initiatives seem to start not with a discussion about technology, but rather with a burning business or mission requirement that government leaders are unable to address with traditional approaches.
* Successful initiatives tend to follow three “Patterns of Deployment” underpinned by the selection of one Big Data “entry point” that corresponds to one of the key characteristics of Big Data – volume, variety and velocity.” (p7).

As well as applications already mentioned, the TechAmerica report also notes big data applications in education, weather forecasting, and threat detection in cybersecurity. “Ultimately, agencies should strive to address the following two questions – “How will the business of government change to leverage Big Data?” and “How will legacy business models and systems be disrupted?”” (p15).

*3. Berkeley Institute for Data Science*

http://newscenter.berkeley.edu/2013/11/13/new-data-science-institute-to-help-scholars-harness-big-data/

November 13, 2013: “The Berkeley Institute for Data Science, to be housed in the campus’s central library building, is made possible by grants from the Gordon and Betty Moore Foundation and the Sloan Foundation, which together pledged $37.8 million over five years to three universities – UC Berkeley, the University of Washington and New York University – to foster collaboration in the area of data science. …

The partnership was announced Nov. 12 at a Washington, D.C., event, “Data to Knowledge to Action,” sponsored by the White House and hosted by John Holdren, assistant to the President for Science and Technology and director of the White House Office of Science and Technology Policy…

UC Berkeley researchers are already at the forefront of data science, as evidenced by the recent creation of the Social Sciences Data Laboratory (D-Lab) for data-intensive social science research; the AMPLab (Algorithms Machines People), which focuses on machine learning; the Simons Institute for the Theory of Computing; and a Masters of Data Science program in the School of Information.”

## D.2 European Union Horizon 2020 Opportunities

In the draft workplans for the EU’s Horizon 2020 Programme, big data attracts investment via several calls, covering both research and infrastructure. The following is a summary of the most obvious opportunities.

ICT 15 – 2014: Big data and Open Data Innovation and take-up (2014; Innovation actions: €39M; coordination actions: €11M). Addressing the general … data challenges that concern entire value chains and/or bridge across borders, languages, industries and sectors. The aim is to improve the ability of European companies to build innovative multilingual data products and services ... One innovation project will establish a European open data integration and reuse incubator for SMEs to foster the development of open data supply chains; the rest will focus on … technology transfer in multilingual data harvesting and analytics solutions and services. Among the coordination actions, there will be a network … of European skills centres for big data analytics technologies and business development.

ICT 16 – 2015: Big data – research (2015; Research & Innovation actions: €38M; coordination actions: €1M). Addressing … fundamental … problems [around] scalability and responsiveness of analytics capabilities (such as privacy-aware machine learning, language understanding, data mining and visualization). To cover: (i) developing novel data structures, algorithms, methodology, software architectures, optimisation methodologies and language understanding … [for] data analytics, data quality assessment and improvement, prediction and visualization … at … scale and with diverse … data; (ii) defining relevant benchmarks in domains of industrial relevance.

EINFRA-1-2014 – Managing, preserving and computing with big research data (2014: €55M) Development and deployment of integrated, secure, permanent, on-demand service-driven, privacy-compliant and sustainable e-infrastructures incorporating advanced computing resources and software are [needed] to increase the capacity to manage, store and analyse … complex datasets, including text mining of large corpora.

EINFRA-5-2015 – Centres of Excellence for computing applications (2015: €40M) Establishing a … number of Centres of Excellence (CoE) is necessary to ensure EU competitiveness in … HPC for addressing scientific, industrial or societal challenges.

EINFRA-9-2015 – e-Infrastructures for virtual research environments (2015: €42M) Capacity building in interdisciplinary research communities to empower researchers through development and deployment of service-driven digital research environments, services and tools tailored to their specific needs.

FETPROACT 1 - 2014: Global Systems Science (GSS) (2014: €10M) Improving the way scientific knowledge [informs] policy and societal responses to global challenges like climate change, global financial crises, global pandemics, and growth of cities – urbanisation and migration patterns. These challenges entangle actions across different sectors of policy and society and must be addressed by radically novel ideas and thinking for producing, delivering, and embedding scientific evidence into the policy and societal processes.

FETHPC 1 - 2014: HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications (2014: €93.4M) Achieving, by 2020, the full range of technological capabilities needed for delivering a broad spectrum of extreme scale HPC systems. The designs of these systems need to respond to critical demands of energy efficiency, new delivery models, as well as to the requirements of new types of applications, including extreme-data applications.