

Yorkshire Universities
Technical Assistance Project (YUTA)

Report on Big Data in the Yorkshire and Humber

University of Leeds

May 2015

Contents

Executive Summary	Page 3
Section A: Big Data background	Page 5
Section B: Funding Opportunities	Page 7
Section C: Research and Innovation Strengths and Assets	Page 11
Section D: Extracts from LEP SEP & ESIF plans	Page 21
Section E: Recommendations and next steps	Page 24

Executive Summary

This report has been commissioned by Yorkshire Universities (YU) with funding from the YU Technical Assistance project (YUTA) to identify opportunities for collaboration between the region's universities and business base in the area of Big Data.

Big Data is a term used to recognise the increasing variety, volume, velocity and veracity of information sources. Examples of big data range from genomic data sets for entire populations to 'crowd-sourced' data of social attitudes and opinions in real time.

In terms of scope, the report covers YU's membership across three LEP areas (Leeds City Region, Humber, and York, North Yorkshire & East Riding), excluding Sheffield Hallam University and the University of Sheffield who are not partners in the YUTA project.

The starting point for the report is a review of the Local Enterprise Partnership European Structural & Investment Fund (LEP ESIF) and the Strategic Economic Plan (SEP) documents which cover the three LEP areas, however only the Leeds City LEP ESIF documentation makes reference to Big Data.

The Centre for Economics and Business Research (CEBR) estimates that the big data marketplace could benefit the UK economy by £216bn and create 58,000 new jobs in the UK before 2017 whilst a recent report from Deloitte estimates that the direct value of public sector information alone to the UK economy is around £1.8bn per annum, with wider social and economic benefits bringing this up to around £6.8bn.

In the context of Yorkshire and the Humber Big Data is important for the following sectors of the economy: Financial Services (Leeds, Halifax, Skipton); health and life sciences (Leeds, York, Huddersfield, Bradford), and the digital and creative sector (Leeds, Huddersfield, York and Airedale)

Funding for university-led research and business-led R&D developments is available from a range of sources including Research Councils UK, the EU commission, Innovate UK and the Treasury. The range of funders for Big Data activities is a reflection of the scope and excitement about this platform.

There is evidence that there is expertise in the area of big data in most HEIs in each of the Yorkshire-based universities. However, this is variable – while some institutions have only a small number of staff involved in Big Data others have made bigger investments and feature whole research groups and institutes devoted to this subject. Nationally the picture is similar and peaks of engagement are highlighted in this report. Nationally, a particular strength are networks such as the ESRC Administrative Data Service (ADS) and Farr Institutes – focussing on local government and health data respectively.

Delivery of the Leeds City Region ESIF strategy will be based around four investment priorities as set out in the ESIF and SEP documents.

These are:

- Unlocking the growth potential of businesses in key economic sectors;
- Making the most of a skilled and flexible workforce;
- Creating a resource smart City Region; and

- Delivering the environment for growth: major centres, housing and transport.

Big Data is identified as a tool for unlocking the growth potential of businesses in key economic sectors:

- Financial and professional services
- Health and Life Sciences
- Digital and creative industries

The Yorkshire universities have strength and critical mass particularly in relation to applied data analytics. This includes major Research Council Centres in both Medical Bioinformatics and Consumer & Behavioural research. Substantive applications to medicine and health are a particular opportunity given the intersection with importance of both government and private sector interests in the local area.

The region has a sound infrastructure basis for big data analytics, for example in the University sector as a hub for the N8 High Performance Computing initiative (n8hpc.org.uk), and in the commercial world in hosting IX3 as the only peer-to-peer internet exchange outside London.

The Yorkshire universities are generating a substantial cohort of graduates in computer science. In order to promote data analytics across the various communities the skills base needs to be extended. This process could be fostered by the creation of new courses with a specific emphasis on data analytics, and which might involve external partners across the full range of activities from design (e.g. advisory boards and consultation) to delivery (e.g. shared projects or even involvement in teaching).

In summary, there is significant potential for closer collaboration and integration between the region's HEIs and private and public sector organisations to fully understand and exploit Big Data.

Section A: Big Data background

Big Data is a term used to recognise the increasing variety, volume, velocity and veracity of information sources. It reflects the fact that businesses, governments and researchers are now all able to capture information at increasing scales, with exceptional speed and frequency, and from a dazzling array of sources. Examples of big data range from genomic data sets for entire populations to 'crowd-sourced' data of social attitudes and opinions in real time.

The UK's open data agenda is world-leading, with over 10,000 public datasets published on data.gov.uk, and the ground-breaking and independent Open Data Institute (ODI). Many organisations from the Health Service to commercial businesses are mining private data.

It should be noted that Research Councils UK (RCUK) have stated that there is 'no official definition for Big Data; however, RCUK describes it as more than just 'large amounts' of data, and it tends to be beyond the ability of typical software tools to capture, store, manage and/or analyse.'¹

Importance for the UK

The Centre for Economics and Business Research (CEBR) estimates that the big data marketplace could benefit the UK economy by £216bn and create 58,000 new jobs in the UK before 2017² whilst a recent report from Deloitte estimates that the direct value of public sector information alone to the UK economy is around £1.8bn per annum, with wider social and economic benefits bringing this up to around £6.8bn. Research by Nesta shows that UK data-driven firms are 40% more likely to report launching products and services ahead of their non-data savvy competitors³.

In November 2012 George Osborne, Chancellor of the Exchequer, outlined the importance of the value of science as a driver of the UK economy. In January 2013 the Government made a commitment to Big Data by classifying it as one of the 'Eight Great Technologies' which would propel the UK to future growth and help it stay ahead in the global race⁴. This was followed by an allocation of £189m from the Autumn Statement 2012 for big data and energy efficient computing. Previously, the government had already invested an extra £150m in e-Infrastructure in October 2011.

Importance for Yorkshire and the Humber

Financial Services are concentrated in Leeds, with major centres also in Halifax and Skipton. Leeds City Region is home to three of the five largest building societies in the UK and 30 national/international banks. Five of the UK's largest law firms, 150 accountancy firms, and leading private equity houses are also based locally. 18,000 people are employed in legal services, and 13,000 in accountancy. Big Data has been identified as one of the

¹ RCUK big data website, <http://www.rcuk.ac.uk/research/infrastructure/big-data/>, accessed 28 April 2015

² CEBR, Data Equity – Unlocking the value of big data, page 4

³ NESTA, Rise of the datavores, Nov 2012

⁴ Department for Business, Innovation & Skills, The Rt Hon David Willetts, speech, 24 January 2013

sector drivers due to the next generation of innovation being data driven.⁵ Leeds plays a significant role in the UK's insurance sector which is the largest in Europe, making a major contribution to the UK economy, employing 320,000 people and contributing almost 3 per cent to GDP.

Health and life sciences are strengths for Leeds City Region with centres of medical research at the universities of Leeds, York, Huddersfield and Bradford. Technological advances, including Big Data have been identified as a sector driver for demand for technical skills and support⁶.

Leeds City Region's **digital and creative sector** is the largest in the UK outside of London and there are hubs of digital expertise and enterprise in Leeds, Huddersfield, York and Airedale. Digital and creative industries are expected to grow due to the next generation of innovation in this area being data driven⁷.

Research and Innovation strengths and assets

Around the subject of Big Data the region's universities have leading national and international research strengths. The University of Leeds is 1st in the UK for carrying out Research Council funded research in Big Data.⁸

Yorkshire universities do not have any of the first generation nodes in the EPSRC-funded Turing Institute. However, given that these are focussed exclusively on theoretical research in mathematics and computation, this still provides a particular opportunity in the applications of data analytics.

⁵ LCR LEP ESIF Strategy 2014-2020, page 32

⁶ Ibid, page 34

⁷ Ibid, page 40

⁸ Ibid, page 50

Section B: Funding Opportunities

The table below highlights specific funding sources.

Funder	Detail
Research Councils (RCUK)	<p>Research Councils invest around £3bn per year in research, covering the full spectrum of academic disciplines.</p> <p>Due to the overarching RCUK requirement to demonstrate the impact of academic research there is significant potential for academia-business collaborations. RCUK makes funding available for such activities.</p> <p>http://www.rcuk.ac.uk/research/infrastructure/big-data/</p> <p>Reflecting the breadth and potential of research aligned with the big data agenda, each research council has developed its own strategy and key opportunities:</p>
	<p>Arts and Humanities Research Council (AHRC): Big data connects to a number of their wider priority areas such as the creative and cultural economy, heritage, digital transformations and the Connected Communities Programme which examines the changing nature of communities in their historical and cultural contexts and the role of communities in sustaining and enhancing our quality of life.</p> <p>Full details can be found at http://www.ahrc.ac.uk/Funding-Opportunities/Research-funding/Themes/Digital-Transformations/Pages/Big-Data.aspx</p>
	<p>Engineering and Physical Sciences Research Council (EPSRC): EPSRC leads the RCUK Digital Economy Theme, supporting research to rapidly realise the transformational impact of digital technologies on community life, cultural experiences, future society, and the economy.</p> <p>EPSRC also has Centres for Doctoral Training in Cloud Computing for Big Data at the University of Newcastle, and in Data Science at the University of Edinburgh.</p> <p>https://www.epsrc.ac.uk/research/ourportfolio/themes/digitaleconomy/</p>
	<p>Biotechnology and Biological Sciences Research Council (BBSRC): BBSRC have set up the Bioinformatics and Biological Resources Fund the purpose of which is to provide proper support for resources such as databases, genetic resources and culture collections which require long term maintenance and curation.</p> <p>http://www.bbsrc.ac.uk/funding/opportunities/2013/2013-bioinformatics-biological-resources-fund/</p>
	<p>Economic and Social Research Council (ESRC): The ESRC is investing in a Big Data Network in order to help optimise this resource. Divided into three phases, the Big Data Network will shape the knowledge of society and help prepare and evaluate better government policies in the future.</p> <p>The three phases are: Administrative Data Research Network (ADRN), Business and Local Government Data, Third Sector and Social Media Data.</p>

	<p>http://www.rcuk.ac.uk/research/infrastructure/big-data/</p> <p>Big data is a priority in the current call for research content within the National Centre for Research Methods (“analysis of online digital and big data”), in the 2015 cohort of Future Leaders projects, and in the third wave of the Secondary Data Analysis Initiative. ESRC anticipates up to £75 million further investment through the Science and Innovation Strategy.⁹</p>
	<p>Medical Research Council (MRC): There are four parts to MRC informatics research funding:</p> <p>e-Health Informatics Research Centres (e-HIRC): To provide training and career development opportunities in using large health datasets for research and carry out cutting edge research using e-health datasets from healthcare and research.</p> <p>The Farr Institute of Health Informatics Research: To provide a new digital infrastructure to facilitate data sharing, development of data standards and enhancement of research facilities incl. data safe havens.</p> <p>UK Health Informatics Research Network: To coordinate training and career development, To develop methodologies and standards, to engage with the public, To provide an interface for collaborations with industry and NHS</p> <p>Medical Bioinformatics Initiative: To support improved linkage between large scale genomic and phenotype data and electronic health records.</p> <p>http://www.mrc.ac.uk/research/initiatives/health-and-biomedical-informatics/</p>
	<p>Natural Environment Research Council (NERC): To provide the cyber-infrastructure to improve access to big data, enable the running of complex environmental models and the integration of large environmental data sets to promote further scientific understanding and working with partners to develop products and services which address the needs of business, government and society.</p> <p>Between 2013 and 2015 NERC will invest £13m capital funding to support 'Big Data', develop the cyber-infrastructure needed to give open access to big data, to provide computers capable of running complex environmental models and to capture real time data from sensors.</p> <p>NERC have also invested £2.5m in the universities of Cranfield, Newcastle, Birmingham and Cambridge to set up a doctoral training centre in Data, Risk & Environmental Analytical Methods (DREAM) to produce a cohort of researchers who can use large or complex datasets to understand and ease the risks posed by a range of societal and environmental changes, such as a rapidly expanding population, limited natural resources, and natural hazards.</p> <p>http://www.nerc.ac.uk/innovation/activities/environmentaldata/bigdatacapital/</p>

⁹<http://www.esrc.ac.uk/news-and-events/announcements/33258/esrc-welcomes-boost-of-up-to-75-million-from-the-government-to-big-data-and-the-information-economy.aspx>

Funder	Detail
EU commission	The EU commission is allocating €83bn to research and development activities under the Horizon 2020 framework.
	<p>Under the industrial leadership/ECSEL work programme there are opportunities for collaboration between the public and private sectors in areas such as Smart Mobility, Smart Society and Smart Energy, utilising big data. ECSEL is running 2014-2024 and €1.18bn of funding will be allocated.</p> <p>https://ec.europa.eu/digital-agenda/en/time-ecsel</p> <p>Another opportunity is H2020-FCT-2015: Fight against crime and terrorism, Forensics topic 1: Tools and infrastructure for the extraction, fusion, exchange and analysis of big data including cyber-offenses generated data for forensic investigation, part of the Secure Societies work programme. Total funding available €42m.</p>

Funder	Detail
Innovate UK (formerly Technology Board, TSB)	<p>TSB has set up a Digital Catapult, a national centre to rapidly accelerate the UK's best digital ideas to market to create new products, services, jobs and value for the UK economy. The Catapult is an autonomous private research and development organisation. Their projects fall into four challenge areas - Closed Organisational Data; Personal Data and Trust; Creative Content and the Internet of Things.</p> <p>https://digital.catapult.org.uk/</p> <p>Specific calls include:</p> <p>Enhancing user experience using personal data: Funding to support feasibility studies looking at innovative ways to improve the user experience in the digital economy. The competition is open to small and medium-sized enterprises singly or in collaboration. Proposals must be led by a business. (Maximum value: £2m)</p> <p>Protecting data in industry: Funding for collaborative research and development (R&D) projects that tackle the growing risks of disruption to internet-enabled businesses and their digital supply chains. Proposals must be collaborative and led by a business. (Maximum value: £4m).</p>

Funder	Detail
Treasury	<p>The Chancellor of the Exchequer, George Osborne, has committed £20m of funding to the 'Health North' initiative, which will be led by the Chairman of the North Health Science Alliance, Ian Greer¹⁰.</p> <p>Health North will build on the North's strengths in health science, enabling better care for patients, and promoting innovation through analysis of data on the effectiveness of different drugs, treatments and health pathways.</p> <p>The Northern Health Science Alliance formed by the leading northern</p>

¹⁰ <http://news.liv.ac.uk/2015/03/19/chancellor-commits-20m-of-funding-to-health-north-initiative/>

	<p>universities, teaching hospitals and Academic Health Science Networks (AHSNs) will set up the world's first partnership using large-scale data to drive public sector reform in health and social care across a 15 million-strong population in the North of England.</p> <p>By analysing integrated information and feeding this back to NHS practitioners, service managers, commissioners, public health professionals, local authority planners, researchers and policy makers, the project teams will be able to identify variations in care and needs.</p> <p>Connected Health Cities will enable new medical discoveries by working with the national Farr Institute of Health Informatics Research (see page 17 below).</p>
--	--

Section C: Research and innovation strengths and assets

The table below summarises the research strengths of the region's universities around Big Data.

University	University of Bradford (www.bradford.ac.uk)
Centres, Institutes, Groups	School of Electrical Engineering and Computer Science / Artificial Intelligence Research Group
Key individuals	Daniel Neagu: Artificial Intelligence techniques applied in toxicology, software engineering, online social networks, data quality, Big Data, healthcare.

University	University of Huddersfield (www.hud.ac.uk)
Centres, Institutes, Groups	Institute of Railway Research (part of the School of Computing and Engineering)
Key individuals	Coen Van Gulijk (RSSB-Huddersfield strategic partnership on Safety in Railways)
Major initiatives, projects and awards	Big Data Risk Analysis (with the Rail Safety and Standards Board), based at the Institute of Railway Research http://www.hud.ac.uk/research/researchcentres/irr/

University	University of Hull (http://www2.hull.ac.uk/)
Centres, Institutes, Groups	Hull University Business School
Key individuals	Richard Vigden: Big Data and business analytics and studying how ICT can be enrolled to support individual and collective behaviour change.

University	University of Leeds (www.leeds.ac.uk)
Centres, Institutes, Groups	Consumer Data Research Centre (CDRC) Medical Bioinformatics Centre (MRC) Leeds Institute for Data Analytics (LIDA) Leeds Q-Step Centre
Key individuals	Mark Birkin: CDRC Alex Markham: MRC
Major initiatives, projects and awards	Leeds Institute for Data Analytics (LIDA) is the home of the Consumer Data Research Centre (CDRC, www.cdrc.ac.uk) and the Medical

awards	<p>Bioinformatics Centre (MRC). LIDA is due to open in summer 2015 and will provide office, data storage and analysis capabilities for around 250 academics, research staff and external partners. The CDRC focuses on analysing consumer-related datasets while the MRC examines health-related datasets. LIDA is the result of a £15m investment from the University and Research Councils UK.</p> <p>Leeds Q-Step Centre is one of 15 centres across the UK for Quantitative Methods in Education to address the shortage of quantitatively trained social scientists, funded by the Nuffield Foundation, ESRC and HEFCE (http://www.nuffieldfoundation.org/q-step). Each centre will support the development and delivery of specialist undergraduate programmes, including new courses, work placements and pathways to postgraduate study.</p>
--------	---

University	Leeds Beckett University (http://www.leedsbeckett.ac.uk/)
Centres, Institutes, Groups	School of Computing, Creative Technologies and Engineering
Key individuals	Victor Chang: web applications, web services, database, grid, cloud, storage/backup, bioinformatics, and financial computing.

University	University of Sheffield (www.sheffield.ac.uk)
Centres, Institutes, Groups	Faculty of Medicine, Dentistry and Health is a partner in the Connected Health Cities project.
Key individuals	<p>Tony Weetman (Connected Health Cities)</p> <p>Helen Kennedy (Sociological Studies): critical approaches to big data analytics, especially big data visualisations, how to make data more accessible to ordinary citizens, or how to make the social life of data more public.</p> <p>Gwilym Price (Sheffield Methods Institute): Formerly Associate Director of the £10m Urban Big Data Centre (UBDC), and continues to lead the UBDC research project: URP3: Integrated Multi-sectoral Model of Employment and Migration. Core research interests are largely in the broad field of urban economics.</p> <p>Farida Vis, Information School: 'Big Data and Social Change', social media, crisis communication, data journalism, citizen engagement and discourses around Big Data.</p>
Major initiatives, projects and awards	<p>Connected Health Cities (The first investment of the government's Health North programme to unlock healthcare innovations in the English regions with the greatest health challenges.)</p> <p>AHRC Digital Transformation Big Data grant (Seeing Data – Making sense of data visualisations, http://seeingdata.org/about-seeing-data/)</p>

University	Sheffield Hallam University
Centres, Institutes, Groups	Centre of Excellence in Terrorism, Resilience, Intelligence & Organised Crime Research (CENTRIC),
Key individuals	Babak Akhgar, (Informatics and Director of CENTRIC) A full list of staff connected to CENTRIC can be found at http://research.shu.ac.uk/centric/index.php/people Kassim Mwitondi: knowledge discovery from multi-faceted data (KDMD) for tackling global challenges and sustainability, including modelling spatio-temporal data dynamics in the big data era and their practical consequences on KDMD.
Major initiatives, projects and awards	Sheffield Hallam/CENTRIC are a partner in the EU FP7 funded Athena project (313220), coordinated by the Police and Crime Commissioner for West Yorkshire, http://www.westyorkshire.police.uk/athena

University	University of York (www.york.ac.uk)
Centres, Institutes, Groups	Department of Computer Science
Key individuals	Mark Nicholson: Safety and Certification of data and data chains, links of data and Big Data to information systems and safety related / safety critical systems.

University	Leeds Trinity (www.leedstrinity.ac.uk)
Centres, Institutes, Groups	Department of Business Management & Marketing
Key individuals	Mostafa Mohamad: e-business & information system, technology management, mobile computing, Big Data analysis

Open Data Institute (www.http://actuaterfutures.com/)	
<p>The region also benefits from the expertise provided by the Leeds Open Data Institute (ODI) and Leeds ODI is the first ODI node outside London.</p> <p>The visions of the ODI network is to establish the ODI as a world-leading centre to innovate, exploit and research the opportunities for the UK created by the Government's Open Data Policy. ODILeeds operates across five cross-cutting areas:</p> <ul style="list-style-type: none"> • Members and partnerships • Events and Training • Challenges • Projects • Start-Up Accelerator 	

A full list of ODI members can be found at <https://directory.theodi.org/members>

Key individual: Paul Connell, founding partner, <http://actuatedfutures.com/team/>

ODI is a key player in the emerging Leeds Data City project whose members also include AQL, TPP, EMIS, KPMG, Bloom, and the University of Leeds.

AQL (<https://www.aql.com/>)

aql specialise in out-of-London (Leeds) colocation based in aql owned/operated facilities.

Aql in collaboration with industry partners have designed and built some of the UK's most efficient and secure datacentres. aql plays a key role in the support of Leeds' digital infrastructure – operating the main Leeds data centres, where much of the regions data flows. The size of its Leeds data centre estate is 110,000 sq ft.

Aql's HQ at Salem Chapel in Leeds is also the home of the region's internet exchange IXLeeds, the only exchange of its kind outside London.¹¹

RedCentric (<http://www.redcentricplc.com/>)

RedCentric is a leading UK IT managed services provider that offers a range of IT and Cloud services designed to support organisations in their journey from traditional infrastructure to the Cloud and hybrid combinations in between. RedCentric's head office is in Harrogate with further locations in Cambridge, London, Reading, Theale and Hyderabad.

Our Data Mutual (www.ourdatamutual.org)

Our Data Mutual will be a mutual society; an organisation owned by its members. The organisation is currently seeking expressions of interest for funding through the start-up process. The aim of Our Data Mutual is to act as a broker between individuals and organisations and their data and organisations that want to buy this data. Our Data Mutual is based in Leeds.

ResearchOne (<http://www.researchone.org/>)

ResearchOne is a health and care research database developed by TPP in partnership with the University of Leeds and the UK Government's Technology Strategy Board. The database consists of de-identified clinical and administrative data drawn from the electronic patient records currently held on the TPP SystmOne clinical system. It has the potential to be one of the largest healthcare research databases in the world, containing up to 28 million records.

The research records contain linked data from a wide variety of settings across both primary and secondary care, including General Practice, Child Health, Community Health, Palliative Hospital, Out-Of-Hours, Accident & Emergency and Acute Hospital.

ResearchOne is entirely self-funded and is run as a not-for-profit enterprise.

¹¹ <https://www.aql.com/news/275/>

At a national level the following institutions have expertise in Big Data:

University, Groups, Institute	Big Data research outside Yorkshire and Humber
University of Oxford	<p>Big Data Institute (BDI) http://www.cggh.org/ The MRC Centre for Genomics and Global Health (CGGH) will be a core part of the Big Data Institute (BDI) when it opens in 2016. CGGH is a partnership between the University of Oxford and the Wellcome Trust Sanger Institute.</p> <p>The BDI will be a centre of excellence for the analysis of large and complex data sets for health research. Main areas of activity will be epidemiology, and genomics and infectious disease surveillance, with the goal of developing innovative statistical and computational approaches for large-scale data analysis across these domains.</p> <p>Oxford Internet Institute (OII) http://www.oii.ox.ac.uk/ The OII was founded in 2001 at the University of Oxford, as an academic centre for the study of the societal implications of the Internet. Their research projects cover the themes of: Internet and society, digital government and politics, information geographies and economies, science, learning and technology and Internet governance, regulation and ethics. Key contact: William H. Dutton</p> <p>Oxford e-Research Centre http://www.oerc.ox.ac.uk/ The centre brings together global experts in the following key technology areas: scientific computing; Internet, Information & Interaction, visual computing; Software, Repositories & Data Management and e-infrastructure. Key contact: David De Roure</p>
University of Bristol	<p>Faculty of Engineering – Intelligent Systems group</p> <p>Nello Cristianini: artificial intelligence, data science, computational social sciences, digital humanities, big data and society.</p>
University of Warwick	<p>Research on big data is carried out in all four of the institution's faculties and through cross-faculty interdisciplinary and multidisciplinary approaches.</p> <p>Warwick maintains a central webpage which summarises their expertise in the area of big data: http://www2.warwick.ac.uk/research/bigdata/academics/</p> <p>Key contact: Jon Coaffee: urban geography, cities; urban resilience; security; counter-terrorism; urban governance; infrastructure; mega-events. Links with New York University's Center for Urban Science and Progress (CUSP, http://cusp.nyu.edu/about/). CUSP observes, analyses, and models cities to optimize outcomes, prototype new solutions, formalize new tools and processes, and develop new expertise/experts.</p>

	<p>Key contact: Stephen Jarvis: computer science; developing predictive techniques capable of determining the detailed execution characteristics and resource use of industrial strength applications running on large high-performance and distributed systems.</p>
<p>University of Southampton</p>	<p>ESRC Administrative Data Research Centre (ADRC)</p> <p>Hosted by the Statistical Sciences Research Institute, working with colleagues in Geography, Engineering and the Environment and the information and communications technology professional services department, iSolutions. The Centre provides state-of-the-art secure facilities with access to high performance computer systems, database management systems and advanced data analysis and statistical tools (http://adn.ac.uk/about/research-centre-england/our-core-team).</p>
<p>University of Edinburgh</p>	<p>ESRC Administrative Data Research Centre (ADRC)</p> <p>Chris Dibben is the Director of the Administrative Data Research Centre for Scotland. He is also the Chair in Geography at the University of Edinburgh and part of the Geography and the Lived Environment research institute and the Director of the ESRC's Administrative Data Liaison Service.</p> <p>Andrew Morris (Medicine) is Director of the Biomedical Research Institute and Vice-Principal Data Science. He is Director of the Farr Institute in Scotland funded by the MRC and nine other funders and Convenor of the UK Health Informatics Research Network, representing a £39m investment in health informatics research.</p> <p>http://adn.ac.uk/about/research-centre-scotland.</p>
<p>Swansea University</p>	<p>ESRC Administrative Data Research Centre (ADRC)</p> <p>David Ford (Health Informatics) specialises in using routinely collected data to support and conduct research. He has been involved in a wide range of large-scale, multi-centre studies of complex interventions and evaluations of policy and service initiatives. He is joint lead of the Health Information Research Unit for Wales, which develops new ways of harnessing the potential of information collected routinely in health and other settings. The Unit's main product is the SAIL Databank, an internationally recognised data linking resource made up of a wide variety of routinely collected data from across Wales.</p> <p>http://adn.ac.uk/about/research-centre-wales</p>
<p>Queens University Belfast</p>	<p>ESRC Administrative Data Research Centre (ADRC, http://adn.ac.uk/about/research-centre-n-ireland)</p> <p>Dermot O'Reilly is Director of the Administrative Data Research</p>

	<p>Centre in Northern Ireland. He is based at the School of Medicine, Dentistry and Biomedical Sciences at Queen's University Belfast, specialises in Public Health and Social Epidemiology and has carried out extensive research into health inequalities. Interests focus on the effective use of routine data collections to further our understanding of disease and social causes and how this affects society.</p>
--	---

<p>University of Essex</p>	<p>ESRC Administrative Data Service (ADS, https://adrn.ac.uk/about/ads)</p> <p>The Administrative Data Research Network gives accredited researchers access to linked, de-identified administrative data in a secure environment. The ADS coordinates the Network, and is the first point of contact for researchers who want access to administrative data. It is based at the University of Essex, with partners at the Universities of Manchester, Oxford, the West of England, and Edinburgh.</p> <p>The Administrative Data Service shares in the expertise of the UK Data Archive, the UK Data Service and its Secure Lab.</p> <p>ESRC Business and Local Government Data Research Centre (http://www.blgdataresearchcentre.com/)</p> <p>This is a national data research centre which applies academic expertise to help companies, local authorities and academics explore data, enhance knowledge and empower society. This service is provided at no cost to the external partner.</p> <p>Key contact: Vania Sena, Director of the ESRC Business and Local Government Data Research Centre</p>
-----------------------------------	---

<p>University College London (UCL)</p>	<p>Department of Geography</p> <p>Paul Longley: development and application of geographic information science, latterly with strong emphasis on the deployment of Big Data and Open Data to create geo-temporal demographic indicators at neighbourhood scales, and to model their distributions over space and time.</p> <p>James Cheshire: the use of "big" and open datasets for the study of social science.</p> <p>John Shawe Taylor: Director of the Centre for Computational Statistics and Machine Learning (CSML). Research interests include statistical learning theory, neural networks, machine learning, graph theory.</p> <p>Philip Treleaven: Director of the Financial Computing Centre. Research interests include the use of computational techniques in finance, automated fraud detection, algorithmic trading, virtual environments and computer graphics (e.g. 3D Body Scanning) for the Clothing Industry. He is the director of the UK National Sizing</p>
---	---

	Survey (SizeUK).
University of Liverpool	<p>Department of Geography and Planning</p> <p>In this theme, spatial aspects of social and economic processes are investigated, focusing on quantitative and computational approaches. The current increase in readily available large data sets, coupled with increases in computer power, allows them to investigate social and economic trends in a number of new and exciting ways.</p> <p>Research covers a wide range of topics, including the analysis of health patterns, the geography of crime, family and fertility dynamics, educational and other lifecourse choices, migration and residential mobility, residential and housing segregation and the effectiveness of the spatial targeting of public policy measures.</p>
University of Manchester	<p>Big Data Community</p> <p>The Big Data Community will seek to build on Manchester University's internationally recognised strengths in Big Data research, addressing common challenges and promoting interdisciplinary collaboration, both internally and externally, in all forms of Big Data research.</p> <p>Expertise is summarised at https://www.informatics.manchester.ac.uk/datascience/Pages/BigData.aspx</p> <p>The University of Manchester is a key partner in RETHINK big, a European project bringing together key hardware, networking and systems architects with the key producers and consumers of Big Data across Europe. The aim is to deliver a strategic roadmap that will maximise European competitiveness in the processing and analysis of Big Data over the next 10 years (http://www.rethinkbig-project.eu/project/about).</p>
University of Glasgow	<p>Urban Big Data Centre (UBDC)</p> <p>The UBDC is a research resource promoting the use of innovative methods and complex urban data to address global city challenges.</p> <p>It works with partners in business, government, academia and more offering a wide range of services that can help organisations make sound data-driven decisions (http://www.gla.ac.uk/schools/education/research/researchprojects/urbanbigdatacentre/)</p>
Farr Institute	<p>The Farr Institute of Health Informatics Research comprises four nodes distributed across the UK and led from the University College London (Farr Institute @ London), University of Manchester (Farr</p>

	<p>Institute @ HeRC N8), Swansea University (Farr Institute @ CIPHER), and the University of Dundee (Farr Institute @ Scotland). With a £17.5m research award from a 10-funder consortium, plus additional £20m capital funds from the Medical Research Council, the Farr Institute aims to deliver high-quality, cutting-edge research linking electronic health data with other forms of research and routinely collected data, as well as build capacity in health informatics research.</p> <p>The Farr Institute aims to provide the physical and electronic infrastructure to facilitate collaboration across the four nodes, support their safe use of patient and research data for medical research, and enable partnerships by providing a physical structure to co-locate NHS organizations, industry, and other UK academic centres (http://www.farrinstitute.org/)</p>
<p>Farr Institute @ HeRC N8 (University of Manchester)</p>	<p>The Health eResearch Centre (HeRC) is a regional health informatics research centre feeding e-health applications into the Northern Health Science Alliance (an alliance of North England's research-intensive universities and teaching hospitals, covering a 15 million population).</p> <p>HeRC is an expansion of the Northwest Institute for Bio-Health Informatics, hosted by the University of Manchester, which has delivered some of the UK's strongest research outputs in health informatics and spun out a services arm, embedded in the NHS, known as Northwest eHealth. Over the past eight years this combination of research and applications has generated in excess of £50m investment in the regional economy, mainly from RCUK and industry.</p> <p>HeRC has a particularly strong engineering focus and is now developing e-infrastructure that could support national-scale sense-making over linked health data (http://www.farrinstitute.org/centre/HeRC/63_About.html)</p>
<p>Farr Institute @ CIPHER (Swansea University)</p>	<p>The Centre for Improvement in Population Health through e-records research (CIPHER) is a multinational research partnership focused on improving the health and wellbeing of the population through the use of health informatics research.</p> <p>CIPHER's research aims to reduce delays between knowledge discovery, intervention assessment and adoption, and population impact by:</p> <ul style="list-style-type: none"> • Developing new methods to anonymise, link and analyse healthcare and similar data. • Establishing collaborations to link previously isolated silos of expertise in observational, interventional, biomedical and social sciences. • Improving knowledge exchange between academic, practitioner and policy leads. • Liberating information trapped in data islands.

	<ul style="list-style-type: none"> • Engagement with the public, policy makers and practitioners <p>CIPHER's research focus has particular emphasis on a number of areas:</p> <ul style="list-style-type: none"> • Chronic diseases • Factors affecting maternal and child health • Infection • Injury • Mental health • Substance use • The built environment <p>http://www.farrinstitute.org/centre/CIPHER/34_About.html</p>
--	---

<p>Farr Institute @ London</p>	<p>The Farr Institute of Health Informatics Research, London is a collaboration between the Higher Education Institutions of UCL, the London School of Hygiene & Tropical Medicine and Queen Mary University of London, the National Health Service organisations in UCL Partners and Public Health England.</p> <p>The aim is to establish research programmes spanning the translational cycle, exploiting our nationally leading informatics strengths in areas of major clinical and public health importance.</p> <ul style="list-style-type: none"> • Develop innovative methodological approaches to the health informatics data cycle, central to high-quality patient care and translational research particularly where innovation can be scaled to and add value to the wider UK Centre Network. • Drive a step change in capacity development, through strategic development across partner organisations of training and career opportunities for clinicians, those in technical and enabling roles, and the wide range of academic disciplines relevant to health informatics research. • Promote a new culture of engagement between the public, patients, clinicians and researchers based on joint development of consent models and electronic health record research to improve patient care and public health. <p>http://www.farrinstitute.org/centre/London/6_About.html</p>
---------------------------------------	--

<p>Farr Institute @ Scotland</p>	<p>The Farr Institute @ Scotland is a collaboration between six Scottish Universities (Dundee - Co-ordinating Centre, Aberdeen, Edinburgh, Glasgow, St Andrews, Strathclyde) and NHS National Services Scotland. The aim is to both improve the health of the Scottish population and place Scotland as a global leader in health informatics research</p> <p>(http://www.farrinstitute.org/centre/Scotland/3_About.html)</p>
---	---

Section D: Extracts from LEP SEP & ESIF plans

This section highlights relevant information from the SEP and ESIF plans:

LEP ESIF Plan	Key information
Leeds City Region	<p>Delivery of the Leeds City Region ESIF strategy will be based around four investment priorities as set out in the ESIF and SEP documents:</p> <p>These are:</p> <ul style="list-style-type: none"> • Unlocking the growth potential of businesses in key economic sectors; • Making the most of a skilled and flexible workforce; • Creating a resource smart City Region; and • Delivering the environment for growth: major centres, housing and transport.¹² <p>Big Data is identified as a tool for unlocking the growth potential of businesses in key economic sectors:</p> <ul style="list-style-type: none"> • Financial and professional services¹³ • Health and Life Sciences¹⁴ • Digital and creative industries¹⁵

In the section on UK and global economic trends and their implications for Leeds City Region, the ESIF strategy highlights the challenge of an **ageing workforce** and a lack of appropriately qualified replacement labour¹⁶. New technology and the use of big data will play a pivotal role in meeting that challenge the document states¹⁷.

LEP SEP Plan	Key information
Leeds City Region	<p>The Leeds City Region (LCR) SEP is the LEP's long term vision for the City Region economy. It is designed to provide the foundations for growth, aligning their plans with those of public and private sector partners and with government for maximum impact. It demonstrates the opportunities for investment in the City region and highlights the significant gains that can be made by all who co-invest in Leeds City Region. The SEP also forms the basis of their Growth Deal negotiations with government¹⁸.</p> <p>The LEP's vision is to unlock the potential of the City Region, developing an</p>

¹² LCR LEP ESIF Strategy 2014-2020, page 12

¹³ Ibid, page 33

¹⁴ Ibid, page 35

¹⁵ Ibid, page 40

¹⁶ Ibid, page 104

¹⁷ Ibid, page 104

¹⁸ LCR SEP Part A, page 3

	<p>economic powerhouse that will create jobs and prosperity.</p> <p>The headline targets by 2021 are:</p> <ul style="list-style-type: none"> • £5.2bn additional economic outputs beyond current projections • 62,000 jobs additional jobs • £675m in benefits savings <p>thereby making the City Region a net contributor to the national economy¹⁹.</p> <p>The overall plan is founded on the following four strategic investment priorities:</p> <ul style="list-style-type: none"> • Supporting growing businesses • Developing a skilled and flexible workforce • Building a resource-smart city region • Delivering the infrastructure for growth²⁰
--	--

Long term the LEP's ambitions over the next decade and beyond, are to:

- enable vibrant private sector growth, based on innovation and exports;
- create a NEET-free City Region, with more and better jobs, and the skilled and flexible local workforce to sustain them;
- become a lean, resource efficient economy underpinned by a 21st century energy infrastructure;
- build a 21st century physical and digital infrastructure that enables the LEP to reach our growth potential;
- and to make the most of the opportunities presented by HS2 – not just the economic gains from this step change in connectivity, but also the regeneration of towns and cities across the City Region, and the jobs, new skills and business opportunities it will bring²¹.

Creating a smart Leeds City Region is one of the ambitions outlined in the SEP:

In building a more innovative, dynamic and better-connected City Region, the LEP also aspires to create a much smarter economy. Our City Region already generates vast quantities of data from events and processes that are monitored and recorded electronically – from bus movements to patient records or air quality. As objects such as cars, domestic appliances and clothes become sources of data themselves in the years to come, the volume of data available will continue to grow exponentially.

'Smart cities' will learn to harness the power of digital technologies and big data to create services for their residents that are both higher quality and more efficient. It is estimated that the value of 'smart cities' services in the UK will be \$20bn by 2020.

The LEP will provide the foundations for the City Region to become a smarter economy through the delivery of superfast and hyperfast broadband across the area. The LEP aims to exploit these great opportunities for business growth and improved public services through programmes in transport, open data, health innovation and energy²².

¹⁹ LCR SEP Part A, page 17

²⁰ LCR SEP Part A, page 24

²¹ LCR SEP Part A, page 18

²² LCR SEP Part A, page 21

The SEP also acknowledges the strengths, assets and opportunities that will be promoted:

- The LCR LEP is the home to two of the world's leading providers of electronic patient record systems, EMIS and TPP;
- Leeds City Region is the UK's largest centre for financial, professional and business services outside London – over a quarter of a million work in these sectors, with 60,000 in financial services alone;
- the City Region's status as a major financial centre stems from its history as the birthplace of the Halifax and Bradford & Bingley building societies, and of three of the UK's current five largest building societies: the Leeds, Yorkshire and Skipton societies;
- The LCR LEP is home to 30 national and international banks (including the Bank of England's only presence outside London), and to leading private equity houses;
- the City Region is also home to large insurance providers such as Aviva and Hiscox, and business intelligence and credit service companies such as Call Credit and Equifax, as well as Engage Mutual Assurance in Harrogate;
- five of the UK's largest law firms and 150 accountancy companies are here too, with 18,000 people employed in legal services and 13,000 in accountancy; and
- the City Region, has a niche strength in data processing to support financial services with large multinational companies such as TSYS (also in York) InTechnology and VodaLink, based in Harrogate²³

In addition, LCR is home to world-leading university research:

The LEP's higher education sector represents one of the largest concentrations of institutions and students anywhere in Europe – there are eight Higher Education Institutions (HEIs) with around 120,000 students, producing 35,000 graduates each year.

There is great quality in the sector as well as scale: more than a third (36%) of the research in our universities is world-class, with 10% world-leading; and City Region universities are in the top five of the research rankings for six of the government's 'eight great technologies': big data, robotics, advanced materials, agri-science, regenerative medicine and satellites²⁴.

²³ LCR SEP Part A, page 36

²⁴ LCR SEP Part A, page 37

Section E: Recommendations and next steps

The Yorkshire Universities have strength and critical mass particularly in relation to applied data analytics. This includes major Research Council Centres in both Medical Bioinformatics and Consumer & Behavioural research. Substantive applications to medicine and health are a particular opportunity given the intersection with importance of both government and private sector interests in the local area. The advanced status of the Leeds Care Record and the size and diversity of the population are also sources of comparative advantage in global research terms and Yorkshire is also a centre for financial services and digital media. In order to harness the substantial research capability and expertise that exists within the region in relation to Big Data platforms and the application of data analytics in particular, the report makes the following recommendations:

1. Future strategies should seek to consolidate expertise through a process of data linkage across both domains and disciplines. Our Data Mutual is a nascent project which seeks to break down both the social and institutional barriers to data access and linkage and is an example of the kind of mechanism which could promote the underlying data resources. The Leeds Institute is already established as a basis for multi-disciplinary collaboration which could be scaled up through replication e.g. on a hub-and-spoke basis.
2. The region has a sound infrastructure basis for big data analytics, for example in the University sector as a hub for the N8 High Performance Computing initiative (n8hpc.org.uk), and in the commercial world in hosting IX3 as the only peer-to-peer internet exchange outside London²⁵. Future development of this infrastructure will require both security and agility to anticipate changing needs. One element of the strategy here could be the active engagement of small and medium sized enterprises in the research and development of big data tools and techniques. An exemplar has already been provided here by the promotion of ResearchOne as a shared patient database which is accessible to many partners through the Yorkshire and Humberside Academic Health Services Network.
3. The Universities in the area are generating a substantial cohort of graduates in Computer Science. In order to promote data analytics across the various communities the skills base needs to be extended. This process could be fostered by the creation of new courses with a specific emphasis on data analytics, and which might involve external partners across the full range of activities from design (e.g. advisory boards and consultation) to delivery (e.g. shared projects or even involvement in teaching). Flexible methods of study and accreditation might also be considered.
4. Universities are also in an ideal position to broker connections between private and public organisations and could provide networking opportunities, perhaps working alongside an organisation such as the ODI. Fellowship programmes to underpin the flow of qualified personnel between the Universities and the outside world (in both directions) could be considered, as well as the potential for co-location of staff within centres or institutes with the capacity and infrastructure to support activities of this type.

²⁵<https://www.aql.com/news/275/>