

Science Landscape Seminar Series: Representative Big Data, Supercomputing and E-Infrastructure

Notes to reader

This document is to inform discussion only and is subject to the following caveats:

- Inclusion, non-inclusion or otherwise is not intended to reflect on the standing of any organisation or infrastructure.
- We did not include classified defence and intelligence assets.
- Assignment of subjects, Eight Great Technologies and Industrial Strategy sectors was based on desk research and may be subject to error. Categories are designed to inform the general discussion and not reflect on individual organisations or infrastructure. Absence of icons in category tables indicates that (in the project team's opinion) a piece of infrastructure cannot be easily categorised.
- This is an internally generated discussion document.
- If any mistakes have been made, please inform the seminar secretariat.

Definition of Big Data, Supercomputing and E-Infrastructure


To frame the group's discussion the GO-Science team has used the following definition of Big Data, Supercomputing and E-Infrastructure in creating this document:

Big Data, Supercomputing and E-Infrastructure is used here in the broadest sense to mean large volumes of data and the methods used to understand them, whether that is a computing facility, a network, or the institutes which house such research.

Included are:

- Research institutions that
 - specialise in data science, informatics and analytics
 - specialise in data privacy and cyber security research
 - generate and/or use big data and supercomputing in their work, such as in the social sciences, genomics and climate modelling
- Large datasets and the institutions or facilities which collate, store and distribute these datasets
- Institutions that develop software, data standards, algorithms and models
- High Performance Computing (HPC) and cloud computing facilities
- Networks and data infrastructure
- Centres and virtual institutes with expertise in the areas listed above

Logo Key

Location		Research Area	Eight Great Technologies		Industrial Strategy
	UK	 Arts and humanities		Energy Storage	 Life Science
	EU	 Biological and medical sciences		Big Data	 Aerospace
	Global	 Earth sciences		Satellites	 Professional Business Services
		 Engineering		Robotics and Autonomous Systems	 Education
Funding					
	Research Councils	 Physical, mathematical and computer sciences		Synthetic Biology	 Nuclear
	Departmental	 Social and economic sciences		Regenerative Medicine	 Oil and Gas
	Private Sector			Agri-Science	 Automotive
	Charity			Advanced Materials	 Offshore Wind
	Academic			Quantum Technologies	 Information Economy
	European				 Construction
					 Agri-tech

List of Infrastructure

The infrastructure identified has been categorised in to the lists below. There will be instances when items could fit in to multiple lists and in these cases we have tried to place the infrastructure in the most appropriate list.

Research Centres and Institutes: Informatics and analytics

The Alan Turing Institute (not yet operational)
Horizon Digital Economy Research
Imperial College London, Data Science Institute
Mathematical and Statistical Research Centres
University of Oxford (The Oxford Internet Institute and Oxford e-Research Centre)
UCL Big Data Institute
Web Science Institute, University of Southampton

Data Privacy and Cyber Security Research

Academic Centres of Excellence in Cyber Security Research, ACEs-CSR
The Research Institute in Science of Cyber Security (RISCS)

Research Centres and Institutes: Applied Big Data Research

Big Data Institute, University of Oxford (not yet operational)
Centre for Agricultural Informatics and Sustainability Metrics (not yet operational)
Centre for Doctoral Training in Financial Computing & Analytics
European Centre for Medium range Weather Forecasts
European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI)
The Farr Institute of Health Informatics Research
Health e-Research Centre (HeRC)
i-sense, EPSRC IRC in Early-Warning Sensing Systems for Infectious Diseases
ISER Institute for Social and Economic Research
The Met Office and the Met Office Hadley Centre
UBDC Urban Big Data Centre, University of Glasgow
The Wellcome Trust Sanger Institute

Government and Public Bodies

The Atomic Weapons Establishment (AWE)
The Catapult Programme
The Data Lab
Defence Science and Technology Laboratory (DSTL)
Government Communications Headquarters (GCHQ)
Northern Ireland Statistics and Research Agency (NISRA)
Office of National Statistics (ONS)
Ordnance Survey (OS)
Public Data Group
Research Councils UK (RCUK)

Supercomputing Facilities

CFMS
Centre for Innovation and CORE, Science and Engineering South (SES)
DiRAC
EPCC
HPC Midlands
HPC Wales
The Hartree Centre
JASMIN
Met Office Supercomputing Facility, Data Archive and MONSoon

MidPlus
N8 High Performance Computer (Polaris)
Partnership for Advanced Computing in Europe (PRACE)

Networks and Data Infrastructure

British Isles continuous GNSS facility
ESA European Centre for Space Applications and Telecommunications (ECSAT)
EUDAT European Data Infrastructure
European Grid Infrastructure EGI
Galileo (GPS)
Global Earth Observation System of Systems (GEOSS)
GridPP
InfraStructure for European Network for Earth System Modelling (IS-ENES) data network
Janet
Jisc

Algorithms, Software, Data Standards and Centres of Expertise

Advanced Computing Research Centre ACRC
International Internet Preservation Consortium (IIPC)
Met Office Unified Model (MetUM)
Numerical Algorithms Group (NAG)
Data standards organisations
Open Data Institute
Research Data Alliance RDA
Software Sustainability Institute
UK Collaborative Computational Projects CCPs

Datasets and Data Services

Administrative Data Research Network (ADRN)
UK Biobank
British Library
Care.data
Cendari
CESSDA
The Clinical Practice Research Datalink (CPRD)
Collections
Data.ac.uk including equipment.data
DATA.GOV.UK
ELIXIR
ESRC Big Data Network
The Genome Analysis Centre (TGAC)
Genomics England
Health and Social Care Information Centre (HSCIC)
Mimas
Museum collections
The National Archives
National Centre for Earth Observation (NCEO)
NERC data centres
NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS)
Social Science Datasets
UK Data Archive
UK Data Service
Virtual Microdata Laboratory
UK Universities Data Repositories

Private Sector

BT Research (Adastral Park)

HP Autonomy

IBM Research

Intel, including Intel Collaborative Research Institute for Sustainable Connected Cities (ICRI)

Microsoft Research Cambridge

Private Sector: 'Internet Giants'

Private Sector: Remote computing services and cloud computing

Private Sector: Internet Service Providers (ISPs) and Mobile Network Operators

Private Sector: 'Tech Giants'

Private Sector: Polling, Surveying and Market Research

Private Sector: Academic publishers

Private Sector: SMEs and Start-ups

Research Centres and Institutes: Informatics, Analytics and Centres of eExpertise

This category covers research institutes which conduct fundamental research in data science, informatics, computer science and the internet, or carry out social science research related to big data.

The Alan Turing Institute (not yet operational)



The Alan Turing Institute will provide a centre to promote advanced research and translational work in the application of data science and the associated computational algorithms. The UK government has committed £42m through EPSRC over the next 5 years to help fund the institute, which will involve a consortium of UK universities. [More info:

<http://www.epsrc.ac.uk/funding/calls/turinginstituteeoinotice>]








Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Horizon Digital Economy Research



Horizon Digital Economy Research was established in 2009 by RCUK, the University of Nottingham and over 100 academic and industrial partners; Horizon research focuses on the role of ubiquitous computing technology. It is both a doctoral training








centre and a research hub. Horizon's aim is to investigate the technical developments needed if electronic information is to be controlled, managed and harnessed for social benefit. [More info: <http://www.horizon.ac.uk/About-us>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Academic	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

Imperial College London, Data Science Institute

Imperial College London

The Data Science Institute acts as a focal point for coordinating data science research at Imperial College. Research focuses on cross-cutting foundations of data science, including statistics, big data, machine learning, modelling, simulation, visualisation and cloud computing. The Institute supports data-driven research in various application domains: astrophysics; particle physics; biology; meteorology; medicine; finance; and healthcare social sciences, amongst others. The Data Science Institute is organized around three types of labs: virtual research labs, joint academia-industry labs and joint academic labs. [More info: <http://www3.imperial.ac.uk/data-science>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic	 Biological Sciences  Physical, mathematical and computer sciences  Social and economic sciences	 Big Data	 Information Economy

Mathematical and statistical research centres

Mathematical research centres have several roles to play in big data research. They have a vital function in providing technical skills training required for data analytics. Additionally, they often perform fundamental and/or applied research, where large datasets are interrogated using novel mathematical and statistical techniques. Examples of institutes performing the latter include:

- Isaac Newton Institute for Mathematical Sciences: National research facility located in Cambridge, where relevant research programmes include statistical network analysis, infectious disease dynamics and statistical theory and methods for complex, high-dimensional data [<http://www.newton.ac.uk/>]
- Turing Gateway to Mathematics: This initiative of the Isaac Newton Institute stimulates the interchange of knowledge and ideas between academics of different disciplines and users of modern mathematics, with research into topics including cyber-security, data linkage and anonymisation, and 'Coping with Big Data - an Analytics and Computational Perspective' [<http://www.turing-gateway.cam.ac.uk/>]
- International Centre for Mathematical Sciences: National research facility located in Edinburgh. Research includes mathematics for health and disease. [<http://www.icms.org.uk/>]
- Heilbronn Institute for Mathematics Research: Centres in Bristol and London, in academic partnership with GCHQ. Research areas include data science, quantum information and dynamic networks and cyber-security

University of Oxford

Oxford Internet Institute, University of Oxford



The Oxford Internet Institute is a multi-disciplinary centre founded in 2001 at the University of Oxford, for the study of the societal implications of the Internet. The Institute looks at this subject with the aim of shaping internet research, policy and practice. It is the UK member of the World Internet Project. [More info: <http://www.oii.ox.ac.uk/about/>]

Oxford e-Research Centre



The University of Oxford e-Research Centre provides an environment that enables collaborative research and provides innovative technologies to academic and commercial partnerships. Over 50 multidisciplinary researchers work in a wide variety of disciplines. The Centre provides research & development computing systems (Microsoft Cluster, Cloud service & GPU Cluster), distributed computing (OxGrid & Volunteer Computing) and high-performance computing, through ARC, the

University's Advanced Research Computing Facility. [More info: <http://www.oerc.ox.ac.uk/about-us>]








Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Academic	 Social and economic sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

UCL Big Data Institute



UCL and Elsevier announced the launch of the UCL Big Data Institute in December 2013; a collaboration to empower researchers to explore innovative ways to apply new technologies and analytics to scholarly content and data. The centre will be based at Mendeley's headquarters in East London. [More info:








http://www.ucl.ac.uk/news/news-articles/1213/UCL_Elsevier_partnership_181213 and <http://www.elsevier.com/connect/university-college-london-and-elsevier-launch-ucl-big-data-institute>]

Location	Funding	Research area	Eight Great	Industrial Strategy		
 UK	 Academic	 Private Sector	 Social and economic sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Web Science Institute, University of Southampton



The interdisciplinary Web Science Institute (WSI) investigates how the World Wide Web is changing the world, and how the world changes the web. The WSI coordinates and puts into practice education, research and enterprise initiatives on web-related developments at the University of Southampton. [More info: <http://www.southampton.ac.uk/webscience/>]

Location	Funding	Research area	Eight Great	Industrial Strategy		
 UK	 Academic	 Research Councils	 Social and economic sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Data privacy and Cyber Security Research







As the rate of data collection and variety of data sources on individuals increases, so does the profile of data privacy and security issues. Institutions in this category perform research in this field.

Academic Centres of Excellence in Cyber Security Research (ACEs-CSR)

EPSRC and GCHQ have set up a scheme to recognise ACEs-CSR until June 2017. They receive an EPSRC a grant in support of their research, sponsored by BIS, CPNI, GCHQ, OCSIA and RCUK. The scheme aims to enhance the quality and scale of academic cyber security research and postgraduate training being undertaken in the UK. 11 ACEs-CSR exist to date. These are:

- University of Bristol
- Imperial College London
- Lancaster University
- University College London
- University of Oxford
- Queen's University Belfast
- Royal Holloway, University of London
- University of Southampton
- University of Birmingham
- University of Cambridge
- Newcastle University

[More info: <http://www.epsrc.ac.uk/funding/calls/acecsr3rdcall/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Physical, mathematical and computer sciences	 Big Data	 Information Economy



The Research Institute in Science of Cyber Security (RISCS)



RISCS is an academic institute that focuses on understanding the overall security of organisations, including their constituent technology, people and processes. It aims to develop cyber security as a science by providing an evidence base. The institute, funded by a £3.8 million grant, is a virtual collaboration between researchers at:

- Imperial College London working with Queen Mary University of London and Royal Holloway, University of London on Games and Abstraction;
- Newcastle University working with Northumbria University working on Choice Architecture;
- Royal Holloway, University of London working on Cyber Security Cartographies; and
- University College London working on Productive Security.

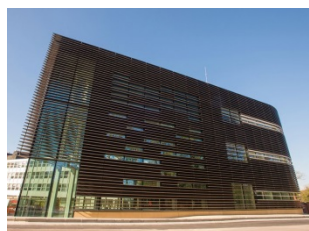
[More info: <http://www.riscs.org.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Research Centres and Institutes: Applied Big Data Research









This category includes data-intensive research which relies on large datasets and high computing power for scientific discovery, without doing fundamental research on the nature of data itself. Examples here include climate, social science and medical (e.g. genomics) research. This list excludes university departments: to be included, institutions must be stand-alone and be partially funded through research councils or other external sources.

Big Data Institute, University of Oxford (not yet operational)



The Big Data Institute is the second phase of development of the Li Ka Shing Centre for Health Information and Discovery, University of Oxford. This is a £90m initiative to put Oxford at the forefront of a healthcare revolution through the use of big data in genetics, epidemiology and human disease, using NHS patient records, DNA sequencing, clinical trials and national registries. [More info:

<http://www.timeshighereducation.co.uk/news/oxford-big-data-centre-to-get-30-million/2003661.article>]


Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic  Research Councils  Private Sector	 Biological and medical sciences	 Regenerative Medicine  Big Data	 Life Science

Centre for Agricultural Informatics and Sustainability Metrics (not yet operational)



£90m will be spent creating Centres for Agricultural Innovation, which will be co-funded by industry. Work on the hub for these centres, the Centre for Agricultural Informatics and Sustainability Metrics, is currently underway, and will be backed by £10m from government and £10m from industry. It will become a repository for, and a single point of access to, relevant data sets, as well as being a centre of excellence for sustainability metrics and big data mining, integration and analysis. [More info:








<http://www.agritechuk.org/Centres>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental  Private Sector	 Biological and medical sciences	 Agri-Science  Big Data	 Agri-tech

Centre for Doctoral Training in Financial Computing & Analytics, UCL



The Centre for Doctoral Training in Financial Computing and Analytics was established at University College London (UCL) in collaboration with academic partners the London School of Economics (LSE) and Imperial College London (IC), supported by partnerships with twenty leading financial institutions. It is the first major collaboration between the financial services, industry and academia. The CDT engages academic advisors in the field of Financial Computing and Analytics. Financial Computing and Analytics encompasses a wide range of research areas including Mathematical Modelling in finance, Computational Finance, Financial IT, Quantitative Risk Management and Financial Engineering. [More info: <http://www.financialcomputing.org/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic  Private Sector	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

European Centre for Medium Range Weather Forecasts (ECMWF)











ECMWF is both a research institute and a 24/7 operational service, producing and disseminating numerical weather predictions to its Member States. This data is fully available to the national meteorological services in the Member States. The Centre also offers a catalogue of forecast data that can be purchased by businesses worldwide and other commercial customers. The supercomputer facility at ECMWF is one of the largest of its type in Europe and Member States can use 25% of its capacity for their own purposes. The centre employs around 270 staff from more than 30 countries. [More info: <http://www.ecmwf.int/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 European	 Earth sciences  Physical, mathematical and computer sciences	 Big Data	

European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI)











EBI is part of European Molecular Biology Laboratory (EMBL) and was formed in 1992. Its Bioinformatics Service gathers collections of biological and chemical information, including EMBL-Bank (DNA and RNA sequences), Ensembl (genomes) and UniProt (protein sequences). It is based at the Wellcome Trust Genome Campus in Hinxton, Cambridge and funded by the 20 member states of EMBL. [More info: www.ebi.ac.uk/about]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental  European  Research Councils  Private Sector	 Biological and medical sciences	 Big Data	 Life Science

The Farr Institute of Health Informatics Research











The Farr Institute comprises four nodes (e-health informatics research centres) distributed across the UK (UCL, University of Manchester, Swansea University and the University of Dundee) which bring together researchers, clinicians and those with an interest in e-health records research to foster collaborations and establish a centre of excellence in innovative health informatics. It is funded by a 10-funder consortium and MRC. Areas of research include drug safety and disease treatment. [More info: <http://www.farrinstitute.org/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic  Private Sector  Departmental	 Biological and medical sciences	 Regenerative Medicine  Big Data	 Life Science

Health e-Research Centre (HeRC)









HeRC is a regional health informatics research centre feeding e-health applications into the [Northern Health Science Alliance](#). It brings together different disciplines to advance data-intensive approaches to health science and care through health informatics. These follow a hub-and-spoke model, where Manchester University hosts HeRC as an expansion of the Northwest Institute for Bio-Health Informatics, with spokes at the Universities of Lancaster, Liverpool and York, and NHS partners across North England. 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. [More info: <http://www.herc.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic  Research Councils  Departmental	 Biological and medical sciences	 Regenerative Medicine  Big Data	 Life Science

i-sense: EPSRC IRC in Early-Warning Sensing Systems for Infectious Diseases






i-sense is a five year £11m EPSRC-funded programme led by UCL and launched in October 2013, with an aim to engineer a new generation of early-warning sensing systems to identify disease outbreaks much earlier than before, using self-reported symptoms on the web and mobile phone-connected diagnostic tests. The capability to detect infections and then wirelessly connect test results to healthcare systems will help patients gain faster access to treatment, and support public health efforts to map indicators of emerging infections in real-time. It also uses the vast amounts of data from Google, Twitter and Facebook to identify indicators of disease outbreaks. [More info: <http://www.i-sense.org.uk/about-i-sense>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental  Academic	 Biological and medical sciences  Physical, mathematical and computer sciences	 Big Data	 Life Science

ISER Institute for Social and Economic Research



ISER was originally established in 1989 at the University of Essex to house the British Household Panel Survey and has grown in to a centre for the production and analysis of longitudinal studies. It offers postgraduate study opportunities, houses a microsimulation unit and runs the tax and benefit model EUROMOD. [More info: <https://www.iser.essex.ac.uk/about>]






Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic  Departmental	 Social and economic sciences	 Big Data	

The Met Office and the Met Office Hadley Centre



A Trading Fund within the Department for Business, Innovation and Skills, which operates on a commercial basis. It employs more than 1,800 people at 60 locations throughout the world. The Met Office is recognised as one of the world's most accurate forecasters, using more than 10 million weather observations a day, an

advanced atmospheric model and high performance supercomputing to create 3,000 tailored forecasts and briefings a day. The Met Office acts as one of six major nodes on the WMO Global Telecommunications Network, processing 106 million observations a day. The Hadley Centre performs climate modelling and researches the attribution and impacts of climate change and is based at the Met Office headquarters in Exeter. [More info: <http://www.metoffice.gov.uk/>]








Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental  Private Sector	 Earth sciences  Physical, mathematical and computer sciences		

Urban Big Data Centre (UBDC), University of Glasgow



The UBDC was established by ESRC to address social, economic and environmental challenges facing cities. It brings together urban social scientists and data scientists from the University of Glasgow and six partner universities, with 56 core members of staff.

Research at UBDC includes methodology in big data management and urban indicators for policymaking. [More info: <http://ubdc.ac.uk/about/overview/>]







Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Academic	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

The Wellcome Trust Sanger Institute



The Sanger Institute is a charitably funded genomics research institute located in Hinxton near Cambridge. It is home to 900 staff and an array of software, databases and expertise. Significant genome sequencing capability is held here; the Sanger made the single largest contribution to the Human Genome Project. Around

£70m is received per year from the Wellcome Trust, accounting for around 85% of the Institute's total funding. [More info: www.sanger.ac.uk/about/]

Location	Funding	Research area	Eight Great		Industrial Strategy
 UK	 Charity	 Biological Sciences	 Regenerative Medicine	 Synthetic Biology	 Life Sciences

Government and Public Bodies

The following government and public bodies play a significant role in big data and supercomputing research, either fundamental or applied. Some have HPC capacity, skills capacity, or enable academia or industry to perform research or innovate.

The Atomic Weapons Establishment (AWE)



AWE holds the UK technical capability in the design, manufacture and support of the UK's nuclear deterrent; Trident. The facility and staff expertise are unique in this regard. Although AWE has two sites, the majority of its 4,500 staff are based in at RAF Aldermaston. AWE hosts several supercomputers, including two of the UK top ten. (www.awe.co.uk/careers/Careers_AWE_Ofbdf.html)

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Physical Sciences	 Engineering	 Aerospace

The Catapult Programme

CATAPULT Catapults are technology and innovation centres where the UK's businesses, scientists and engineers can work side by side on research and development, transforming ideas into new products and services to generate economic growth. Catapults add an important new dimension to complement existing research and development programmes established by Innovate UK. Several catapults make a contribution to Big Data expertise and/or hardware. These include:

- Connected Digital Economy Catapult, including nodes in Sunderland, Brighton and Bradford which will collaborate on projects from March 2015
- Future Cities Catapult
- Satellite Applications Catapult
- Transport Systems Catapult

[More info: <https://www.catapult.org.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Physical, mathematical and computer sciences		

The Data Lab








The Data Lab is industry-led and was set up with an £11.3m grant from the Scottish Funding Council. It has recently offered funding opportunities of up to £50,000 for projects which demonstrate collaboration between Scottish universities and companies (or public sector bodies). [More info: <http://www.thedatalab.com/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Council	 Physical, mathematical and computer sciences	 Big Data	

Defence Science and Technology Laboratory (DSTL)



DSTL is a Trading Fund of the Ministry of Defence which provides government with research and technical support for UK defence and security. DSTL research is wide-ranging, including for example big data, medical and forensic sciences with defence applications, as well as defence technology. It has four bases, with headquarters at Porton Down, Wiltshire, and employs around 3,300 people. [More info: www.dstl.gov.uk]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Physical, mathematical and computer sciences Engineering	 Advanced Materials Big Data Automation and Robotics Satellites	 Aerospace

Government Communications Headquarters (GCHQ)



GCHQ is the government intelligence and security organisation. Its heritage can be traced back to signals intelligence at Bletchley Park in WWII. GCHQ addresses threats from terrorism, the spread of nuclear weapons and conflict resolution. Within GCHQ, CESG provides policy and assistance on the security of communications and electronic data, working in partnership with industry and academia. [More info: <http://www.gchq.gov.uk/Pages/homepage.aspx>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Social and economic sciences Physical, mathematical and computer sciences	 Big Data	 Information Economy

Northern Ireland Statistics and Research Agency (NISRA)



NISRA is the principle source of official statistics and social research on Northern Ireland and an agency of the Department of Finance and Personnel. These statistics and research inform public policy and associated debate in the wider society. It provides registration, statistics and research service, handles census data amongst others and hosts the Central Survey Unit. [More info: <http://www.nisra.gov.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Departmental	 Social and economic sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Office of National Statistics (ONS)



ONS is the UK's largest independent producer of official statistics and is the recognised national statistical institute for the UK. It is responsible for collecting and publishing statistics related to the economy, population and society at national, regional and local levels. It also conducts the Census in England and Wales every ten years. [More info: <https://www.ncas.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Departmental	 Social and economic sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Ordnance Survey (OS)



Ordnance Survey is the national mapping agency for Great Britain, and provides up to date geographic information to government, businesses and individuals. OS is one of the world's largest producers of maps. It is a trading fund of the government. It has a research department that is active in the area of geographical information science, and engages in collaborative research with academia. OS uses remote sensed data and other data sources. [More info: <http://www.ordnancesurvey.co.uk/about/>]








Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Departmental	 Earth sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Public Data Group

The Public Data Group (PDG) brings together 4 organisations (Companies House, Land Registry, Met Office, Ordnance Survey) that excel in the collection, management and distribution of vital data sets, under one banner.

As well as fulfilling their own core public service roles, members of the PDG drive growth in the economy and enable the improvement of services across the public sector by:

- making high quality and reliable data available for national and international businesses and the public sector
- supporting the development of UK SMEs and other developers through effectively communicating the availability and accessibility of data and providing support and advice on how to use it
- driving efficiencies and digitisation and supporting the public sector to deliver more efficient public services
- providing high profile and nationally important data sets as Open Data

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Earth sciences  Physical, mathematical and computer sciences  Social and economic sciences	 Big Data	 Information Economy



Research Councils UK (RCUK)



RCUK is the strategic partnership between the UK's seven research councils, which (between them) fund around £3 billion in research covering the full spectrum of academic disciplines. Big data research is interdisciplinary in nature; therefore, all research councils play an important role in supporting big data research financially and enabling it by providing many of the facilities listed in this document. The research councils are:

- Arts & Humanities Research Council (AHRC)
- Biotechnology & Biological Sciences Research Council (BBSRC)
- Engineering & Physical Sciences Research Council (EPSRC)
- Economic & Social Research Council (ESRC)
- Medical Research Council (MRC)
- Natural Environment Research Council (NERC)
- Science and Technology Facilities Council (STFC)

[More info: <http://www.rcuk.ac.uk/research/areas/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils	All	All	






Supercomputing Facilities

These High Performance Computing (HPC) facilities are among the top in the country. Unlike, for example, the supercomputers hosted by AWE, the ones listed here are available for use by researchers in academia or the public or private sectors. Large and complex datasets require powerful computing facilities to extract their full value. Computing technology is constantly improving and computing infrastructure has a short lifespan before it becomes obsolete, or at least is no longer at the cutting edge of research. This category is rapidly changing as academic research exploits new forms of scientific computing, for example, commercial cloud services.

CFMS



CFMS is an independent centre for modelling and simulation based at the Bristol & Bath Science Park. It enables companies of all sizes to carry out modelling and simulation in standalone and collaborative projects. The centre's core services include high performance computing, secure infrastructure and expertise. [More info: <http://www.cfms.org.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Private Sector	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Centre for Innovation and CORE, Science and Engineering South (SES)



SES is a consortium of research intensive universities in the UK (universities of Oxford, Cambridge and Southampton, Imperial College and UCL). Amongst other things, they share e-infrastructure, operating under two consortia within SES:

- The Centre for Innovation (Oxford, UCL, Southampton, Bristol and STFC):
 - **Iridis 4** CPU system, a HPC facility hosted by Southampton
 - **Emerald** GPU system, a HPC facility hosted by STFC Rutherford Appleton Laboratory (RAL)
- **CORE** (Cambridge and Imperial), provides flexible HPC and data analytics capability. The Cambridge facility is part of both CORE and DiRAC.

[More info: <http://www.ses.ac.uk/e-infrastructure/>]

Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Research Councils	 Academic	 Physical, mathematical and computer sciences	 Big Data	 Information Economy







DiRAC



DiRAC is an STFC facility established in 2009. It is a distributed HPC service, providing a variety of computer architectures for users, matching machine architecture to the algorithm design and requirements of the research problems to be solved. It is a national facility available to both academic and non-academic users. The 5 system architectures are:

- Cambridge HPC service: Data analytic cluster
- Cambridge COSMOS SHARED MEMORY service
- Durham ICC Service: Data Centric Cluster
- Edinburgh BlueGene/Q
- Leicester IT Services: Complexity Cluster

[More info: <http://www.dirac.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Academic	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

EPCC










EPCC, formerly the Edinburgh Parallel Computing Centre, houses a range of supercomputers and employs 75 staff. It hosts and administers a number of national facilities for use by UK researchers on behalf of various research funding councils. The national facilities are:

- ARCHER (Cray XC30 HPC facility provided by RCUK) [More info: <http://www.archer.ac.uk/>]
- BlueGene/Q (the world's most energy efficient petascale machine, part of DiRAC HPC facility provided by STFC) [More info: <http://www.dirac.ac.uk/>]
- UK Research Data Facility UK-RDF (national data facility provided by RCUK).

EPCC also hosts INDY, a mid-range industry standard HPC cluster for industrial users in science and engineering.

[More info: <http://www.epcc.ed.ac.uk/about>]








Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils	 Earth sciences  Physical, mathematical and computer sciences  Social and economic sciences	 Big Data	 Information Economy

HPC Midlands



HPC Midlands is a centre of excellence providing supercomputing on demand for regional industry and academia, with flexible licensing for commercial software. It is built around Hera, a new supercomputer supported by an

EPSRC grant and based at the Universities of Loughborough and Leicester. [More info: <http://www.hpc-midlands.ac.uk/about/>]







Location	Funding	Research area	Eight Great	Industrial Strategy		
 UK	 Research Councils	 Private Sector	 Academic	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

High Performance Computing (HPC) Wales



High Performance Computing (HPC) Wales is an innovative collaboration between Welsh universities, Government and Fujitsu which provides businesses and researchers with access to world-class, secure and easy to use high performance computing (HPC)

technology. HPC Wales consists of HPC capacity (a pan-Wales distributed network of computer clusters), a HPC institute for research, and a HPC skills academy. [More info: <http://www.hpcwales.co.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy		
 UK	 Research Councils	 Private Sector	 Academic	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

The Hartree Centre

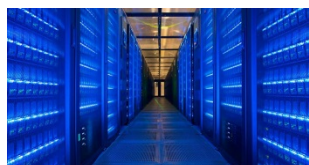


The Hartree Centre is a UK high performance computing (HPC) facility based at the Science and Technology Facilities Council (STFC) Daresbury Laboratory near Warrington. It is a venture by STFC and IBM with a 150-strong scientific computing team. Their HPC equipment includes Blue Joule (BlueGene/Q), Blue Wonder, a world-class iDataPlex cluster comprising over 8000 processor cores ideal for driving optimal value from 'big data', and state-of-







the-art visualisation suites. They focus on many sectors and support industrial, government, academic and research organisations. [More info: <http://www.stfc.ac.uk/Hartree/default.aspx>]

Location	Funding	Research area	Eight Great	Industrial Strategy		
 UK	 Research Councils	 Earth sciences	 Physical, mathematical and computer sciences	 Social and economic sciences	 Big Data	 Information Economy

JASMIN



The JASMIN super-data cluster at the STFC Centre for Environmental Data Archival (CEDA) at the Rutherford Appleton Laboratory is funded by the Natural Environment Research Council (NERC) and the UK Space Agency (UKSA). It is responsible for two main functions: infrastructure for CEDA, including the British Atmospheric Data Centre; and providing a platform for data-intensive scientific computation for environmental researchers across the UK. By the end of phase 4 of its introduction (later this year), JASMIN will offer a 15 petabyte storage infrastructure. [More info: <http://www.jasmin.ac.uk/>]







Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Earth sciences  Physical, mathematical and computer sciences	 Big Data	

Met Office Supercomputing Facility, Data Archive and MONSooN



The Met Office uses an IBM supercomputer which can do more than 1000 trillion calculations a second. Its power allows it to take in hundreds of thousands of weather observations from all over the world which it then takes as a starting point for running an atmospheric model containing more than a million lines of code. MONSooN is a shared service jointly funded by Met Office and

NERC to facilitate collaborative research using the supercomputing capability [More info: <http://www.metoffice.gov.uk/news/in-depth/supercomputers> and <http://www.jwcrp.org.uk/documents/monsoon-leaflet.pdf>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Earth sciences  Physical, mathematical and computer sciences	 Big Data	

MidPlus



MidPlus is a Centre of Excellence for Computational Science, Engineering and Mathematics that was established in 2011 as a joint venture between the Universities of Warwick, Birmingham and Nottingham, and Queen Mary University of London. It provides a collaborative environment that enhances computational research and provides access to high-end e-infrastructure for University, Industrial and Commercial users. It hosts a HPC cluster, a high throughput computer cluster, data storage, archiving and visualisation facilities. [More info: http://www2.warwick.ac.uk/fac/cross_fac/midplus/]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic  Research Councils	 Physical, mathematical and computer sciences	 Big Data	

N8 High Performance Computer (Polaris)



High Performance Computing

The N8 High Performance computer (Polaris) is a shared HPC facility for the 8 northern research-intensive UK universities that form the N8 Research Partnership. N8 HPC currently offers one

facility, a shared 'Tier 2' HPC facility (an SGI High Performance Computing Cluster), underpinning research, enabling collaborations between institutions and creating the opportunity to engage more with business and the community. [More info: <http://n8hpc.org.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils	 Earth sciences  Physical, mathematical and computer sciences  Social and economic sciences	 Big Data	 Information Economy

Partnership for Advanced Computing in Europe (PRACE)



PRACE provides a persistent pan-European Research Infrastructure providing leading HPC resources. It is operated in collaboration with national HPC facilities. PRACE resources can be accessed by scientists and researchers around the world, and industrial users with head offices or substantial R&D activity in Europe. [More info: <http://www.prace-ri.eu/about-prace-ri>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 European  Research Councils	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Networks and Data Infrastructure

This category is for infrastructure used to communicate, transfer, share and enable big data use in research. This includes satellite, fibre optic networks and grid computing, where a grid is a distributed system using a collection of computer resources in multiple locations, creating a 'super virtual computer'. This category also includes organisations who facilitate research networks.

British Isles Continuous GNSS Facility



The NERC funded British Isles Continuous GNSS Facility (BIGF) supports research scientists with archived RINEX format GNSS (Global Navigation Satellite System) data, metadata and derivative products. This facility is hosted at the Nottingham Geospatial Institute. The archive comprises RINEX data from GPS and GLONASS satellites, from a high density network of around 160 continuously recording stations, sited throughout mainland Britain,

Northern Ireland and Ireland. [More info: <http://www.bigf.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Academic	 Earth sciences  Physical, mathematical and computer sciences	 Satellites	

ESA European Centre for Space Applications and Telecommunications (ECSAT), Harwell



The European Space Agency (ESA) studies the Earth's immediate space environment, the solar system and the universe, and develops satellite technologies. The UK hosts the ESA European Centre for Space Applications and Telecommunications (ECSAT) at Harwell. ECSAT will support activities related to telecommunications, integrated applications, climate change, technology and science. Around 100 ESA employees will be based in Harwell by 2015. [More info:

http://www.esa.int/About_Us/Welcome_to_ESA/ECSAT]






Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental  European	 Earth sciences  Physical, mathematical and computer sciences	 Satellites	 Aerospace

EUDAT European Data Infrastructure



EUDAT is a project which aims to contribute to the production of a Collaborative Data Infrastructure. The project's target is to provide a pan-European solution to the challenge of data proliferation in Europe's scientific and research communities, to allow researchers to share data within and between communities. EUDAT comprises 25 European partners, including data centres, technology providers, research communities and funding agencies from 13 countries. [More info:

<http://www.eudat.eu/>]






Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 European	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

European Grid Infrastructure (EGI)



The European Grid Infrastructure (EGI) is a series of efforts to provide access to high-throughput computing resources across Europe using grid computing techniques. It links centres in different European countries to support international research in many disciplines. [More info:


<http://www.egi.eu/about/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 European	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Galileo (not yet operational)

Galileo is a €5 billion global navigation satellite system (GNSS) project currently being built by the EU and ESA to reduce EU dependence on American GPS and Russian Glonass (with which Galileo will be fully compatible). Satellite-based navigation is vital to modern forms of navigation and positioning, and is set to expand with the advent of the Internet of Things.






[More info: <http://www.gsa.europa.eu/galileo/why-galileo>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 European	 Physical, mathematical and computer sciences	 Satellites	 Information Economy

Global Earth Observation System of Systems (GEOSS)








This 'system of systems' will proactively link together existing and planned observing systems around the world and support the development of new systems where gaps currently exist. It will promote common technical standards so that data from the thousands of different instruments can be combined into coherent data sets. It is comprised of 80 governments and the European Commission and a wide range of participating organisations. [More info: <http://www.earthobservations.org/geoss.php>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Multinational	 Earth sciences	 Physical	 Big Data

GridPP



GridPP (UK Computing for Particle Physics) is an STFC-funded collaboration of particle physicists and computer scientists from the UK and CERN. This facility includes a distributed computing grid across the UK for particle physicists. There is currently a working grid across 17 UK institutions contributing the equivalent of almost 40,000 PCs to the worldwide Large Hadron Collider (LHC) computing grid, helping to process the data from the LHC. [More info: <http://www.gridpp.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

InfraStructure for European Network for Earth System Modelling (IS-ENES2) data network






The ENES network gathers together the European modelling community working on understanding and predicting climate variability and change. IS-ENES2 combines expertise in climate modelling, computational science, data management and climate impacts. The INES Consortium comprises 23 partners from 11 countries, including the University of Reading, STFC and the Met Office. [More info: <https://verc.enes.org/ISENES2>]

Location	Funding	Research area	Eight Great	Industrial Strategy		
 EU	 Research Councils	 Departmental	 European	 Earth sciences	 Big Data	 Information Economy

Janet










Primarily funded through BIS and BIS Partner Organisations, Janet is an academic fibre network and the dominant activity of Jisc, formerly the Joint Information Systems Committee. Janet connects Research Councils and all further and higher education institutions in the UK. This enables Jisc to supply a suite of network and IT services such as super-fast broadband, institutional email and electronic access to journals. [More info: www.ja.net/about-Janet/about-us]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

Jisc



Jisc, historically the Joint Information Systems Committee, is a registered charity that provides leadership and support in the use of digital services in UK education and research. It runs the Janet network, the EDINA data centre (including the Digimap Collections), runs a digitisation programme to digitise cultural heritage and scholarly materials, provides advisory and development services, and has various other further and higher education digital support functions. [More info: <http://www.jisc.ac.uk/about>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

Algorithms, Software, Data Standards and Centres of Expertise





This category covers algorithms and software used in big data analytics, which may or may not run on supercomputing facilities. Perhaps more importantly this category includes the centres which host expertise in this field; software can become obsolete, skills to update and create new code are only going to increase in demand. This category also includes institutes promoting or developing data and metadata standards. This category does not include universities, although they too are important for present and future skills capability.

Advanced Computing Research Centre (ACRC)



The Advanced Computing Research Centre was established in 2013 by HEFCE and the University of Sheffield. It is a dedicated industry-facing research centre focused on fast forwarding technological research into business benefit by developing innovative products and services exploiting the knowledge base of the University of Sheffield in the areas of data and text analytics,

complex simulation, virtual reality and quality assurance and testing. Projects include NHS decision supporting tools, financial modelling and a virtual rainforest simulator. [More info: <http://www.acrc.com/>]







Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Private Sector  Academic	Various	 Big Data	 Information Economy

International Internet Preservation Consortium (IIPC)



The IIPC is a membership organisation dedicated to improving the tools, standards and best practices of web archiving while promoting international collaboration and the broad access and use

of web archives for research and cultural heritage. It was formally chartered in 2003 and now has members from over 25 countries, including national, university and regional libraries and archives. [More info: <http://netpreserve.org/about-us>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Private Sector  Departmental	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Met Office Unified Model (MetUM)

The Met Office Unified Model (MetUM) provides seamless prediction systems for operation and research in weather forecasting, climate prediction and climate change projection. The UM system constitutes around 2million lines of code and represents an investment of about £100m in critical research infrastructure. The UM system is currently used by over 1000 scientists in the UK (including 500 at the Met Office). Its future development is supported by the Met Office/NERC Joint Weather and Climate Research Programme (JWCRP). [More info: <http://www.metoffice.gov.uk/research/modelling-systems/unified-model>]






Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental  Research Councils	 Earth sciences  Physical, mathematical and computer sciences	 Big Data	

Numerical Algorithms Group (NAG)



NAG is a British software company which provides methods for the solution of mathematical and statistical problems, and offers services to HPC users. It is a not-for-profit organisation with users from large companies, universities, HPC facilities and independent software developers, with offices in the UK and international

locations. NAG produced their first library in 1971, which is now the largest commercially available collection of high quality mathematical and statistical algorithms, containing more than 1600 routines. They also provide a compiler and consulting and HPC services. [More info: http://www.nag.co.uk/about_nag.asp]

Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Private Sector	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Open Data Institute








The Open Data Institute was set up in 2012 by Tim Berners-Lee and Nigel Shadbolt as a non-profit private company limited by guarantee. It brings together world-class experts to collaborate and promote innovation, with a goal of spreading and developing the principles of open data technology. It offers public training courses and is currently funding two research projects on the publishing of open data and the use of open data. A global network of ODI 'Nodes' are a collaboration of the ODI with similar institutes around the world [More info: <http://theodi.org/>]

Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Departmental	 Private Sector		 Big Data	 Information Economy

Data Standards Organisations

Data standards ensure the quality and interoperability of data, allowing different users on different platforms to make use of it. This can include the standardisation of metadata, format, structure and tagging. Many organisations exist to develop data standards and/or promote their use in open data, making datasets available to researchers. Examples of these include:

- Data Documentation Initiative (DDI): DDI is an initiative to create an international standard for describing data from the social, behavioral, and economic sciences.
- EucoCRIS: EuroCRIS is a European organization responsible for publicising work on Current Research Information System (CRIS) systems.
- Consortia Advancing Standards in Research Administration Information (CASRAI): CASRAI uses standards committees and review circles to develop and maintain data standards for use by researchers and funders.
- International Standards Organisation (ISO)





Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Global	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Research Data Alliance (RDA)



The Research Data Alliance is a social and technical infrastructure to enable open sharing of data. It works through focussed working groups and interest groups formed of experts around the world from academia, industry and government. It was started in 2013 by a core group of interested agencies, including the European Commission, the US National Science Foundation and the

Australian Government's Department of Innovation. It now has 1600 members from 70 countries. [More info: <https://rd-alliance.org/about.html>]





Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Global		 Big Data	 Information Economy

Software Sustainability Institute







The Software Sustainability Institute is a national facility for cultivating world-class research through software. Software sustainability means that current software can be improved and supported in the future. The institute is based at the

Universities of Edinburgh, Manchester, Oxford and Southampton, and draws on a team of experts with a breadth of experience in software development, project and programme management, research facilitation, publicity and community engagement. [More info: <http://www.software.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic	 Physical, mathematical and computer sciences		 Information Economy

UK Collaborative Computational Projects (CCPs)

CCPs provide a software infrastructure on which individual research projects can be built. They support both the R&D and exploitation phases of computational research projects. They develop software which makes optimum use of the whole range of hardware available to the scientific community, from desktop computers to powerful national HPC facilities. The main activities of CCPs are to carry out flagship code development projects and maintain and distribute code libraries. There are currently 14 CCP projects. [More info: <http://www.ccp.ac.uk/about.html>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 European	 Physical, mathematical and computer sciences		 Information Economy

Datasets and Data Services

This broad category contains datasets along with the services that collate, archive and distribute those datasets. Models which rely on big data and algorithms are also included.

Administrative Data Research Network (ADRN)



ADRN is an ESRC-funded partnership between universities, government departments and agencies, national statistics authorities, funders and the wider research community. ADRN makes it possible to carry out social and economic research using administrative data to benefit society. Administrative data collections are built up by government departments and agencies during their day-to-day activities, for example when registering people, carrying out transactions or in record keeping. It will be operational from November

2014. ADRN consists of an Administrative Data Research Centre in England, Wales, Scotland and Northern Ireland, and is coordinated by the Administrative Data Service (ADS) within the UK Data Archive. [More info: <http://www.adrn.ac.uk/about>]

Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Research Council	 Social and economic sciences	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

UK Biobank



UK Biobank is a long-term study into the respective contributors of genetic predisposition and environmental exposure to the development of disease. It recruited 500,000 people aged between 40-69 and will collect health information on these individuals for the next 25 years. The study began in 2006 with initial funding of £62m.

Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Departmental	 Research Councils	 Biological and Medical Sciences	 Big Data	

British Library



The British Library is a major research library, holding around 170 million items, including around 14 million books, and is the largest library in the world by number of items catalogued. As a legal deposit library, it receives copies of all books produced in the UK and Ireland. The British Library is also creating the UK Web Archive, archiving websites from the UK domain for future access. The library has an annual budget of £142m.

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Arts and humanities	 Big Data	

Care.data



NHS England is commissioning a modern data service from the Health and Social Care Information Centre (HSCIC) on behalf of the entire health and social care system. Known as care.data, this programme will build on existing data services and expand them to provide linked data that will eventually cover all care settings, both inside and outside of

hospital. [More info: <http://www.hscic.gov.uk/gpes/caredata>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Biological and medical sciences  Social and economic sciences	 Big Data	

CENDARI



CENDARI is a research infrastructure project aimed at integrating digital archives for the medieval and World War One eras. The project brings information and computer scientists together with leading historians and existing

historical research infrastructures (archives, libraries and other digital projects) to improve the conditions for historical scholarship in Europe through active reflection of, and considered response to, the impact of the digital age on scholarly and archival practice. It is a 4-year, European Commission-funded project led by Trinity College Dublin, in partnership with 14 institutions across 8 countries. [More info: <http://www.cendari.eu/about-cendari/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 Research Councils  Academic  European	 Arts and humanities  Physical, mathematical and computer sciences	 Big Data	

CESSDA



CESSDA was established in 1976 as an umbrella organisation for the European national data services; it became a permanent legal entity in June 2013 and is financed by the governments of member states. It is hosted in Norway and has 13 member countries. The UK Data Archive forms the UK branch. Its major objective is to provide seamless access to data across repositories, nations, languages and research purposes. CESSDA encourages the use of and actively develops standards for data and metadata. [More info: <http://www.cessda.net/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 Research Councils  European		 Big Data	 Information Economy

The Clinical Practice Research Datalink (CPRD)



The Clinical Practice Research Datalink (CPRD) is the new English NHS observational data and interventional research service, jointly funded by the National Institute for Health Research (NIHR) and the Medicines and Healthcare Products Regulatory Agency

(MHRA). CPRD services are designed to maximise the way anonymised NHS clinical data can be linked to enable many types of observational research and deliver research outputs that are beneficial to improving and safeguarding public health. [More info:

<http://www.cprd.com/intro.asp>]





Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental  Research Councils	 Biological and medical sciences	 Big Data	

Data.ac.uk



Data.ac.uk is a site which provides a single point of contact for linked open data development for academia. It provides access to expertise to create linked data and data aggregation sites and also enables access to large aggregated data sets. Data.ac.uk also

informs national standards and assists the development of national data aggregation subdomains. Equipment.data is an EPSRC-funded subdomain of this site. It provides data on the use of UK research equipment and is the RCUK preferred medium for national equipment data sharing [More info: www.data.ac.uk/ and <http://equipment.data.ac.uk/>]







Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils	Various	 Big Data	 Information Economy

DATA.GOV.UK



Data.gov is the portal through which the government releases public data under the Government Open Data licence. Datasets include raw data from all central government departments and a number of other public sector bodies and local authorities. It is a

project created by the Open Knowledge Network. Categories of data includes transport, health, society, mapping, education and economic. As the project matures, more data will be available through the site. [More info: <http://data.gov.uk/about>]








Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Physical, mathematical and computer sciences  Social and economic sciences	 Big Data	 Information Economy

ELIXIR



The goal of ELIXIR is to orchestrate the collection, quality control and archiving of large amounts of biological data produced by life science experiments. Some of these datasets are highly specialised and would previously only have been available to researchers within the country in which they were generated. [More

info: <http://www.elixir-europe.org/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 EU	 European  Research Councils	 Biological Sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

ESRC Big Data Network



In April 2013 the ESRC announced new funding of £64m to be invested in a Big Data Network. The core aim of this network is to facilitate access to different types of data, stimulating new methods to undertake research. The network is supporting the establishment of Local Government Data Research Centres, which will have 5-yearly budgets of up to £7m.








Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Council	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

The Genome Analysis Centre (TGAC)



The Genome Analysis Centre™

located at the Norwich Research Park, is a research centre for data intensive science in biology. It was established as a national facility to promote the use of genomics in advancing bioscience research and innovation in the UK. Its expertise is in high-throughput sequencing, with supercomputing capability. It has sequenced the genome of the fungus responsible for ash dieback, amongst others. It supports academic and industrial investigators with a team of over 70 staff. [More info: <http://www.tgac.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Biological and medical sciences	 Agri-Science  Big Data	 Life Science

Genomics England



Genomics England is sequencing 100,000 genomes from NHS patients, to be completed by 2017. It is a company owned by the Department of Health and was set up to deliver the 100,000 genomes project. It focuses on patients with rare diseases and common cancers and is currently in its pilot phase. The aim of the project is to set up a data service to enable new scientific discovery and insight and to kickstart the development of a UK genomics industry. [More info: <http://www.genomicsengland.co.uk/>]

project is to set up a data service to enable new scientific discovery and insight and to kickstart the development of a UK genomics industry. [More info: <http://www.genomicsengland.co.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Biological and medical sciences	 Regenerative Medicine  Big Data	 Life Science

Health and Social Care Information Centre (HSCIC)



The Health and Social Care Information Centre (HSCIC) was set up as an Executive Non Departmental Public Body (ENDPB) in April 2013. It collects, analyses and presents national health and social care data. Additionally, it has responsibility for setting up and managing national IT systems to collect such data, to reduce paperwork for doctors, and to help health and care organisations improve the quality of data that they collect. [More info: <http://www.hscic.gov.uk/whoweare/>]

it has responsibility for setting up and managing national IT systems to collect such data, to reduce paperwork for doctors, and to help health and care organisations improve the quality of data that they collect. [More info: <http://www.hscic.gov.uk/whoweare/>]


Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Biological and medical sciences  Social and economic sciences	 Big Data	 Life Science

Mimas



Mimas, part of the digital resources division at Jisc, hosts many of the UK's research information assets and builds applications to help people make the most of this resource. As part of the UK Data Service, Mimas provides access to, and specialist support for, the databanks of organisations such as the International Monetary Fund, and also aggregates statistics from the 1971 to 2011 UK censuses. [More info: <http://mimas.ac.uk/>]

and also aggregates statistics from the 1971 to 2011 UK censuses. [More info: <http://mimas.ac.uk/>]


Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

Museum Collections



The UK has irreplaceable collections which are used in research and funded through government departments, such as the British Museum (historical and contemporary material culture), Victoria and Albert Museum (design), National Gallery (fine art) and the Natural History Museum. Collections are often digitised; for example, the Victoria and Albert museum has digitised much of its collection, providing access to over 1.1 million catalogue records and 293,000





images of objects.

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Departmental	 Arts and Humanities  Social Sciences		

The National Archives






The National Archives is the official archive and publisher for the UK Government and for England and Wales, and an executive agency of the Ministry of Justice. Its collection holds over 11 million historical government and public records, including photographs, maps, paintings and websites, some of which date back over 1,000 years.

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Arts and Humanities	 Big Data	

National Centre for Earth Observation (NCEO)



NCEO is a partnership of over 100 scientists from 26 institutions whose mission is to unlock the full potential of Earth observation data. The scientists use data from Earth observation satellites to monitor global and regional changes in the environment, to learn more about the Earth system and to improve predictions of future environmental conditions. NCEO works closely with the UK satellite industry and international space organisations and operates with an annual budget of £6.5m. [More info: <http://www.nceo.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Council	 Earth sciences  Physical Sciences	 Satellites	

NERC Data Centres



NERC has a network of environmental data centres that provide a focal point for its scientific data and information. The NERC Data Catalogue Service provides an integrated, searchable catalogue of the data holdings of the data centres. NERC supports seven data centres covering the following discipline areas:

- Marine - British Oceanographic Data Centre
- Atmospheric - British Atmospheric Data Centre (BADC)
- Earth observation - NERC Earth Observation Data Centre (NEODC)
- Solar and space physics - UK Solar System Data Centre (UKSSDC)
- Terrestrial and freshwater - Environmental Information Data Centre
- Geoscience - National Geoscience Data Centre
- Polar and cryosphere - Polar Data Centre







Location	Funding	Research area	Eight Great	Industrial Strategy	
 UK	 Research Councils	 Earth sciences	 Physical, mathematical and computer sciences	 Big Data	

NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS)



NEODAAS is funded by NERC to support UK environmental scientists with remote sensing data and information. It was formed by the merger of two long established services, the Dundee Satellite Receiving Station, University of Dundee, and the Remote Sensing Data Analysis Service which operates within Plymouth Marine Laboratory. Together they have the capability to automatically receive, archive, process and map global data from

multiple polar-orbiting sensors in near-real time, and also receive and process data from multiple geostationary satellites. [More info: <http://www.neodaas.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy		
 UK	 Research Councils	 Earth sciences	 Physical, mathematical and computer sciences	 Satellites	 Big Data	

Social Science Datasets








The UK has a portfolio of the longest-running longitudinal studies in the world. Apart from the most recently established, these studies are irreplaceable resources to policy making and social research. Three of the largest of these are managed by the Centre for Longitudinal Studies in London (1958 National Child Development Study, 1970 British Cohort Study, Millennium Cohort Study). The Cohort and Longitudinal Studies Enhancement Resources (CLOSER) programme integrates these with an additional six studies, including the current Understanding Society and Life Study. HALCyon brings together interdisciplinary scientists working on nine UK cohort studies, including cohorts from Scotland and Wales. The UK also invests in other types of social science data – for example the European Social Survey is hosted at the Centre for Comparative Social Surveys (CCSS) in London, and the new Administrative Data Research Network (ADRN) integrates data across government departments. [More info: e.g. <http://www.closer.ac.uk/>, www.halcyon.ac.uk/?q=cohorts]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils	 Social and economic sciences	 Big Data	

UK Data Archive



The UK Data Archive curates the largest collection of digital social science data in the United Kingdom. Founded in 1967, the Archive is an acknowledged centre of expertise in the areas of acquiring, curating and providing access to data. In 2005 it was designated a Place of Deposit by the National Archives, which allows the curation of public records. It is largely funded by ESRC, Jisc and the University of Essex. The UK Data Archive is the UK branch of CESSDA. [More info: <http://www.data-archive.ac.uk>]


Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Councils  Academic	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

UK Data Service

UK Data Service





The UK Data Service provides users with access to a wide range of data resources to facilitate social and economic research and education. It archives, provides access and provides data management and support. They aim to provide a single point of access for all data types. It was established in 2012 by ESRC; UK Data Archive, Mimas and several academic departments are amongst its host organisations. [More info: <http://ukdataservice.ac.uk/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Research Council	 Social and economic sciences  Physical Sciences	 Big Data	 Information Economy

UK Universities Data Depositories

UK universities all hold research datasets, literature, library materials and other data in digital archives. Whilst some of this data is hosted and shared through other services listed here, some of it is uniquely stored at the respective university location. Universities often make this data accessible to external users. This resource represents an important repository of historic and recent data.

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Academic	All		

Virtual Microdata Laboratory



The VML (Virtual Microdata Laboratory) is ONS's facility for providing secure access to sensitive detailed data for statistical research purposes that serve the public good. Access is provided for Approved Researchers working on defined and approved projects. Available data sources include business surveys, social surveys and Census samples. [More info: <http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Departmental	 Social and economic sciences  Physical, mathematical and computer sciences	 Big Data	 Information Economy

Private Sector







The private sector is crucial in making big data and supercomputing research a reality, and hundreds of companies are involved in the data revolution. Listed individually are those with a noticeable research presence in the UK, for example, having a research campus or working in collaboration with academia. Other private sector involvement in big data and supercomputing research is listed by category at the end; these are 'internet giants', 'tech giants', internet and mobile service providers, polling and market research companies, and big data SMEs and start-ups; all of these sectors play an important role in the big data revolution. Although private sector data and facilities are not always openly available to the academic research community, it has an important role to play in nurturing UK skills.

BT Research (Adastral Park)



Undertaking research in internet, telecoms and networks, BT's main research capability is based at Adastral Park, near Ipswich. Employing around 4,000 R&D staff, the research park supports UK capability in cyber and comms resilience and is an important component of the UK's technological advantage in the telecoms sector. They spent €811m on research in 2012, 3.8% of revenue. [More info:






www.adastralvision.com/News/BT_SETS_ITS_SIGHTS_ON_BRIGHT_FUTURE_FOR_ADASTRAL_PARK.html & EU R&D Scoreboard 2013]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Private Sector	 Physical Sciences  Engineering	 Big Data	 Information Economy

HP Autonomy








HP Autonomy, formerly Autonomy Corporation PLC, is a multinational enterprise specialising in software that processes human information, or unstructured data, from the internet. Its product portfolio helps companies with search analytics, business process management and OEM operations. It was founded in Cambridge in 1996. [More info: <http://www.autonomy.com/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Private Sector	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

IBM Research








IBM Research was established in 1945 and is the research and development branch of IBM (which spends 6% of its revenue on R&D), consisting of 12 laboratories worldwide. IBM Research's global network of scientists work on a range of exploratory research projects to advance the capabilities of technology, as well as applied research projects such as IBM's healthcare initiative. High profile developments include the relational database, ATM, DRAM and Watson, the artificial intelligence computer. IBM also builds the Blue Gene supercomputer. [More data: <http://www.research.ibm.com/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Private Sector	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Intel, including Intel Collaborative Research Institute for Sustainable Connected Cities (ICRI)







Intel Corporation is an American multinational corporation. It is one of the world's largest and most highly valued semiconductor chip makers. Intel also designs processors for supercomputers. The ICRI is a collaboration between Intel, UCL and Imperial College. It is concerned with how to enable future cities to become more connected and sustainable, using an interdisciplinary research to drive new frontiers in the application of computing technologies to advance the social, economic and environmental well-being of cities. [More info: <http://www.cities.io/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 Global	 Private Sector	 Physical, mathematical and computer sciences	 Big Data	 Information Economy

Microsoft Research Cambridge



Microsoft Research Cambridge was set up in July 1997 with three researchers. Today it has over 100 researchers, who are mostly from Europe, engaged in computer science research at the lab. It conducts basic computer science research on a wide variety of topics, including machine learning, security, and information retrieval, and maintains close ties to the University of Cambridge and the University of Cambridge Computer Laboratory. [More info: <http://research.microsoft.com/en-us/labs/cambridge/>]

Location	Funding	Research area	Eight Great	Industrial Strategy
 UK	 Private Sector	 Physical, mathematical and computer sciences		 Information Economy

Private Sector: Remote Computing Services and Cloud Computing

Remote computing services, also known as web services, are offered over cloud computing platforms on a commercial basis. These services offer the use of remote servers ('infrastructure as a service') and are typically charged according to the required computing power and time, in order to provide computing capacity for companies who lack their own hardware. They are increasingly commonly used, for example by websites and governments, and are powerful computing facilities distinct to HPC. Examples include:

- Amazon Web Services e.g. Amazon EC2 and Amazon S3
- Google Compute Engine
- IBM cloud computing
- Microsoft Azure

Private Sector: 'Internet Giants'

Websites have the ability to collect and store huge volumes of data, with the largest, the 'internet giants' such as Google and Facebook, being well known for collecting disorganised 'big data' such as text, images, videos, searches and location data. Companies engaged in such activities are expected to be at the cutting edge of big data analytics. The 10 most popular websites in the UK as of October 2014 (<http://www.lexiconnect.co.uk/top-20-uk-websites.html>) are as follows:

1. Google
2. Facebook
3. YouTube
4. BBC Online
5. eBay UK
6. Amazon
7. Yahoo
8. Wikipedia
9. LinkedIn
10. Twitter

Private sector: Internet Service Providers (ISPs) and Mobile Network Operators

The internet is arguably the most important infrastructure for big data generation and big data research: it is the platform upon which big data is created and collected. ISPs bring this service to homes and businesses across the UK. Mobile data provision is becoming increasingly more important in generating big data, not only because of the increasing popularity of mobile devices such as smart phones, but also as more everyday objects become connected to the internet with the advent of the Internet of Things.

Private Sector: 'Tech Giants'

The 'tech giants' create the physical technology infrastructure on which big data and supercomputing (or indeed, all computing) ultimately depend on, and so are absolutely vital to the research landscape. These companies develop and manufacture computing hardware, including HPC systems vital for analysing data, software, data storage infrastructure, server and networking solutions, and mobile devices (such as smartphones) which generate much of the world's data. These companies often are multinational with significant activity in the UK and invest heavily in research and development. Examples in this category with UK-based headquarters and/or research and development facilities include:

- Cray Inc. and Cray UK Limited
- Google
- Hitachi
- HP
- IBM
- Intel
- Microsoft
- Oracle
- Samsung (Samsung Electronics Research Institute SERI)
- Sony (SCEE R&D)

Private Sector: Polling, Surveying and Market Research

Another group which collects and analyses large volumes of data in the UK are polling, surveying and market research companies. Examples of UK companies in this category include:

- YouGov
- Ipsos MORI
- Which?
- ComRes
- Populus
- Dunnhumby
- Datamonitor
- Kantar

Private Sector: Academic Publishers

Academic publishers host large volumes of research data through the huge number of academic papers, books and other scholarly content published globally; many of these items present original data, often in the format of a digital data repository. Academic publishers are therefore key in hosting and disseminating data across all research areas. Some publishers carry out research using their data resource (see UCL Big Data Institute). Examples include:

- Elsevier
- Springer
- Thomson Reuters
- Non-profits and University Presses

Private Sector: SMEs and Start-ups

These firms are important to the UK big data and supercomputing research landscape, primarily because they attract a high proportion of workers with specialist skills in computing, data analytics and informatics. Expertise is concentrated in small companies and new start-ups, sometimes partnered with other organisations such as the ODI. Such companies are found across the UK, though many are concentrated in clusters. This sector provides services to industry, government and academia. Smaller companies tend to have less hardware and access to HPC facilities; instead, they may use cloud computing techniques and foster associated skills. Although important as a sector, SMEs and start-ups are small by definition, so are considered vital infrastructure as a group rather than individually. Examples of local initiatives supporting high tech start-up companies, or geographical regions with clusters of such companies, include:

- Tech City UK: an organisation supporting the emerging 'silicon roundabout' tech cluster in East London
- SETsquared Partnership: an enterprise collaboration between the universities of Bath, Bristol, Exeter, Southampton and Surrey.
- Croydon Tech City: This organisation supports early-stage digital and tech startups in Croydon, south London
- Tech North: This tech hub was recently announced by Tech City UK and the deputy prime minister to coordinate existing digital technology expertise in Manchester, Leeds, Sheffield, Liverpool and the North East tech cluster
- 'Silicon Fen' or 'Cambridge cluster' tech cluster in Cambridgeshire: more than 1000 companies with a focus on software, electronics and biotechnology
- Edinburgh TechCube: A 'technology incubator' in Edinburgh supporting Scotland's tech startups

Annex 1: UK top 12 supercomputers as published in the The 43rd TOP500 List, June 23rd, 2014

Global rank	UK rank	Site	System	Vendor	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
19	1	ECMWF	Cray XC30, Intel Xeon E5-2697v2 12C 2.7GHz, Aries interconnect	Cray Inc.	83,160	1,552.00	1,796.30	
20	2	ECMWF	Cray XC30, Intel Xeon E5-2697v2 12C 2.7GHz, Aries interconnect	Cray Inc.	83,160	1,552.00	1,796.30	
23	3	Hartree Centre	Blue Joule - BlueGene/Q, Power BQC 16C 1.60GHz, Custom	IBM	131,072	1,431.10	1,677.70	657
25	4	EPCC	ARCHER - Cray XC30, Intel Xeon E5 v2 12C 2.700GHz, Aries interconnect	Cray Inc.	76,192	1,367.50	1,645.70	
33	5	EPCC	DiRAC - BlueGene/Q, Power BQC 16C 1.60GHz, Custom	IBM	98,304	1,073.30	1,258.30	493.1
40	6	AWE	Spruce A - SGI ICE X, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR	SGI	44,520	958.7	997.2	855.1
49	7	AWE	Spruce B - SGI ICE X, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR	SGI	35,640	767.5	798.3	684.5
59	8	EPCC	HECToR - Cray XE6, Opteron 6276 16C 2.30 GHz, Cray Gemini interconnect	Cray Inc.	90,112	660.2	829	
60	9	ECMWF	Power 775, POWER7 8C 3.836GHz, Custom Interconnect	IBM	24,576	635.1	754.2	1,386.90
61	10	ECMWF	Power 775, POWER7 8C 3.836GHz, Custom Interconnect	IBM	24,576	635.1	754.2	1,386.90
78	11	Met Office	Power 775, POWER7 8C 3.836GHz, Custom Interconnect	IBM	18,432	476.3	565.6	1,040.20
97	12	Met Office	Power 775, POWER7 8C 3.836GHz, Custom Interconnect	IBM	15,360	396.9	471.4	866.8