



Research Councils UK

RCUK and big data

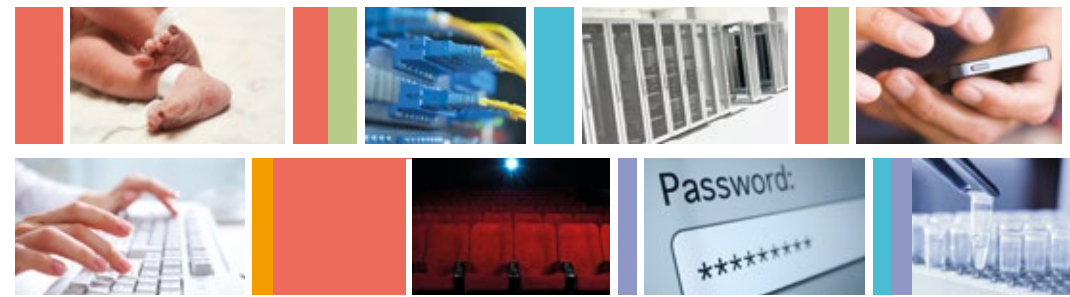


Research funded by the Research Councils makes a vital contribution to the UK's economic growth, prosperity and well-being.

We take a variety of approaches to support innovation and deliver impact from research, including the development of collaborative research programmes, investment in major research capabilities, such as national research facilities, and the support of impact-related capabilities.

Often the impact of research is realised through the combination of several investments over time. The Research Councils seek to ensure that the outputs and outcomes of their funded research have significant long-term benefits for the economy and society. This timeline, one of a series, highlights how investments made in research over the long term combine to create a significant impact in particular areas. In addition, research in one area can combine with that from another to drive innovation and make a key contribution to UK growth. For example, big data has had an impact on almost all sectors and across all research disciplines from as early as the 1960s when Research Council funded researchers developed one of the world's most powerful computers. The UK has also invested in a breadth of longitudinal cohort studies that have helped us understand and track social and medical changes over the decades, right up to today. UK researchers have had a massive impact on the software and world wide web industries whilst big data has also enabled UK researchers to understand the nature of climate change and predict weather more accurately.

A key part of the Government's Industrial Strategy is supporting technologies where the UK has the depth of research, expertise and the business capability to develop and exploit them commercially. Synthetic biology is one of 'Eight Great Technologies' identified by the Chancellor of the Exchequer in autumn 2012 when he announced an additional £600 million to help support their development. These eight are: Big data and energy-efficient computing; Satellites and commercial applications of space; Robotics and autonomous systems; Synthetic biology; Regenerative medicine; Agri-science; Advanced materials and nanotechnology; and Energy and its storage.



RCUK and big data

Big data is a term for a collection of datasets so large and complex that it is beyond the ability of typical database software tools to capture, store, manage, and analyse them. 'Big' is not defined as being larger than a certain number of 'bytes' because as technology advances over time, the size of datasets that qualify as big data will also increase. Research Council funded research over the decades has not only improved the software which analyses the data but has also enabled mass datasets to be captured which helps us understand and interpret our society and world around us.

From the fastest computers being built in the 1960s by Research Council funded researchers at Manchester University that could store both data and programmes, to the Rutherford Appleton lab Tier-1 Grid computing service which sustains record breaking data transfer to CERN, UK researchers have been at the heart of technological developments to sustain and enable big data to develop.

Big data has played an essential part of understanding our society and the world around us. From the 1970s Research Council funded researchers have been using cohort studies to collect data on births and families. The British Cohort Study showed that assessments of school-aged children suggest that their development is not affected by whether or not their mother worked during their first year of life.

Over time the Avon Longitudinal Study of Parents and Children (ALSPAC) and then Research Councils' Millennium Cohort Study have provided a vast amount of data over the years which have assisted scientists across the world with research into health and social problems. More recently, Research Council funded researchers for the Understanding Society project have provided approximately three billion data points of information, and the 2014 Life Study, the UK's largest national birth cohort study so far will gather data on 100,000 babies born in the UK – all of which will help to inform policy-makers policymakers and governments in the years to come.

The UK's Research Councils have played a vital role in the development of big data – and this in turn has had a significant impact on placing the UK as a world leading authority in many areas. From understanding our environment, where the first map of ice thickness was published by the British Antarctic Survey which reduced the global estimates on freshwater by 10 per cent, to enhancing our technology so research into genetic advancements can be made, researchers funded by the Research Councils have played a significant role in worldwide advancements, and will likely do so for years to come.

1980: The UK hosts and funds the European Bioinformatics Institute, which maintains the world's most comprehensive and up-to-date molecular databases, with more than 1 million separate users every month.

1989: The first RCUK - Cognitive Function and Ageing Study (CFAS) is established with the initial aim of investigating dementia and cognitive decline in a representative sample of more than 18,000 people aged over 65 years.

1991: The RCUK Avon Longitudinal Study of Parents and Children (ALSPAC) recruits more than 14,000 mothers in order to follow the health and development of their children.

1997: The Million Women Study, funded by RCUK and Cancer Research UK, is launched. Investigating the effect of Hormone Replacement Therapy (HRT) on women's health, it has now recruited 1.3 million women over the age of 50 and is the largest cohort study in the UK.

2000: The first draft of the human genome is published.

2005: By analysing millions of possible structures, the RCUK e-materials project predicts a previously unknown crystal structure of a molecule of piracetam - a drug for Alzheimer's disease.

2007: The RCUK e-Science centre allows UK scientists to screen a total of 140 million arrangements between drug compounds and target proteins to identify the most promising drug compounds for fighting malaria.

2007: Constellation Technologies is formed using expertise from the UK's particle physics programme and CERN. It provides cloud-based bioinformatics services and products to life science companies.

2008: The combined supercomputing power of the UK and US enables UK scientists to simulate the efficacy of a drug in blocking a key protein on the HIV virus.

2010: UK Biobank completes its recruitment of 500,000 people aged between 40-69 who agree to have their health monitored. This major national health study, partly RCUK-governed, will help scientists discover why some people develop particular diseases and others do not.

2011: The International Mouse Phenotyping Consortium is launched, with the aim of developing an encyclopaedia of mammalian gene function.

2011: The oldest birth cohort study in the UK, the National Survey for Health and Development (NHS) celebrates its 65th anniversary. The NHS has a data archive of over 18,000 variables collected over the lifetime of the study.

2012: 1000 genomics project launched - this is an integrated map of genetic variation from 1092 human genomes drawn from 14 populations around the world, including Europe, the Americas, East Asia and Africa.

2012: The UK government announces funding for a programme to fully sequence the genomes of 100,000 patients with cancer and rare diseases.

2013: 10 year anniversary of Human Genome project. Commercial sequencing of the whole human genome now costs \$3,000 to \$5,000 and takes just one to two days.

2013: Researchers at King's College London are awarded £2.52 million of RCUK funds to build a computational model of the brain networks involved in epilepsy using data from brain scans and electrical recordings of brain waves to understand the mechanisms which result in seizures.

2013: An RCUK-led consortium of 10 UK government and charity funders makes a historic £17.5 million investment to establish four e-health research Centres of Excellence. These Centres (HIRCs) will train researchers to use large and complex sets of patient data and carry out high-quality cutting edge research using patient records.

2013: The GEUVADIS (Genetic European Variation in Health and Disease) project, funded by the European Commission's FP7 programme, publishes the largest-ever dataset linking human genomes to gene activity at the level of RNA.

2013: RCUK-supported researchers in collaboration with Microsoft are using cloud computing to cut the time taken by chemists working on new cancer drugs to process data from five years to ten hours.

2013: Three new RCUK Interdisciplinary Research Collaborations are announced, focusing on research that can create or use new ICT applications and technologies to determine and/or sense physical and environmental factors, and integrate, analyse and interpret this data to inform decisions.

1964: The Biological Records Centre is founded.

1980: The European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Data Library (now known as EMBL-Bank) is established. Funded in part by RCUK, this is the world's first nucleotide sequence database.

1991: The first Samples of Anonymised Records (SAR) is released. SARs are UK datasets consisting of samples of individual records from national censuses.

1994: The European Bioinformatics Institute (EBI), largely funded by RCUK, is established at the Wellcome Trust Genome Campus in Cambridge.

1994: The General Practice Research Database (GPRD) is formed, giving researchers access to NHS information in secure, ethical ways.

2001: RCUK undertakes the 2001 Census of Population to represent the 200th anniversary of census-taking in the UK.

2003: The UniProt database is launched as a collection of information on protein sequences. UniProt is a collaboration between the EBI, Swiss Institute of Bioinformatics, and the Protein Information Resource. UniProt holds information on 41 million proteins including 13 billion amino acid residues.

2004: The National e-Infrastructure Service facilitates UK research by providing computer and data services and becomes the foremost provider of international e-infrastructure for the UK.

2006: The RCUK-funded Data Preservation and Sharing Initiative is launched. This aims to improve accessibility to anonymised scientific data across the research community.

2007: The Organisation for Economic Cooperation and Development (OECD) publishes its 'principles and guidelines for access to research data from public funding' to facilitate cost-effective access to digital research data from public funding. RCUK leads the UK delegation in responding to the guidelines.

2008: The Sequence Read Archive (SRA) is established as a public repository for next-generation sequence data as part of the International Nucleotide Sequence Database Collaboration (INSDC).

2009: RCUK invest £9m in four regional hubs across the UK to enhance access to next-generation sequencing equipment and technical expertise to allow biomedical and clinical researchers to capitalise on high throughput, large-scale genetic analysis.

2009: The RCUK and Met office Monsoon supercomputer launches which provides greater access to atmospheric and meteorological data.

2011: The SRA surpasses 100 terabytes of open-access genetic sequence read from next generation sequencing technologies, and has grown faster than the ability to expand storage capacity.

2011: An RCUK-led mapping exercise on the needs of UK researchers using large and complex health datasets is undertaken to understand whether the UK has the research capability to exploit health records in research.

2012: The GPRD becomes the Clinical Practice Research Datalink (CPRD).

2013: RCUK funding is announced for the Farr Institute, a UK health informatics research institute to provide new insights into the causes of ill health and help cement the UK's reputation as a world leader in research using large electronic health data.

2013: The Health and Social Care Information Centre (HSCIC) is set up as an Executive Non Departmental Public Body (ENDPB): the trusted national provider of high-quality information, data and IT systems for health and social care.

2013: An RCUK-led consortium of 10 UK government and charity funders makes a £1.5m investment in a UK Health Informatics Research Network which will harness the expertise of the UK informatics research community.

1958

1991

2001

2007

2012

2013

1978: The first Cray-1 supercomputer in the UK is installed at the Daresbury Laboratory and reigns as the world's fastest computer from 1976 to 1982.

1985: UK scientists at the Rutherford Appleton Laboratory play a significant role in creating a series of international web and computer graphic standards, essential worldwide in the early adoption of these technologies.

1985: UK researchers collaborate with IBM and Microsoft to develop the first version of MS-Windows. MS-Windows version 3.0 was then released in 1990, and was the first Microsoft Windows version to achieve broad commercial success.

2004: Creation of the Digital Curation Centre, now a world leading centre that supports data management across the UK's higher education research community.

2001: The UK e-Science Programme begins as a £250 million five-year initiative involving RCUK and the Department of Trade and Industry. It has funded the development of e-Science technologies and demonstrated their use in a variety of applications.

2008: A small filter is developed by RCUK-funded engineers to help predict weather more accurately. The optical filters, 30mm in diameter and 1/100mm thick, will be installed in European Space Agency satellites, producing data to forecast weather, pollution and climate change.

2008: At 25 petabytes a year, the Large Hadron Collider at CERN generates more data than any other experiment in the world. In 2012, analysis of this dataset proved the existence of the Higgs boson.

2006: Solexa, a Cambridge spinout based on RCUK-funded research, launches its first sequencer, the Genome Analyser, which can sequence around 1Bn base pairs per day. Solexa is later sold to biotechnology company Illumina for \$600 million.

2011: The Imperial Antibiotic Prescribing Policy (IAPP) smartphone application is launched. This includes a decision-support system to help prescribers choose the right anti-infective treatment for patients, and a clinical calculator to help with accurate and safe prescribing of medicines.

2010: The Environmental Research Group at King's College London launch the London Air iPhone App which displays the latest air pollution levels recorded at over 100 monitoring sites in the London Air Quality Network.

2010: UK scientists, with Research Council support, commercialise a software product, Ikinema, which is revolutionising the games and animation industries.

2012: RCUK-funded researcher develops PEEK, a portable eye examination kit, using smartphone technology to undertake ophthalmic tests in remote and resource-poor locations. The data is stored on the phone and can be shared with specialists anywhere in the world to provide expert diagnosis and treatment plans.

2012: Oxford NanoPore Technologies, a spinout based on RCUK-funded work, reveal their third generation DNA sequencing technology, which can sequence the 3Bn base pairs of the human genome in a few hours.

2012: The £75 million Large Facilities Capital Fund (LFCF) investment in EBI (announced as part of the BIS strategy for UK life sciences) includes storage of over 125 petabytes of freely available molecular life science data by 2020.

2013: The MRC-NIHR Phenome Centre opens. Making use of the 2012 Olympics' state-of-the-art drug testing laboratory, it will examine around 100,000 blood and urine samples every year.

2013: The Oxford University Centre for Doctoral Training in Cyber Security is launched to cover some of the most pressing challenges our society faces today including the security of big data.

2013: RCUK signs an agreement endorsing the permanent establishment of the European Life-Science Infrastructure for Biological Information (known as ELIXIR) which supports European R&D by linking biological data, analytic tools and scientific literature, and ensuring they remain freely available to scientists of all disciplines.

2013: RCUK-supported scientists design GPS tracking collars to study cheetahs in Botswana and record the first data on the animals acceleration and manoeuvres which will enable them to see how managing habitats will have an impact on predators and hunting. The same technology is used to track 50 domestic cats in a Surrey village for a BBC programme to showcase scientific research methods to the public.

2013: The UK has 25 of the world's 500 most powerful computers. These include BlueJoule at Daresbury Laboratory ranked 13th and DiRAC in Edinburgh ranked 23rd in the world.

2013: Gung-Ho is a project to design and build a next-generation weather forecasting model for the UK. It is a collaboration between the Met Office and Research Councils and is one of the first major projects to benefit from STFC's new Hartree Centre.

1960: UK researchers at the Rutherford Appleton Lab establish computer facilities to visualise scientific data as animated films. As the most advanced facility in the UK, these were used to develop computer graphics.

1958: The National Child Development Study (NCDS) is established, following the lives of 17,000 people born in England, Scotland and Wales in a single week of 1958.

1970: The 1970 British Cohort Study (BCS70) begins as the British Births Survey (BBS), collecting data on the births and families of just under 17,200 babies born in a particular week in April 1970.

1967: SSRC Databank established. Over 58,000 datasets were delivered to the UKDS last year and the service has 24,000 registered users.

1979: The 1979 Academy Award for visual effects is given to the Ridley Scott film 'Alien'; the first major film to include CGI, which was generated on computers at the Rutherford Appleton Lab.

1973: The IBM 360/195 becomes the first and most powerful machine outside the USA to connect to the Advanced Research Projects Agency Network (ARPANET), an early form of the Internet.

1989: UK scientist, Tim Berners-Lee hands a document, 'Information Management: a Proposal' to his supervisor at CERN. This was the start of the World Wide Web.

1985: Rutherford Appleton Lab hosts the first UK website and later installs one of the first 50 Web servers in the world.

1992: Analysis of Large and Complex Datasets programme commences to advance and improve statistical analysis and modelling.

1991: British Household Panel Study (BHPS) launched as the UK's first socio-economic household panel survey. Surveyed 5000 people annually until 2008.

1996: An early RCUK spinout Nominet is created which manages the co.uk domain, one of the world's largest Internet registries. Nominet now manages 10 million UK business domain names, employs over 100 staff and in 2012 had a turnover of over £25 million.

2001: Research Methods Programme launched to develop qualitative and quantitative methods within the context of substantive research and effective dissemination of good practice through a range of related training activities that would have a direct impact on the methodological skill base of social science research in the UK.

2000: The Millennium Cohort Study (MCS), the fourth of Britain's world-renowned national longitudinal birth cohort studies established.

2000: The first map of ice thickness is published by the British Antarctic Survey, reducing the global estimates on freshwater by 10 per cent.

2004: Created from the investment by the cross-Council e-Science programme, the RCUK-funded e-Social Science programme commences which focusses on building tools and technologies to support new forms of social science.

2004: The RCUK-funded National Centre for Research Methods (NCRM) is established to provide a focal point for research, training and capacity building activities aimed at promoting a step change in the quality and range of methodological skills and techniques used by the UK social science community.

2005: Climateprediction.net reports the results of their first study in Nature. This experiment used the spare capacity of 95,000 PCs and revealed that a doubling of pre-industrial, atmospheric carbon dioxide could lead to temperature rises more than double those previously thought.

2009: Understanding Society commences. A unique and valuable study that captures important information every year about the social and economic circumstances and attitudes of people living in 40,000 UK households. Approximately 3 billion data points of information.

2009: The UK Climate Projections published using a tool developed by the British Atmospheric Data Centre. These have over 75,000 users.

2012: Cohort and Longitudinal Studies Enhancement Resources (CLOSER) established to bring together and maximise the use, value and impact of the UK's world class portfolio of birth cohort and longitudinal studies.

2010: Ministerial approval given for Life Study which will be the UK's largest national birth cohort study collecting data on up to 100,000 babies born in the UK.

The seven Research Councils are:

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

Research Councils UK
Polaris House, North Star Avenue
Swindon, Wiltshire SN2 1ET
United Kingdom
Tel: +44 (0) 1793 444420
Fax: +44 (0) 1793 444009
Email: info@rcuk.ac.uk
Web: www.rcuk.ac.uk

www.rcuk.ac.uk

Research Councils UK is the strategic partnership of the UK's Research Councils.

We invest annually around £3 billion in research. Our focus is on excellence with impact. We nurture the highest quality research, as judged by international peer review providing the UK with a competitive advantage. Global research requires we sustain a diversity of funding approaches, fostering international collaborations, and providing access to the best facilities and infrastructure, and locating skilled researchers in stimulating environments.

Our research achieves impact – the demonstrable contribution to society and the economy made by knowledge and skilled people. To deliver impact, researchers and funders need to engage and collaborate with the public, business, government and charitable organisations.

Image credits:

Autosub
National Oceanography Centre

Nissan Leaf
Department of Engineering Science, Oxford University