

Dynamics of Data Science: Models and mechanisms

This pack is an accompaniment to the Royal Society report, [*Dynamics of data science skills: How can all sectors benefit from data science talent?*](#)

There is a clear need for collaborative, sustainable initiatives to develop and nurture data science talent across all sectors. We have identified a range of models and mechanisms that already work to enable this. This includes outreach programmes, enrichment and fellowship schemes, capability-building programmes, informal or peer-to-peer mechanisms, collaborative events and data centres.

This booklet features exciting or innovative models and mechanisms that are already in place and that could be spread more widely. It features a variety of tried and tested ideas from across the UK, which require minimal to major resource support and can be led by individuals as well as organisations. Examples of good practice have been collected with the input of members of the data community from across academia, industry and the public sector.

The aim of the models and mechanisms booklet is to inspire scale-up and cohesion so that institutions and individuals can read about opportunities and resources in data science training and practice. This booklet can be used by people who are:

- concerned about the recruitment of data scientists
- involved in developing data science talent at all levels
- considering (re)training as data scientists
- making decisions around skills funding on a local, regional and national scale
- seeking to ensure that data they hold is used for societal benefit

Contents

Area for action: Developing foundational knowledge and skills

Models and mechanisms

3

Area for action: Advancing professional skills and nurturing talent

Models and mechanisms

5

Area for action: Enabling movement and sharing of talent

Models and mechanisms

7

Area for action: Widening access to data in a well-governed way

Models and mechanisms

10

AREA FOR ACTION:

Developing foundational knowledge and skills

MODELS AND MECHANISMS

Through our workshops and interviews we identified some key models and mechanisms that can help to address these priority needs. The mechanisms include integration of data skills into the school curriculum, apprenticeships to develop practical skills, alternative degree pathways and outreach programmes. The specific examples below serve as models of what can and has worked in specific contexts.

Mechanism

Integration of data skills into the school curriculum.

Model

Data Skills in Geography programme, Royal Geographical Society with the Institute of British Geographers

Recent changes in curricula in schools and at university, along with a recognised skills gap, have brought renewed emphasis on students being trained in data skills (the collection, analysis and presentation of data) in geography at GCSE, A Level and in undergraduate courses – and within Higher Education geography has been recognised by HEFCE as a ‘part-STEM’ subject. The shift is presenting new challenges for many school teachers, particularly those with little prior experience of such skills. In response to these changes and challenges, the Royal Geographical Society (with IBG) is leading a programme Data Skills in Geography, supported by funding from the Nuffield Foundation. RGS has also established new networks and strengthened existing ones, including creating a Data Champions scheme made up of teachers who are data enthusiasts with different levels of experience.

Mechanism

Apprenticeships to develop practical skills.

Model

The Office for National Statistics: Data Science Campus

The Office for National Statistics (ONS) is leading the way in developing a degree level apprenticeship in data science on behalf of the public and private sectors. The course is one of a number of new offerings at the new Data Science Campus¹. The campus is based at the ONS headquarters in Newport and part of a £17 million investment in statistics and data by the UK government to modernise and improve the statistics it produces. The funding was allocated after the Bean Review highlighted concerns around the use of administrative data².

All the training programmes at the campus cover three levels: raising awareness, embedding core skills and developing expertise. The ONS set up the scheme because it felt that there were no suitable existing programmes available in the university sector. The need to develop its own courses was important to equip data scientists with better insights into the UK economy for policymakers. All the applicants to the campus come from very different and diverse backgrounds. Some are mathematicians, physicists, social scientists, and artists. School-leavers and career-changers are also welcome. The ONS wanted a diverse array of educational pathways so that everyone can fulfil their potential, realising that work-based learning has a key role in delivering that ambition.

-
1. Poluck M. 2017 ONS opens £10 million hub to exploit digital data. *UK authority*. 27 March 2017. See <http://www.ukauthority.com/data4good/entry/7009/ons-opens-%C2%A310-million-hub-to-exploit-digital-data#> (accessed 18 April 2018).
 2. Chu B. 2015 Bean review tells ONS to be more “intellectually curious”. *Independent*. 3 December 2015. <https://www.independent.co.uk/news/business/news/bean-review-tells-ons-to-be-more-intellectually-curious-a6758226.html> (accessed 18 April 2018).

Mechanism

Alternative degree pathways.

Model

Q-Step, Nuffield Foundation, ESRC and HEFCE (now Office for Students (OfS))

The Q-Step programme, launched 2013/14 aims to address the quantitative skills gap in social sciences in the UK. With an initial six-year £19.5 million investment from the Nuffield Foundation, the ESRC and HEFCE, around 60 new academic staff were employed to use innovative approaches to embed quantitative methods and data analyses into social science teaching. There are currently 18 universities across the UK funded to deliver specialist undergraduate programmes, including new courses, work placements and pathways to postgraduate study.

Over 1,000 undergraduates a year start a Q-Step degree pathway, whilst almost ten-times as many experience enhanced quantitative teaching by taking one or more Q-Step modules. Over 420 employers are involved with offering placements to more than 860 students, ranging from private/public to the third sector and government departments. The programme is having a notable and positive impact on the data skills of social scientists taking up postgraduate research studentships. Q-Step will continue to be funded by the Nuffield Foundation and the ESRC for two more years (2019/20 and 2020/21) and is undergoing an independent impact evaluation.

Mechanism

Outreach programmes.

Model

Code First: Girls

Code First: Girls supports young adult and working age women to develop further personal and professional skills. Code First: Girls runs free coding courses; it also connects women to a community of other talented and like-minded women and companies who can support and accompany them through their professional development. Code First: Girls helps companies train their people, recruit new people, and develop their talent management policies and processes so they don't miss out on female tech talent.

AREA FOR ACTION:

Advancing professional skills and nurturing talent

MODELS AND MECHANISMS

To address these priority needs, mechanisms including enrichment and fellowship schemes, capability-building programmes and informal/ peer-to-peer mechanisms such as online courses, meet-ups and forums can prove effective. The following models are illustrations of how they can work in practice.

Mechanism

Enrichment and fellowship schemes

Model

Alan Turing Institute, Enrichment Scheme and Data Study Groups

One of the major goals of the Alan Turing Institute is to train new generations of data science and artificial intelligence leaders with the necessary breadth and depth of technical and ethical skills to match the UK's growing industrial and societal needs. The Enrichment Scheme offers students currently enrolled on a doctoral programme at a UK university the opportunity to join its research body. Doctoral students typically in their second or third year of study can undertake a 6, 9 or 12 month placement at the Institute's headquarters in London. Joining a community of more than 400 senior academics, early career researchers and PhD students, enrichment students have the opportunity to boost their skills and experience, enrich their research and make new collaborations during their time at the Alan Turing Institute.

The Data Study Groups are five-day 'collaborative hackathons', which bring together organisations from industry, government and the third sector, with talented multi-disciplinary researchers from academia. At each event, several organisations bring their real-world problems to be tackled by small groups of highly talented, carefully selected researchers, with a diversity of thought. Researchers brainstorm and engineer data science solutions, presenting their work at the end of the week. Organisations get to quickly prototype possible solutions to their data science challenges, and researchers get an opportunity to put knowledge into practice and go beyond individual fields of research to solve real-world problems. Knowledge is exchanged among groups, and participants from both academia and the organisations posing challenges rapidly learn new skills during the week – from how to work in secure analysis environments to learning new data science methods and techniques, and tools for doing data science collaboratively in groups.

Mechanism

Graduate development programmes.

Model

Office for National Statistics advanced training

For the more advanced level training (eg PhD level), partnerships were made between the ONS, the Alan Turing Institute and several universities. The ONS has built relationships with many departments of the UK government which have set up Data Science 'hubs', allowing them to regularly exchange and communicate on the needs and skills needed. This should improve the flow of data science talent across the UK.

Model

Faculty Fellowship

The Faculty Fellowship (formerly ASI Data Science Fellowship) exists to ensure that the brightest academics get a chance to immerse themselves in working life, learn about artificial intelligence (AI) in business and help build the future of operational AI.

Since its founding in 2014, Faculty has trained and transitioned over 250 PhD STEM graduates into data scientist roles in industry. Taking place three times a year (January, May and September), the Fellowship is highly competitive and receives applications from 5 to 10% of the UK's physics, mathematics and engineering postgraduate research students. In part, this is because alumni go on to work for big names like Google, DeepMind, Facebook and Deliveroo.

Faculty believes that AI is the most important technology of our age, but that it is only valuable when applied in the real world – enhancing products, improving services, and saving lives. To apply AI, organisations need the right strategy, software and skills. This is why fellows are given the opportunity to build their AI skills while working on tangible, real-world problems during a six-week placement.

Model

Pivigo data science training

Pivigo is a data science marketplace and training company based in London. It helps organisations to innovate through data science by connecting them with their own community of data scientists. Its Science to Data Science programme trains and graduates some of the world's top scientific PhD talent in data science, with three programmes each year. Pivigo runs a five-week programme at its London campus or online. It works with large multinationals, charities, SMEs, and start-ups to help learners gain practical experience with data science technologies and technical skills in a commercial environment. The scheme offers students the opportunity to boost their skills, grow their networks and work alongside researchers.

Mechanism: Capability-building programmes

Model

The Cross-Government Data Science Accelerator

The Data Science Accelerator is a 12 week skills-building programme that gives analysts and aspiring data scientists from across the public sector, including central and local government, the opportunity to develop their data science skills. Created in 2015, this award-winning programme has been recognised for its impact on increasing data capability across the Civil Service. More than 150 participants from across the country have delivered a variety of projects, many of which have made a substantial difference to their public sector organisations³.

3. Government Digital Service, Government Office for Science and Office for National Statistics. 2019. Introduction to the Data Science Accelerator programme. See <https://www.gov.uk/government/publications/data-science-accelerator-programme/introduction-to-the-data-science-accelerator> (accessed 15 April 2019).

AREA FOR ACTION:

Enabling movement and sharing of talent

MODELS AND MECHANISMS

There are a number of ways to make moving between sectors a natural part of the data science career path. Some mechanisms exist to foster long-term collaborative and business engagement networks and close interdisciplinary links. These include centres for doctoral training, industry fellowships, data residencies and innovative approaches to commercialisation. Here are some examples of existing models which highlight how the mechanisms are working in practice.

Mechanism

Centres for doctoral training

Model

Centre for Doctoral Training in Data Science and AI, The University of Edinburgh, and beyond

In 2013, the Engineering and Physical Sciences Research Council (EPSRC) invested £500 million in 115 Centres for Doctoral Training (CDTs), matched by more than £450 million from business, universities and other stakeholders. The University of Edinburgh has hosted a CDT in Data Science since 2014.

The CDT has two types of studentships including one with a PhD project in collaboration with an industry partner. For both courses the first year provides Masters level training in the core areas of data science along with a significant project. In years 2 – 4, students carry out PhD research in data science. In 2017, the UK's Science and Technology Facilities Council announced £10 million to train the next generation through supporting eight new CDTs in data intensive science. The centres include industrial partners and will offer comprehensive training in data intensive science through cutting edge research projects and a targeted academic training programme.

Model

DISCnet

The Data Intensive Science Centre is an STFC Centre for Doctoral Training providing a platform to train a new generation of data intensive scientists. The innovative education, training and research is delivered by a consortium of five universities from the South East Physics Network (SEPnet). The centre trains postgraduate students via world-leading research projects in particle physics and astrophysics and explores the untapped potential of these big data skills in diverse applications across a spectrum of industries. DISCnet currently has 70 non-academic partners.

Mechanism

Industry fellowships

Model

EPSRC Research Software Engineer Fellowships

The Research Software Engineering Fellowship is awarded to exceptional individuals in the software field, who demonstrate leadership and have combined expertise in programming and a solid knowledge of the research environment. As well as having expertise in computational software development and engineering, RSE Fellows should be ambassadors for the research software community and have the potential to be a future research leader in the RSE community.

Model

UKRI Future Leaders Fellowships

The Future Leaders Fellowship (FLF) is UKRI's flagship talent scheme, which aims to develop the next generation of research and innovation leaders in the UK. It will recruit and retain rising stars by attracting the brightest and best from at home and across the world. The FLF scheme will provide long-term funding for each Fellow (up to £1.2 million over an initial four years, with an option to extend to seven), allowing them to tackle difficult and novel challenges. A total of 550 FLFs will be awarded from 2019/20 to 2021/22 across six separate rounds, marking a significant investment to grow the UK's research and innovation base. Although the FLF scheme is not prescriptive of the research and innovation areas it supports, and Fellowships will be awarded on a competitive basis, there are several features that should make it attractive to those working in data science and AI. Whereas most existing Fellowship schemes fund only academic researchers, FLF also supports individuals in industry, as well as those working at the interface of academia, industry and the public sector, encouraging a new paradigm in career path that is mobile across all three. Operating across the breadth of UKRI will allow Fellows to take the most cross-cutting and interdisciplinary approaches to research and innovation. The open remit of the call allows for Fellowships to be held across a spectrum – from those with a background in AI wishing to apply their skills to a wide range of disciplines and challenges, to those who are from different disciplinary backgrounds, where AI could make a transformational contribution to that discipline or where that discipline could be brought to bear on the development of AI.

Model

Royal Society Entrepreneur in Residence scheme

The Royal Society Entrepreneur in Residence (EiR) scheme aims to increase the knowledge and awareness in UK universities of cutting-edge industrial science, research and innovation⁴. The scheme provides opportunities for outstanding industrial scientists and entrepreneurs to spend time working in a university to expose university staff and students to state-of-the-art industrial research and development, and the scientific challenges faced by industry. The scheme also allows universities to gain from expert advice aimed at promoting innovation and the translation of research by universities, and build confidence and understanding of business and entrepreneurship among staff and students.

Model

Royal Society Industry Fellowship scheme

The Royal Society Industry Fellowship is a paid secondment scheme for academic scientists who want to work on a collaborative project with industry and for scientists in industry who want to work on a collaborative project with an academic organisation⁵. Providing a basic salary for the researcher and a contribution towards research costs, the Fellowship aims to enhance knowledge transfer in science and technology between those in industry and those in academia in the UK. The scheme supports researcher-mobility and has run for over 30 years, bridging industry and academia for hundreds of scientists.

Model

Royal Commission for the Exhibition of 1851

The Royal Commission for the Exhibition of 1851 awards three-year research fellowships to early career postdoc scientists or engineers of exceptional promise⁶. The Fellowship, which was founded in 1891 and has initiated the careers of thirteen Nobel laureates to date, is open to all nationalities and fields of science, including physical or biological sciences, mathematics, applied science, and any branch of engineering. The Commission also awards Industrial Fellowships to encourage industry – academia collaboration at doctoral research level, Industrial Design Studentships for postgraduates and, in partnership with the Royal Academy of Engineering, graduate Enterprise Fellowships for entrepreneurs.

Mechanism

Residencies

Model

Microsoft AI Residency program (US/UK)

The Microsoft AI Residency program is a 12-month role designed to advance a career in machine learning research and engineering. The goal of the AI Residency is to help residents become creative and productive AI researchers, scientists and engineers. Residencies are open to BSc, MSc, and PhD graduates with substantial coursework in, but not limited to: computer science, electrical engineering, data science, mathematics, physics, economics, human – computer interaction, and computational biology.

4. Royal Society. (no date) Entrepreneur in Residence. See <https://royalsociety.org/grants-schemes-awards/grants/entrepreneur-in-residence/> (accessed 15 April 2019).

5. Royal Society. (no date) Industry Fellowships. See <https://royalsociety.org/grants-schemes-awards/grants/industry-fellowship/> See (accessed 15 April 2019).

6. Royal Commission for the Exhibition of 1851 (no date). Our awards. See <https://www.royalcommission1851.org/> (accessed 15 April 2019).

Model

Uber AI Residency (US)

Established in 2018, the Uber AI Residency is a 12-month training programme for recent college and Master's graduates, professionals who are looking to reinforce their AI skills, and those with quantitative skills and interest in becoming an AI researcher at Uber AI Labs or Uber Advanced Technologies Group (ATG). Uber AI Residents have the opportunity to pursue interests across academic and applied research. Uber is committed to an open and inclusive research mission that benefits the community at large, including contributing papers to top conferences and taking part in open-source projects.

Mechanism

Approaches to commercialisation

Model

IP free zone, Department of Computer Science and Technology, the University of Cambridge and beyond

The IP free zone at the University of Cambridge is part of a more general framework set up by the former Head of Department of Computer Science and Technology, Professor Andy Hopper FRS. The strategy has been to minimise barriers to the formation of new companies while aligning incentives for staff and students, avoiding IP issues, providing mentoring, being helpful in every possible way, and not picking winners. Furthermore, this has been a cradle-to-grave approach ranging from undergraduate lectures to the maintenance of an industrial business club beyond the department. A total of 270 companies have been formed by staff and students (including Raspberry Pi), of which 50% are active with revenues of \$1 billion, and 18% sold for over \$40 billion.

In the US, Carnegie Mellon and the University of Washington are currently working on a set of recommendations for commercial companies meant to provide a way for universities and companies to share talent more equally.

Model

UKRI Impact Acceleration Funding

Impact Acceleration Accounts (IAAs) are strategic awards provided to research organisations to support knowledge exchange and accelerate the impact of research. IAAs allow organisations to respond in more flexible, responsive and creative ways appropriate to their strategic priorities, enabling impact to be achieved in an effective and timely manner, for example, through secondments and exchanges, user engagement, proof of concept, and by building capacity for work across disciplines.

AREA FOR ACTION:

Widening access to data in a well-governed way

MODELS AND MECHANISMS

Mechanisms such as collaborative events and partnerships, data stores and APIs, offices of data analytics and data centres/institutes are important ways to bring data scientists and data together and the models below show how this can work, even for organisations that do not have the resources to hire data scientists themselves.

Mechanism

Collaborative events and partnerships

Model

Charity DataDives, DataKind UK

A hackathon brings together a range of people to generate an outcome, usually software projects. They can be associated with computer programmers, software developers, data scientists and, often, subject-matter-experts. The charity DataKind supports charities and social enterprises large and small across a variety of issue areas. It runs hackathon events called 'DataDives' where charities and social enterprises work alongside teams of volunteer data scientists, analysts, developers and designers using data to gain insight into their programmes and to increase their impact.

Hackathon style events can stimulate more engagement between the academic community and social projects as there are lots of skills within universities that are both expensive and in short supply within the third sector.

DataKind also run longer-term engagements over 6 – 9 months to build a data science solution (DataCorps) and have monthly office hours which any non-profit or social change organisation can sign up to for advice.

Model

Royal Statistical Society – Statisticians for Society

The Statisticians for Society initiative was launched in 2014, to help statisticians offer their skills to charities and other socially useful initiatives that need their professional expertise⁷. Many third sector organisations are keen to explore the use of data for decision making and service improvements. There is a growing need for them to provide evidence of their impact, but due to lack of capacity and appropriate skills needed for data analysis, some charities are unable to fully demonstrate the value of their work. As one of the leading voices for promoting the importance of data and evidence, the Royal Statistical Society supports statisticians in helping charities in making a difference. Volunteers can provide the tools and guidance for undertaking data analysis. Statisticians collect, analyse and interpret data across a wide range of industries and topics; they are skilled at designing methods for collecting data and regularly tasked with analysing data to spot patterns and trends; and they can manipulate data to identify relationships and make future predictions. Following this, they produce reports and summaries that communicate their findings.

Mechanism

Data stores

Model

The London Datastore

The London Datastore is a free and open data-sharing portal where anyone can access data relating to the capital⁸. It is one of the Greater London Authority's (GLA) flagship projects and is a platform through which many of the Smart London Plan objectives are delivered. Researchers are encouraged to visualise or build apps from the data available on the site.

Model

ONS API

The Office for National Statistics API makes datasets and other data available programmatically, allowing researchers to filter datasets and directly access specific data points.

7. Royal Statistical Society. 2019 Statisticians for Society. See http://www.rss.org.uk/RSS/Get_involved/Statisticians_for_Society/RSS/Get_involved/Statisticians_for_Society.aspx?hkey=c7977c58-1558-495a-9e5a-e99d64ea9cfd (accessed 15 April 2019).

8. London Datastore 2019. Welcome to the Datastore. See <https://data.london.gov.uk/> (accessed 15 April 2019).

Model

UK Data Service

The UK Data Service is the UK's only nationally funded research infrastructure for the curation and provision of access to social science data and its practices, especially around secure access to data and data curation, have been influential across the world. Funded by the Economic and Social Research Council (ESRC) to meet the data needs of researchers, students and teachers from all sectors, its unique collection of social science data resources includes major UK government-sponsored surveys, cross-national surveys, longitudinal studies, UK census data, international aggregate, business data, and qualitative data. It brings together several important past investments including the Economic and Social Data Service, Question Bank, Qualidata and Census Programme.

Climate change, ageing, security threats, the provision of better public services and a more productive society all call for policies and solutions that are informed by evidence-based research and innovation. Access and training services through the UK Data Service enable impactful research to influence national and regional policy and develop research excellence across all sectors. Its excellence in collecting, storing, analysing and sharing collections of complex data has helped it build a trusted reputation and become a critical part of the UK's research infrastructure.

Mechanism

Data stores

Model

Transport for London API

Transport for London (TfL) is the local public body responsible for public transport in London⁹. Every year, TfL ensures the transportation of 1.37 billion people, with a network length of 402 km, which is equivalent to 83.6 million km travelled per year¹⁰. Over the past ten years TfL has made a significant amount of data accessible to the public free of charge, including timetables, service status and disruption information. This has allowed the market to develop exponentially with the introduction of new products and services. TfL is now considered as a leader in publishing open data through APIs, the Cloud, the internet and across its physical network. It has created over 700 jobs and brought £14 million per year in GVA, enabling development of UK's skills in data.

Mechanism

Offices of Data Analytics

Model

Nesta's programme of Offices of Data Analytics

The Offices of Data Analytics (ODA) programme helps cities and regions join up, analyse and act upon data from multiple sources to reform public services¹¹. As of December 2018, there were nine initiatives that classify as Offices of Data Analytics across the UK. The model allows multiple organisations to join up, analyse and act upon data sourced from multiple public sector bodies to improve services and make better decisions.

Mechanism:

Data Centres / Institutes

Model

Health Data Research UK (HDR UK)

Health Data Research UK is uniting the UK's health data to make discoveries that improve people's lives. By bringing together the sharpest scientific minds, and providing safe and secure access to rich health data, it aims to better understand diseases and discover new ways to prevent, treat and cure them¹². Its vision is for large-scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and improve public health. To achieve this, HDR UK is leading an ambitious training and talent programme, and will create a cohort and network of thousands of health data scientists spanning all career stages, from school-leaver to senior research manager and international opinion leaders. The UK has a rich and diverse scientific talent base, thanks to the strength of the NHS, its academic institutions and innovative scientific and digital industries. HDR UK plans to harness this, bring on board international peers, to create an intelligent cohort of health data scientists that will dramatically change medical research, and open up new, faster, smarter pathways to patient care.

9. Transport for London 2019. Facts & figures. See <https://tfl.gov.uk/corporate/about-tfl/what-we-do/london-underground/facts-and-figures> (accessed 15 April 2019).

10. Deloitte 2017. Assessing the value of TfL's open data and digital partnerships. See <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> (accessed 15 April 2019).

11. Eaton M and Bertocin C 2018. State of Offices of Data Analytics (ODA) in the UK. See: https://media.nesta.org.uk/documents/State_of_Offices_of_Data_Analytics_ODA_in_the_UK_WEB_v5.pdf (accessed 15 April 2019).

12. Medical Research Council. 2019. Health Data Research Institute. See <https://mrc.ukri.org/about/institutes-units-centres/uk-institute-for-health-and-biomedical-informatics-research/> (accessed 15 April 2019).

Model

National Innovation Centre for Data (NICD)

The National Innovation Centre for Data (NICD) is a unique new facility that delivers data analytics skills into industry and the public sector by exploiting the knowledge and expertise currently locked within universities. A flexible rolling programme of collaborative projects focused on organisations' specific challenges and opportunities will transfer practical data skills into the workforce of those organisations. These projects will be supported by a range of related activities, including awareness-raising events, themed business and technical seminars and technical training courses. As a result of engagement with the Centre, organisations will be able to increase their productivity by optimising their existing operations, and to grow by launching new data-driven products and services.