

# Data skills for all

## Note of discussions at a Royal Society and STEM Learning workshop on data science skills

Friday 13 September 2019, National STEM Learning Centre, University of York

### The Royal Society and STEM learning

The Royal Society is the UK’s national academy of sciences. The Society’s fundamental purpose, reflected in its founding Charters of the 1660s, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

STEM Learning is the UK’s largest provider of education and careers support in science, technology, engineering and mathematics (STEM). Their state-of-the-art venue, based within the prestigious University of York campus, has a range of labs, teaching rooms and a resource centre. Their website is home to thousands of free-to-access, quality-assured resources to support teaching and learning.

### Dynamics of data science skills

Following the successful launch in May 2019 of the Royal Society’s policy report, *Dynamics of Data Science Skills*<sup>1</sup>, the Society is now taking forward a series of events and activities to explore ways to advance the report’s four areas for action and specific recommendations. The report, led by Professor Andrew Blake FREng FRS, looks at the current demand for data professionals and how this varies across industrial sectors and UK regions.

It identifies four major areas for action to meet the growth in demand for data science skills:

1. Develop foundational skills;
2. Advance professional skills and nurture talent;
3. Enable the movement and sharing of data science talent;
4. Widen access to data in a well-governed way.

### Developing data skills

The workshop focused on Action Area 1, developing foundational skills. The agenda for the workshop was based around a framework created for the report which highlights five areas to focus on to develop basic foundational skills (fig 1).

*Data Skills for All* was a joint workshop between the Royal Society and STEM Learning. This report is a summary of discussions at the workshop and does not present the views or positions of either organisation.

Figure 1

### A framework for developing data skills



1. *Dynamics of data science skills* (The Royal Society, 2019). See <https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf> (accessed 08 October 2019).

## Data-enabled citizens

Dave Gibbs, Senior Computing and Technology Specialist at STEM Learning, set the scene by highlighting how much data is being generated online every minute (fig 2).

Young people need to learn about the implications of how much data they produce in one day. The foundations for learning about this lie in mathematics and computing, but there is also a natural linkage to citizenship and other areas. Learning *about* data is equally important as learning how to handle it.

Children in primary school today will enter the workforce in roles that don't yet exist because of the way data, and data-enabled technologies such as artificial intelligence, are transforming the economy. This is leading to a dramatic shift in the demand for data skills nationwide at all levels from everyday user to expert.

There is a need for people with data skills across the spectrum. Young people should be encouraged to ask questions about data even if they haven't got the skills needed to process and analyse it. Not everyone needs to become a data scientist, but all young people need the ability to become informed and critical activists, particularly in an era of misinformation and 'fake news'.

## Data jobs

People in the workplace need to know how to work with data to make decisions and interpret quantitative information. Sufficient understanding of how data can be collected and used is necessary to make informed judgments in people's daily lives.

As the job market is changing and people are moving around more and more, skills requirements are also changing. There is increased demand for data professionals in the job market. The demand for workers with specialist data skills has more than tripled over five-and-a-half years and there is particularly high demand for people with specialist data science and data engineering skills (The Royal Society, *Dynamics of Data Science Skills*, 2019).

The STEM Ambassadors scheme has had a positive impact on schools and students through the provision of learning resources, but also on industry volunteers who reported increased job satisfaction. It is one of several mechanisms designed to bring much needed discipline-specific expertise into schools.

Emma McCoy, Professor of Statistics and Vice-Dean (Education), Faculty of Natural Sciences, Imperial College London, is a member of the Working Group for the Dynamics of Data Science Skills report, and also Chaired the 'Data Skills for All' workshop. She spoke about the changing nature of skills.

Some students who enter university with top grades in Mathematics and Further Mathematics (A\* at A-level) may still lack a strong understanding about data. This is a challenge because of the ways that the statistics profession has moved on. Historically, statisticians make inferences from well-designed trials; now the abundance of real-time data means that observational studies are more widespread. The analysis of observational data requires an understanding of the difference between causation and correlation. In addition, the increase of Machine Learning prediction tools based on historical data may result in bias, so we need to have an understanding of fairness and ethics. Like machine learning and AI, data management and computing are also important elements of modern data science. Consequently, a new set of skills is required to be able to handle big, messy datasets, and there must be horizon-scanning for new curricula.

Figure 2

### What happens in an internet minute



Image credit: This is what happens in an internet minute (2019) © Lori Lewis

## Diversity in the profession

A diverse pipeline of data scientists is important because diversity in the profession increases the likelihood of picking up and responding to inadvertent biases in algorithms that can impact many different types of people. Many data science roles are highly paid and so a more diverse pipeline could also help to close the gender pay gap, as well as lead to a more equitable distribution of the control of economic production.

In 2019 at school level girls outperformed boys in STEM subjects at GCSE for the first time in three years at the top A\*/A grades, despite the number of students awarded A and A\* grades falling overall. At GCSE, boys are still outperforming girls in maths, but it seems that their progress may have stalled slightly. The proportion of boys who got A\* was 8.2%, 0.7 percentage points higher than girls (7.5%). In physics, boys also outperform girls, but in chemistry and biology, girls do better. Male students are twice as likely to take A-level Mathematics and over eight times more likely to take computer science.

Women make up a disproportionately small fraction of the educational pipeline associated with data science positions, and further efforts are needed by all stakeholders to address diversity more broadly (beyond gender disparities).

Figure 3

### DfE Revised A-level and other 16 – 18 results January 2019

	Female	Male	% Female
Maths	32,586	49,484	40
Further Maths	4,081	10,213	29
Computer Science	1,077	8,162	12

## The curriculum in England

Vanessa Pittard, Deputy CEO and Stella Dudzic, Programme Leader (Curriculum and Resources), both at Mathematics in Education and Industry (MEI) spoke about the place of data science in the curriculum in England.

At Key Stages 1 – 4 (pupils between the ages of 3 – 16) many of the curriculum building blocks are in place for young people to develop pre-requisite skills in mathematics, the sciences and computing, and elsewhere<sup>2</sup>. At this level there is no strong case for (potentially unwelcome) curriculum change, though a more detailed, updated and coherent computing curriculum is needed. This could be done through non-statutory means and the National Centre for Computing Education (NCCE) is currently looking into developing curricula and resources. This is already starting to have an impact on the teaching and learning of computing.

At A-level, there are great examples of where quantitative content has been embedded in A-level subjects such as geography or psychology, where statistics is made explicit and assessed<sup>3</sup>. However, the 16 – 18 curriculum has tended to narrow due to funding constraints and accountability pressures, when it should be broadening to provide essential mathematics and statistical reasoning for all. It has been left to schools to foster this curriculum intent, so there is a case for wider curriculum change at this level. There are positive developments at A-level, however studying a wider range of subjects post-16 since curriculum has slightly narrowed due to funding constraints and accountability. Traditionally non-STEM subjects will need to use data, which is an important policy line to note.

A wider issue is that England is bottom of the league tables regarding post-16 participation in mathematics<sup>4</sup>. The current policy for requiring students who fail GCSE Mathematics to take resits is ineffective and the implications of the policy should be revisited. There needs to be a curriculum change to encourage schools to offer mathematics post-16 and improve students' attainment.

2. *The integration of data science in the primary and secondary curriculum* (Pittard, 2018). See <https://royalsociety.org/~media/policy/Publications/2018/18-07-18-The-Integration-of-Data-Science-in-the-primary-and-secondary-curriculum.pdf> (accessed 08 October 2019).

3. *Embedding statistics at A-level* (ACME and the Royal Statistical Society, 2015). See <https://royalsociety.org/~media/policy/Publications/2015/embedding-statistics-at-a-level-07-2015.pdf> (accessed 08 October 2019).

4. *Jobs are changing, so should education* (The Royal Society, 2019). See <https://royalsociety.org/~media/policy/Publications/2019/12-02-19-jobs-are-changing-so-should-education.pdf> (accessed 08 October 2019).

## Support for teachers

The impact of the data revolution touches everybody, including teachers who may need data skills to teach their subject and to inform their own teaching through more sophisticated analysis of student data.

Educators need to think about how to foster quantitative reasoning and problem solving across subjects, but there is a gap between curriculum requirements and what is being taught in schools.

Barriers to progress include low confidence among teachers in using and applying mathematics; and lack of experience in formal mathematics among teachers of other qualitative subjects. It is important to understand what teachers know about data, data management and artificial intelligence. However, it may be challenging for teachers to gain depth of understanding of the building blocks of these areas.

Computing teachers need access to up-to-date information on relevant conceptual and technical aspects of computing, machine learning and AI and data management. These need to be addressed through professional development programs and more detailed guidance and support to teachers. There is a lot of hype around data science – some of what we consider to be AI is standard computational analysis.

This is all situated in a difficult context for recruiting and retaining good teachers in certain subjects. Therefore, it will also be important to consider new strategies for incentivising good teachers to join and remain in the profession.

## Motivating pupils

Miles Berry, Principal Lecturer in Computing Education at the University of Roehampton, highlighted the importance of providing motivating contexts to engage students in data-related activities.

Data science is an interesting context for motivating students in schools since it is about real-world situations. This can be seen in the tripartite framework proposed for computing education in the Royal Society report *Shut down or restart?* that includes foundations (computer science), applications (information technology) and implications (digital literacy)<sup>5</sup>. This framework could also be applied for data science. School curricula should include topics in probability and statistics, although currently the English primary curriculum does not include probability. Pupils might also apply their spreadsheet and programming skills to problems arising in analysing and communicating data. They should also consider ethical aspects of the pervasiveness and permanence of data, not least in relation to their own privacy.

## Innovation in Scotland

Kate Farrell, Director of Curriculum Development at the University of Edinburgh and the Data Education in Schools project, presented data-driven curriculum innovations in Scotland to address the skills shortages in subjects such as health and social care, tourism, creative arts, finance and fintech.

The universities in the Edinburgh area have committed to training 10,000 more students in data science as part of their degree over the next decade. Scotland was interested in how to boost data education in schools across all relevant subjects. The Data Education in Schools project looked at trialing teaching resources with schools and raising awareness of the importance of data. The National Progress Award (NPA) in Data Science is an example of a qualification at levels 4, 5 and 6 (equivalent to GCSE and A-level in England). In this qualification, data science can be embedded in other subjects which are themed (or flavoured) differently, depending on the contexts, datasets and case studies used.

## The IDSSP: towards a new international qualification in data science

Neil Sheldon, Chair of The Teaching Statistics Trust, presented a new pre-calculus data science curriculum (International Data Science in Schools Project: IDSSP) that has been developed with a range of overseas partners.

This initiative involves two frameworks for introducing Data Science, including topics and learning outcomes. The first is a curriculum framework to underpin the design of courses to teach secondary school students, and the second serves a parallel purpose by preparing teachers to teach Introductory Data Science to their students. The curriculum was developed by a broad-based international team of computer scientists and statisticians. It would depend on teacher professional development specifically in data science. In England, the curriculum framework might be best implemented as an A-level. Alternatively, a more vocationally oriented course might sit within the T-levels currently being developed. In either case, it would be important to recognise and accept that a skills-based curriculum of this type would be best assessed by carrying out projects at the time of learning. Terminal examinations of the kind that dominate A-levels would be much less appropriate.

---

5. *Shut down or restart? The way forward for computing in UK schools* (The Royal Society, 2012). See <https://royalsociety.org/~media/education/computing-in-schools/2012-01-12-computing-in-schools.pdf> (accessed 16 October 2019).

## Discussion: Overcoming challenges

There are several implementation challenges to embedding data skills training into education. The afternoon session required participants to work together to answer six fundamental questions regarding data science education focusing on primary, secondary or other aspects of education, e.g. parents, teachers, and older learners.

### Challenge 1

#### How can known constraints be overcome to achieve data skills for all?

These could include: curriculum overcrowding, a lack of teacher time, poor infrastructure, a focus on assessment and accountability, and a lack of support to develop teacher knowledge and skills.

- Provide subject-specific teacher CPD with buy-in from senior leadership teams, including online opportunities.
- Use consistent language – develop a glossary and identify the difference between data skills, data literacy and data science within the limits and scope of each.
- Make data skills part of Senior Leadership teams' development plans.
- Help teachers enliven what they are teaching by introducing data science by fun and stealth. For example, tiny tasks for primary – dodgy data / daily data / data dilemma discussions.
- Provide opportunities for informal learning in addition to core curriculum.

### Challenge 2

#### How can we make sure that data skills are taught cross-curriculum and who should take the lead?

- Devise real-world contexts for lesson content.
- Highlight safeguarding issues around use of personal data.
- Focus on ethics of personal data use.
- Embed data as a core skill (like reading and writing) across all subjects e.g. science, maths and humanities.
- Run 'unplugged' activities – do not solely focus on technology (but recognise that tech is integral to data science).
- Appoint digital skills champion in all schools.
- Encourage teachers to communicate with each other and share resources.

### Challenge 3

#### What can/should other organisations do?

- Regulators – Ofsted and Ofqual – and policy makers should make data training a priority in accountability and inspections.
- Consider devising a 'data-literacy' mark for schools.
- Engage Initial Teacher Training (ITT) providers.
- Push for curriculum reform, detailed curriculum framework with age-related expectations.
- Produce resources – e.g. '10 big ideas about data science' and banks of easy-to-use resources.
- Develop a cross-organisation 'data science working group' to ensure consistency of messaging regarding resources.
- Produce guidance for schools on purchasing tech and resources so that limited funds are not wasted (NCCE).

#### Challenge 4

What are the highest priority issues and how can these be overcome?

- Address infrastructure in schools including assess schools' budgets – there is a lack of computers and software has a high cost.
- Design easy to use software for teachers and students.
- Ensure teachers have time to develop resources and curriculum content.
- Improve access and awareness of large, curated datasets.
- Introduce qualifications or promote integration into exam framework.
- Push the message to schools that computer science is not just about coding to improve uptake.
- Develop a post-16 curriculum and qualification similar to Scotland across the rest of the UK.

#### Challenge 5

How can data literacy in the home be promoted?

- Offer parental engagement activities e.g. how to avoid 'fake news'.
- Use social media to increase awareness.
- Encourage children to take their learning home.
- Open conversations between families and children about what data they are happy to share or to post on social media.
- Encourage exploration of data use in accessible investigative news shows.
- Expand use of 'makerspaces' and community outreach opportunities especially for disadvantaged and vulnerable students.
- Focus on data in everyday life – not just for careers.
- Encourage families to take part in Citizen Science projects, engaging with libraries and community centres.
- Promote data capital within science capital.

#### Challenge 6

What are the basic elements that need to be incorporated into lessons?

- Build key definitions – e.g. what is data science?
- Actively implement practice in statistical interpretation across all subjects.
- Support education and awareness regarding real world contexts and 'fake news'.
- Provide links to careers and information guidance – highlight what employers want, support data scientists going into schools, make links to Gatsby career benchmarks, and highlight prospects and opportunities.
- Identify the different roles of data scientists – e.g. collection of data, data cleaning, interpretation and machine learning.
- Consider a framework of skills that could be developed at each level of learning<sup>6</sup>.
- Explore the relationships between data collection, raw data, the back story and data visualisation.
- Use the cycle of enquiry: problem – planning – data collection – analysis – conclusion.
- Show links between data and AI and machine learning.

6. *Fostering data literacy competencies in school* (Grillenberger, 2019). See [http://computingeducation.de/pub/2019\\_Grillenberger\\_HelloWorld.pdf](http://computingeducation.de/pub/2019_Grillenberger_HelloWorld.pdf) (accessed 09 October 2019).

## Participants

Samantha Ahern, Digital Education Innovation and Development Officer, UCL

George Boukeas, Learning Manager, Raspberry Pi Foundation

Kisha Bradley, Science Communicator, Bright Box Makerspace

Mike Cargill, Managing Director, UK STEM Ltd

Jo Cox, Schools Engagement Manager, The Royal Society

Fran Dainty, Head of STEM Content and Expertise, STEM Learning

Emily Dreimann, Project Manager, David and Jane Richards Family Foundation

Stella Dudzic, Programme Leader (Curriculum and Resources), Mathematics in Education and Industry (MEI)

Juliane Fagotti, PhD Student, Imperial College

Kate Farrell, Director of Curriculum Development, University of Edinburgh, Data Education in Schools project

Sue Finnigan, eLearning Consultant, Sheffield City Council

Matthew Forshaw, Data Skills Policy Leader, The Alan Turing Institute

Zeireen Fuzurally, Project Co-ordinator, The Royal Society

Claire Garside, Co-founder and postgraduate researcher, Foundation for Digital Creativity / Leeds University

Laura Gemmell, PhD Researcher / Founder, University of Bristol / Taught by Humans

Dave Gibbs, Senior Computing and Technology Specialist, STEM Learning

Bernard Ginns, David and Jane Richards Family Foundation

Sway Grantham, Teacher and Consultant

Steve Grundy, Programme Manager, Q-Step, Nuffield Foundation

Jon Hale, Head of Biology, Beaulieu Convent School

Helen Harth, Senior Policy Adviser, The Royal Society

Carole Haynes, Education Manager, Royal Statistical Society

Kathryn Horan, Teacher and Consultant, That Science Lady

Zeeniya Kamil, Research Associate, UWE Bristol

Tommy Lawson, Schools Technology Officer, University of Edinburgh

Natasha McCarthy, Head of Policy (Data), The Royal Society

Emma McCoy, Professor of Statistics, Imperial College London

Ian McDaid, Teaching and Learning Lead for Science, Horizon Community College

Maria McGrory, Teacher, Hudson Road Primary School

Bill Mitchell, Director of Policy, BCS – The Chartered Institute for IT

Jennifer Panting, Policy Adviser, The Royal Society

Vanessa Pittard, Deputy Director, Mathematics in Education and Industry (MEI)

Versha Prakash, Postdoctoral Research Associate, Royal Holloway University of London

Chris Robbins, Owner, Grallator

Andrew Robson, Head of Computing, Gosforth Central Middle School

Paul Scott, Curriculum Innovation Manager, BMDC

Neil Sheldon, Chair, The Teaching Statistics Trust

Sarah Taylor, Teacher, Winchcombe Primary School

Jane Waite, Research and teacher training, Queen Mary University of London

Michel Wermelinger, Senior Lecturer, The Open University

Jon Witts, Director of Digital Strategy, Queen Margaret's School