THE INTEGRATION OF DATA SCIENCE IN THE PRIMARY AND SECONDARY CURRICULUM

FINAL REPORT: To the Royal Society Advisory Committee on Mathematics Education

PREPARED BY: Dr Vanessa Pittard

ACKNOWLEDGEMENTS

This research was undertaken by **Dr Vanessa Pittard**.

The Royal Society Advisory Committee on Mathematics Education's (ACME) data science work is being overseen by **Professor Emma McCoy**, Vice-Dean (Education) for the Faculty of Natural Sciences and Professor of Statistics in the Mathematics Department at Imperial College London, and **Sir David Spiegelhalter OBE FRS**, Winton Professor for the Public Understanding of Risk at University of Cambridge.

The work was carried out in 2018 and the report published in July 2018.

1. Contents

2. Executive summary

2.1. Scope of the review

This review looks at England's primary and secondary national curricula (Key Stages 1–4) to identify the extent to which data science (or elements of it) exist within computing, mathematics, science, geography, history and other data-rich subjects.

Where data science recognisably exists, it considers how data science skills are developed across subjects up to the end of Key Stage 4 and identifies specific topics that are conducive to nurturing data science skills and enhancing teaching of statutory curriculum content.

2.2. Headline findings: primary curriculum

Core knowledge and skills are developed mainly through the statistics strand of mathematics at Key Stage 2 and algorithms and use of digital technology in computing at Key Stages 1 & 2.

At lower Key Stage 2, science curriculum requirements to 'work scientifically' highlight the application of mathematics, but this is not fully reflected in guidance on attainment targets.

Neither history nor geography at Key Stages 1 & 2 include mathematics application or use of data, focusing instead on building a base of subject knowledge.

Despite this, overall, if taught well, pupils are likely to gain a reasonable introduction to underpinning elements of data science in the primary curriculum.

2.3. Headline findings: secondary curriculum

In secondary mathematics statistics content is carefully sequenced to support conceptual progression. The GCSE contains relevant content at both Foundation and Higher tiers.

Key Stage 3 computing contains substantial content relating to data science and represents a step-up from Key Stage 2. GCSE Computing qualifications offer a detailed grounding in the fundamentals of computing but do not build on this nor include machine learning and AI.

Secondary science emphasises the application of mathematics and the centrality of the classic data cycle. This includes content on appraising claims based on data.

Key stage 3 and 4 geography strongly emphasises the collection and analysis of quantitative data. Data analysis in GCSE Psychology is integrated into empirical or experimental psychological research methods, much in the same way as the natural science data cycle.

The secondary history curriculum is essentially non-quantitative and offers very little scope for nurturing data science skills.

2.4. Barriers

At primary level, the main challenges to the coherent integration of data science are:

- A lack of specificity in the computing programme of study at Key Stage 2, meaning that little detailed content is set out on algorithms and data presentation and analysis and this is therefore left to the discretion of teachers.
- A potential disjunction in the Key Stage 2 science curriculum between statements about the importance of applying mathematics and the study of scientific curriculum topics. Linkage is dependent on teachers and teaching.

A major curriculum challenge at secondary level is the apparent lack of systematic progression in relevant aspects of computing. Computational modelling at Key Stage 3 may be difficult to study in a meaningful way and needs careful thought. It is not clear how computational modelling is built on at GCSE.

While Key Stage 4 mathematics and science include content on making accurate claims using data, there is insufficient focus in the curriculum overall on critiquing arguments based on data.

It may be difficult to teach content in the humanities, sciences and social sciences drawing on real data, as this is often too complex to be brought into the sphere of secondary study.

2.5. Overcoming barriers

Most of the curriculum issues highlighted can be overcome without recourse to potentially unwelcome curriculum change, but a more detailed, updated and coherent computing curriculum is needed – this could be through non-statutory means.

Other barriers relate to teacher confidence in applying mathematics to domain questions, and in conceptual and technical aspects of computing. These issues could be addressed through professional development opportunities and more detailed guidance to teachers.

Simplified 'intermediary' datasets and models with 'conceptual veracity' in secondary education would be helpful to bring real data into the analytical sphere of students.

2.6. Further opportunities

Further opportunities for integrating data science in the primary curriculum include:

- The application of mathematics to specific questions in science to enhance subject understanding.
- Using meaningful data in computing when applying technology in the analysis and presentation of data – linked to other areas of the curriculum.
- Enriching knowledge-based history and geography curricula at primary level through informative infographics and data visualisation.

The secondary curriculum offers a wide range of possibilities for incorporating real uses of secondary data across the sciences and humanities, including:

- Biology content at Key Stage 4/GCSE - how, at a high level, scientists use such data to develop and test theories and hypotheses and draw valid conclusions (for example on the relationship between health and disease and the importance of biodiversity).
- Geography in-depth case study at GCSE – for example, using quantitative indices and examining related data to understand economic development at country level.
- Developing geographic investigation using GDS/GIS data – relatively underplayed in the curriculum currently.
- Emphasising appraisal and critique of arguments based on data in biology, geography history and the social sciences.

3. Introduction

3.1. The field of data science

As Sir Adrian Smith's review of 16-18 mathematics observed, 'data science' has become both a significant academic discipline and a source of employment, with demand for data science experts far outstripping that for other employment sectors.[1] Much of this demand arises from the way that digital technologies are fundamentally changing how people live and the nature of work. Commercial, industrial and academic uses of data have expanded considerably over recent years. In the commercial sector, real-time analysis of online transactional data has become central to achieving competitive advantage in some sectors and has led to new applications of predictive analytics and machine learning to address business problems.

Data scientists are found across a wide range of domains, applying a wide range of statistical and computational approaches. In common, to be effective, they learn from data and require understanding of problems and goals within the relevant domain. They can require substantial subject matter expertise, particularly in the sciences where there are strong moves towards making scientific data openly available. Such moves are likely to expand the scope and impact of data science and revolutionise research paradigms[2].

3.2. Definition for this review

The term 'data science' has existed for several decades and definitions have evolved over time, partly as a reflection of changes in technology and data handling[3]. During the late 1990s and early 2000s there was a move in the US to equate the term with statistics and re-label the discipline as 'data science', but newer uses of the term do not reflect this perspective. The field exceeds the boundary of statistics and represents the bringing together of statistics and computing (notably machine learning) to learn from 'big data'.

While this helps to frame the focus of this review, it is important to note that there is no consensus on the definition of data science. There may be a need for the Royal Society to develop a definition applicable across the breadth of its data science-related work. For this review a working definition has been adopted to reflect the following:

- The focus is not on how Key Stage 1-4 schooling creates data scientists. Data science sits in the realm of advanced study – people choose related specialisms well after Key Stage 4.
- Nonetheless, Key Stage 1-4 study should provide a grounding to ensure that all young people develop underpinning knowledge and skills, regardless of whether they aspire to become data scientists.
- This grounding will be supported both through mathematics and computing (technical conceptual grounding) and study within other data-rich subjects (domain application).
- All students should understand how data contributes to advances in knowledge, be able to critique claims based on data and understand the role of data in the modern world.

3.3. Approach taken

This review identifies:

- Knowledge and techniques in mathematics, statistics and computing used in the development, analysis and use of data.
- Study in other subjects which incorporates the use or analysis of data or draws on mathematical and computational ideas and techniques.

[1] Report of Professor Sir Adrian Smith's review of post-16 mathematics (see https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/630488/AS_review_report.pdf)

[2] Donoho, David (September 2015). "50 Years of Data Science". Based on a talk at Tukey Centennial workshop, Princeton NJ Sept 18 2015.

[3] https://en.wikipedia.org/wiki/Data_science

- Broader opportunities to engage students in studying and understanding data science.

The analysis is based on published National Curriculum programmes of study at Key Stages 1 & 2 for mathematics, computing, science, history and geography. At Key Stages 3 & 4 it reflects both published programmes of study and GCSE specifications for mathematics, computing, science subjects, humanities and the social sciences. Curriculum content is highlighted in Annex 1.

The central questions posed and dealt with are: 'How well does the school curriculum provide a foundation for data science?' and 'How might it need to adapt in light of the increasing importance of data science?'. The resulting analysis highlights both potential curriculum issues and opportunities.

4. Development of data science through the primary curriculum

4.1. Core knowledge and skills

While elements of data science feature in the primary curriculum at Key Stage 1, this is restricted mainly to an introduction to categorising and counting and algorithms and simple programs as instructions. There is little at Key Stage 1 by way of application of basic mathematics in science and no data science elements within Key Stage 1 geography or history.

The core knowledge and skills relating to data science are developed predominantly through introductory 'statistics' in mathematics at Key Stage 2 and the study of algorithms and use of digital technology in computing at Key Stage 1 and Key Stage 2, with a potential contribution from Key Stage 2 science, depending on how this is taught.

Mathematics

At Key Stage 2 the mathematics curriculum introduces simple representations of quantitative data and basic analysis, such as making comparisons. This is built on through Key Stage 2 with a focus mainly on representing data in basic tables and charts and undertaking simple analyses of univariate quantitative data.

Mathematics content at Key Stage 1 and lower Key Stage 2 places significant emphasis on concepts in number and fluency in number operations. Reflecting this, Key Stage 1 statistics content focuses on simple tasks such as categorisation, tallying and quantifying. Pupils are introduced more formally to data and statistics at year 3, when they are taught to:

- *Interpret and present data using bar charts, pictograms and tables;* and
- *Solve one-step and two-step questions using information presented in scaled bar charts and pictograms and tables*

Associated non-statutory guidance at year 3 states that pupils should *continue to interpret data presented in many contexts*. Given the strong emphasis on other aspects of mathematics in early Key Stage 2 and the breadth of the mathematics curriculum, sufficient exposure to such examples is dependent on approaches taken to teaching in other subjects, notably science.

Years 4 and 5 introduce important early concepts in the representation and interpretation of quantitative data, such as the concept of 'discrete' and 'continuous' data and the representations appropriate for different types of data. At year 6 pupils move from basic presentation and interpretation of data into descriptive statistical analysis. They are expected, among other things, to use graphical representations of data to solve problems, understand and calculate the mean of a dataset and draw graphs relating two variables *from their own enquiry and other subjects*.

Computing

The Key Stage 2 computing curriculum develops pupils' understanding of algorithms and skills in programming, focusing on logical reasoning to explain how simple algorithms work. Pupils are required to:

- *Solve problems by decomposing them into smaller parts;*
- *Use sequence, selection, and repetition in programs;* and
- *Work with variables and various forms of input and output.*

They are also required to:

- *Analyse, evaluate and present data and information using a variety of software*.

Neither area is specified in great detail, however, and therefore how these areas are taught, rather than the curriculum statements, is critical to the development of core knowledge and skills.

The computing programmes of study were deliberately under-specified to avoid over-prescription, which might stifle innovation and currency, and reduce risks that what children study becomes dated. Helpful resources, schemes and programmes do exist for primary schools, some of which (such as the BCS primary computing curriculum and programmes) set out in some detail what pupils should know and be able to do in relation to data.

But given the lack of detail in the curriculum, the knowledge and skills acquired by pupils will reflect the choices made by schools about *which* schemes to use and their quality. This presents the risk that there will be gaps in pupils' understanding of underpinning aspects of data science.

Science

The introduction to the primary science curriculum emphasises the central importance of mathematics and data within science, stating that pupils should:

- *Apply their mathematical knowledge to their understanding of science, including collecting, presenting and analysing data;* and
- *Seek answers to questions through collecting, analysing and presenting data.*

At lower Key Stage 2, curriculum requirements to 'work scientifically' reflect the application of mathematics, including:

- *Presenting data in a variety of ways to help in answering questions;* and
- *Recording findings using bar charts, and tables.*

At upper Key Stage 2, the application of mathematics content in science is further developed as part of 'working scientifically' through:

- *Recording data and results of increasing complexity tables, scatter graphs, bar and line graphs;* and
- *Reporting and presenting findings from enquiries, including conclusions, causal relationships and explanations*.

Detailed science content is set out separately from the statements on 'working scientifically'. This is not necessarily helpful. Domain content is listed as subject 'attainment targets' in areas such as 'living things and their habitats' and this is accompanied by non-statutory guidance below each area, which *should* state how pupils might work with data in relation to these areas.

Unfortunately very little is set out on the application of mathematics and use of data in these sections. This highlights a disjunction between statements about the application of mathematics at the start and recommended practice within the study of topics. There is more on this issue in the section below.

History and Geography

Curriculum statements for primary history and geography (both Key Stage 1 and Key Stage 2) focus entirely on building a base of subject knowledge. Neither programmes of study make mention of the application of mathematics nor cite the use and analysis of data; this is left to the discretion of the school or the teacher.

Despite these issues, overall, if taught well, pupils are likely to gain a reasonable introduction to underpinning elements of data science in the primary curriculum. Where there are issues, these focus more on sequencing and coherence.

4.2. Sequencing and coherence

In mathematics in particular, data science elements are carefully sequenced – built on in small steps with the aim that pupils gain a secure grasp of them.

As the computing curriculum is not specified in such detail as mathematics, it is less clear whether this is the case for the study of algorithms and digital technology for presenting and analysing data – this will depend, as highlighted above, on the quality of published schemes adopted by schools and how this area is taught.

The quality of sequencing in detailed schemes for Key Stage 2 computing used by schools is something that should be looked at in more detail, to establish how well such schemes support a secure understanding of algorithmic concepts and a provide a sound foundation in the presentation and analysis of data using technology.

In terms of coherence, presentation of data and information using a variety of software in computing presents a good opportunity for cross-curricular study through the use of data from other curriculum subjects. This is not something highlighted in currculum statements, nor in published schemes, meaning there are risks that meaningful examples are missed and these exercises become trivial.

Problems with non-statutory guidance on ‘working scientifically’ were highlighted in the section above. Without guidance, it may be difficult for primary teachers to judge where in the Key Stage 2 science curriculum it is fruitful to incorporate the use of data and which mathematics and statistics are suited to given domain topics and questions.

It would be helpful to publish detailed examples for Key Stage 2 teachers in the effective application of mathematics in science. Similarly, it would be helpful for Key Stage 2 teachers to have access to detailed examples of the application of technology in computing to the analysis of meaningful data from other curriculum areas.

Finally, the curriculum makes no explicit link between mathematics and computing content. Making links of this kind could be helpful in enhancing the rigour of aspects of computing, and should ideally be reflected in published schemes.

5. Development of data science through the secondary curriculum

5.1. Progression from primary

In mathematics, progression from primary to secondary in content relevant to data science is carefully planned. Key Stage 3 content in the 'statistics' strand consolidates and builds on what is taught at Key Stage 2 in small steps, helping support understanding that is secure. The same is true for related Key Stage 4 content in mathematics, which reinforces Key Stage 3 statistics content and builds on this in small steps.

Again because of a relative lack of detail, precise progression in elements of data science in computing is less clear. It is evident, however, that the level of challenge is higher at Key Stage 3 than Key Stage 2, and therefore does represent a progression.

Crucially, the step is made at Key Stage 3 to *study computational abstractions that model the state and behaviour of real-world problems and physical systems*. Practical techniques for handling data are also taught, including algorithms for sorting and searching. Given this considerable 'step up', careful sequencing of this study is critical – something raised below.

The Key Stage 3 science curriculum further develops students' skills in the analysis of data and the application of statistical techniques. This is set at an appropriate level, reflecting the mathematics that students have been taught at Key Stage 2 and are being taught at Key Stage 3.

Given the focus in geography and history at Key Stage 2 on a grounding in domain knowledge, Key Stage 3 is the first real point at which students encounter data analysis in the humanities. This is found in geography rather than history. Data systems (notably geographical information systems) and the analysis and interpretation of data are introduced.

5.2. Core knowledge and skills

Mathematics

In Key Stage 3 mathematics students progress to more complex analysis of univariate data (measures of central tendency and spread), deepen their understanding of how to represent data of different kinds (for example, frequency tables, bar charts, pie charts and bar charts) and move on to simple bivariate analysis.

Importantly, Key Stage 3 mathematics requires students to *describe simple mathematical relationships between two variables (bivariate data) in observational and experimental contexts and illustrate using scatter graphs*, offering an important opportunity within mathematics to address questions from other domains.

Key Stage 4 reinforces and builds on this content, placing emphasis on more complex analysis and representation of univariate and bivariate data (for example, box plots; correlations).

The important step is taken in mathematics at Key Stage 4 of introducing concepts relating to (underpinning) sample analysis and inferential statistics. The new mathematics GCSE contains significant statistics content at both Foundation and Higher tiers, offering all pupils a grounding in these areas, including what can and cannot be concluded from statistical analysis of different sorts.

Computing

Key Stage 3 is the first point at which students engage seriously with aspects of data science in computing, notably formal study of algorithms, computational modelling, data representations and data structures. The curriculum states that at Key Stage 3 students should *design, use and evaluate computational abstractions that model the state and behaviour of real-world problems and physical systems* and *understand several key algorithms that reflect computational thinking, and make appropriate use of data structures*.

Computing curriculum requirements at Key Stage 4 are minimal, reflecting the fact that GCSE Computing is optional. Students not taking GCSE Computing fulfil these requirements by further developing their computing capability, analytic and design skills elsewhere.

Students taking GCSE Computing generally study a formal computer science curriculum, focused on offering a detailed grounding in the fundamentals of computing theory and practice and developing formal programming skills.

There are differences between Computing GCSE qualifications in the treatment of content of relevance to data science. All offer a fairly 'traditional' introduction to database management systems (DBMS) and data querying, focused on theory rather than practical use, but exam boards differ in their treatment of other aspects of data science – for example, in relation to 'data representation': focusing variously on low-level (machine-level) representation, structured versus unstructured datasets, and data file types (for example, HTML).

There is no focus on machine learning or artificial intelligence in any specification. The overall feel of GCSE Computing specifications is that they are dated.

It would be helpful for awarding organisations to review GCSE Computing specifications in light of developments in data science. In particular, they should consider introductory content on machine learning and computational modelling, and strengthening content on data representation.

Science

The Key Stage 3 science curriculum is set out in a similar way to that at Key Stage 2, separating data use and application of mathematics from statements on specific subject content. As a result, it is difficult to establish with certainty what the science curriculum at Key Stage 3 and Key Stage 4 delivers in terms of core knowledge and skills relating to data science, as this will depend on teaching.

Nonetheless, it is a strength that data is positioned as central to the secondary science curriculum, which states that *quantitative analysis is a central element both of many theories and of scientific methods of inquiry* and requires the *application of mathematical knowledge to the understanding of science*, setting out detail on this area.

A further strength is that at both Key Stage 3 and Key Stage 4 the science curriculum emphasises the classic data cycle of collecting, presenting and analysing empirical data. This stresses important aspects of data literacy through requirements to:

- *present reasoned explanations, including explaining data in relation to predictions and hypotheses* and *evaluate data, showing awareness of potential sources of random and systematic error;* and
- *identify further questions arising from their results*.

The cycle focuses predominantly on the analysis of datasets derived from the students own measurements and empirical research rather than secondary datasets.

Humanities and social sciences

The aims of geography at Key Stage 3 reflect those of the primary curriculum in relation to the skills needed to 'collect, analyse and communicate with a range of data'. Additionally, however, students are required to *interpret a range of sources of geographical information, including Geographical Information Systems (GIS)* and *communicate geographical information in a variety of ways, including through numerical and quantitative skills*.

Much of Key Stage 3 geography content is amenable to using and analysing quantitative data, including analysis from and of large secondary datasets. As is commonly the case, however, this is left to the discretion of teachers.

Humanities and social sciences are not compulsory as such at Key Stage 4, though students have an entitlement to study either GCSE Geography or GCSE History. History includes very little quantitative analytical content; most of the action is in Geography.

Every GCSE Geography specification includes the application of mathematics and analysis of data, setting out these requirements in detail. They also include content and domain 'problems' amenable to the exploration of data. In particular, specifications include in-depth case studies, setting out topics that offer good opportunities for students to use real datasets and their analysis.

There are however differences between the exam boards in their approach to applying mathematics to geography – each with strengths. OCR is the only exam board to emphasise competence in using GIS. Others are specific about links to publicly available data sources in relation to geographical issues, such as data on the economic development of countries.

Data analysis in GCSE Psychology is integrated into empirical/experimental psychological research methods, much in the same way as the natural science data cycle. Related mathematics content is drawn from the statistics strand of mathematics at Key Stage 3 and GCSE. GCSE Psychology specifications differ in the level of mathematical challenge and role of data. AQA's qualification places greater emphasis than others on knowledge of psychological theory and has less emphasis on practical research and psychological investigations than other qualifications.

5.3. Sequencing and coherence

Computational modelling of real problems and systems at Key Stage 3 is a significant step up from Key Stage 2 computing content and as such demands skilled teaching. A lack of detail in what should be taught within this is a worry and could lead to poor teaching, misconceptions and gaps in students' understanding. The required level of challenge is also unclear from curriculum statements, leading to a risk that, inappropriately, students tackle complex ideas in a shallow way and fail to develop more fundamental concepts.

Ideally, work is needed to establish the detailed and effective teaching sequences required at Key Stage 3 for introducing computational modelling, with the aim of building a secure grasp of fundamental concepts and skills in this area.

While Key Stage 3 computing feels like a step up, data science elements in GCSE Computing feel like a step down. It may be that, due to poor teaching and lack of formal assessment at Key Stage 3, the GCSE has to help students play 'catch-up' in computing. There is some evience of repeated content from Key Stage 3 – in one GCSE example certain taught algorithms are simply repeated. Content relating to data science – data handling, algorithms, databases and the like – feels dated in the context of web transactions, modern software, big data and a range of related developments.

The focus in science at Key Stages 3 & 4 on the classic data cycle is a strong feature of the curriculum. As indicated earlier, the datasets arising from this kind of inquiry are likely to be small univariate or bivariate sets to which limited analysis can be applied. This is not a problem in its own right, but there is a case for the classic data cycle to be expanded to reflect that other people's data and models might be used to answer the question of interest.

While Key Stage 4 mathematics and science both include content to help students make accurate conclusions from data, there is insufficient focus in curriculum statements across quantitative subjects on appraising and critiquing claims based on data. It's unlikely that students gain a coherent foundation in this important aspect of data literacy at this stage.

6. Barriers to coherent integration of data science

6.1. Barriers at primary level

The main *curriculum* challenges to the coherent integration of data science have been outlined above. The summary below highlights key issues.

In terms of curriculum design (as opposed to implementation) these appear to be:

- A lack of specificity in the computing programme of study at Key Stage 2, meaning that there is little detailed content set out on:
    - algorithms
    - presentation and analysis of information and data using technology.
- A potential disjunction in the Key Stage 2 science curriculum (depending on how taught) between:
    - statements about the importance of applying mathematics; and
    - scientific curriculum topics and associated guidance on working scientifically.

Further barriers relate to making coherent links between mathematics and computing – this is left to teachers. Primary teachers can sometimes lack confidence in mathematics and numeracy and generally lack experience in conceptual and technical aspects of computing. Putting these issues together, primary teachers could find it difficult to judge the best ways to integrate mathematics into computing as well as science without helpful support and specific guidance.

6.2. Barriers at secondary level

A major curriculum challenge at secondary level is an apparent lack of systematic *progression* in aspects of computing:.

- There is some evidence of overlap in content relevant to data science between Key Stage 3 curriculum statements and GCSE specifications.
- The introduction of *computational modelling of real situations and physical systems* at Key Stage 3 is (on the face of it) challenging – difficult to study in a meaningful way at this level (not to suggest that it should not be covered, but needs careful thought).
- It is not clear now *computational modelling of real situations and physical systems* is built on at GCSE.
- GCSE Computing qualifications appear to neglect introducing important aspects of content related to data science, such as machine learning and online transactions.

A likely further barrier to integrating the use of secondary datasets in the humanities, sciences or social sciences is, arguably, the complexity of linking taught content with relevant data or data analysis. Specifically:

- Much real data may be too complex to bring into the sphere of secondary curriculum study.
- Relevant questions at Key Stages 3 & 4 may be at a level at which little useful data exists, as they are long-resolved.

6.3. How can barriers to data science be overcome?

Most of the curriculum barriers highlighted and discussed in this report can be overcome without recourse to potentially unwelcome curriculum change. The issues relating to lack of specificity in the computing curriculum can be overcome through more unified and coherent curriculum – based on, for example, a curriculum panel with a role to provide guidance to schools. This will have greater impact if endorsed by government (this could act like current endorsement of the PSHE curriculum published by the PSHE Association).

Integrating modern data science and making it relevant to curriculum study at secondary level is likely to depend on the creation of good 'intermediary' datasets and models with conceptual veracity in the secondary curriculum, and which can be brought into the data-analytical or mathematical sphere of students.

6.4. Support for teaching and enrichment

There is no doubt that for secondary teachers of numerate subjects, the realm of 'real data science' is something unfamiliar to them. To advance their understanding, support and input will be needed, including detailed curriculum-matched examples of the use and analysis of data.

For primary teachers, the central need is for detailed guidance and support to understand how to apply mathematics in other areas of the curriculum and to teach relevant computing concepts rigorously and systematically, making coherent links across to mathematics.

7. Additional opportunities for data science in the curriculum

7.1. Opportunities at primary

The main additional opportunities for data science in the curriculum relate to:

- The application of mathematics and use of data in specific areas of the science curriculum to enhance subject understanding.
- Using meaningful data in computing when applying technology in the analysis and presentation of data – linked to other areas of the curriculum.

There is also potential to enrich the curriculum at primary through engaging infographics and visualisation, drawing on real data to enhance the study of knowledge-based subject curricula – notably, history and geography.

Science subject understanding

There are several areas where technologies, tools and approaches of data science can enhance subject understanding, particularly where concepts may be abstract – generally but not exclusively at higher Key Stage 2. Examples might be:

- At year 4 – pupils are asked to find patterns between the volume of a sound and the strength of the vibrations that produced it, and recognise that sounds get fainter as the distance from the sound source increases. Technology tools can be used to measure and record these data efficiently and in volume. Pupils might analyse the data in basic ways, but pre-defined macros could be used to represent the relationships between these data graphically to aid understanding.
- At year 5 – content relating to air resistance and friction. Pupils are asked to identify the effects of air resistance, water resistance and friction, that act between moving surfaces. These effects are difficult to study directly at Key Stage 2, offering little scope for the collection and analysis of data. Examples of simplified aerodynamic modelling and data analysis could be helpful to both support subject understanding and introduce the notion of computer modelling.

Computing – analysis and presentation of meaningful data

At primary level, teaching in this area can lack imagination, focusing on creating tables and simple charts in documents and presentations. This could be enhanced in two ways:

- Constructing an algorithm to translate a small dataset according to a rule (for example, multiply by x and add a constant) to solve a mathematics problem or answer a question.
- Taking a set of data from another curriculum area – for example, a class geography investigation – and using visualisation software to present different representations of it, discussing their appropriateness and merit.

Use of infographics for knowledge-based curricula

History and geography are rich in data. While the topics studied at primary level represent a simple subject grounding, some areas of study could easily draw in existing data and information. This could be presented in memorable ways through the use of quantitative infographics. Examples of curriculum relevant areas at Key Stage 2 include:

- Land-use patterns in the UK, understanding how some of these aspects have changed over time; and
- Human geography: types of settlement and land use, economic activity including trade links.

7.2. Opportunities at secondary

Conducive content in biology, geography and history

The secondary curriculum offers a wide range of possibilities for incorporating real uses of data science.

Biology content at Key Stage 4/GCSE is particularly amenable to incorporating the content from data models. While students would not be able or expected to engage with the analysis underpinning such models, they could start to build understanding of how datasets are combined and used, what variables are reflected in this, and how, *at a high level*, scientists use such data to develop and test hypotheses and draw valid conclusions.

Relevant content includes:

- The relationship between health and disease;
- The importance of biodiversity;
- Measures of distribution, frequency and abundance of species within a habitat; and
- Positive and negative human interactions with ecosystems.

The in-depth case study in GCSE Geography is also ripe for incorporating real uses of data science. A good area of study is economic development at country level. There are many quantitative indices relating to this, backed up by published datasets which include, for example: gross national income; birth and death rates; infant mortality; literacy rates; and life expectancy. Students could explore real models constructed by geographers and economists or apply analysis from the mathematics curriculum to interrogate simplified data to answer questions.

GDS data and GIS in appear to be relatively under-utilised in the study of geography. Given the ease of use of these systems and availability of large amounts of data (for example from web transactions) there is significant scope for students to develop geographic investigation using related datasets.

Reflecting the nature of the subject, there are fewer opportunities for bringing data analysis directly into the history curriculum. There are, however, opportunities in history to explore the relationship between data and historical claims and therefore foster data literacy – for example, claims relating to historical trends over time or causes of events (for example, the Great Depression).

Annex 1

Development of data science through the primary curriculum				
Key Stage & Year	Elements of data science			
	Mathematics	Computing	Science	Humanities
KS1 Y1 & Y2	Categorisation, tallying and quantifying.	Understand what algorithms are, how they're implemented as programs and that they follow precise and unambiguous instructions. Create and debug simple programs.	Practical scientific methods, processes and skills, including identifying and classifying and gathering and recording data to help in answering questions. (Non statutory Y1): Make tables and charts about the weather; and displays of what happens in the world around them, including day length, as the seasons change.	No elements of data science are specified in geography at KS1. No elements of data science are specified in history at KS1.
KS2 Y3	Quantities in pictograms, bar charts and tables; answer one- and two-step questions. Interpret data 'presented in many contexts'.	Design, write and debug programs that accomplish specific goals, including controlling or simulating physical systems; solve problems by decomposing them into smaller parts. Use sequence, selection, and repetition in programs; work with variables and various forms of input and output. Use logical reasoning to explain how some simple algorithms work and to detect	'Working scientifically' is described separately but must always be taught through and clearly related to substantive science content. It includes: Gathering, recording, classifying and presenting data in a variety of ways to help in answering questions. Recording findings using simple scientific language, drawings, labelled diagrams, keys, bar charts, and tables. (Non-statutory Y3/4) With help, pupils should look for changes,	No elements of data science are specified in geography at KS2. No elements of data science are specified in history at KS2.
KS2 Y4	Discrete and continuous data in bar charts and time graphs, inc. comparison, sum and difference problems.			

KS2 Y4 cont.		and correct errors in algorithms and programs. Select, use and combine a variety of software (including internet services) on a range of digital devices to design and create a range of programs, systems and content that accomplish given goals, including collecting, analysing, evaluating and presenting data and information. Use technology safely, respectfully and responsibly.	patterns, similarities and differences in their data in order to draw simple conclusions and answer questions. (Non-statutory Y3/4) With support, they should identify new questions arising from the data, making predictions for new values within or beyond the data they have collected. (Non-statutory Y4) Pupils might observe and record evaporation over a period of time, for example, a puddle in the playground.	
KS2 Y5	Comparison, sum and difference problems inc. continuous data in line graphs. Decide which representations of quantitative data are most appropriate and why.		'Working scientifically' is described separately but must always be taught through and clearly related to substantive science content. It includes: Recording data and results of increasing complexity using scientific diagrams and labels, classification keys, tables, scatter graphs, bar and line graphs. (Non-statutory Y6) Pupils might work scientifically by exploring scientific research about the relationship between diet, exercise, drugs, lifestyle and health.	
KS2 Y6	Interpret and construct pie charts and line graphs to solve problems. Calculate and interpret the mean as an average and know when it is appropriate to find the mean of a data set. Draw graphs relating from two variables arising from own enquiry and other subjects.			

Development of data science through the secondary curriculum				
Key Stage & Year	Elements of data science			
	Mathematics	Computing	Science	Social Science/Humanities
KS3 Y7, Y8 & Y9	Describe, interpret and compare observed distributions of a single variable through: appropriate graphical representation involving discrete, continuous and grouped data; and appropriate measures of central tendency (mean, mode, median) and spread (range, consideration of outliers). Construct and interpret appropriate tables, charts, and diagrams, including frequency tables, bar charts, pie charts, and pictograms for categorical data, and vertical line (or bar) charts for ungrouped and grouped numerical data Describe simple mathematical relationships between two variables (bivariate data) in observational and experimental contexts and illustrate using scatter graphs.	Design, use and evaluate computational abstractions that model the state and behaviour of real-world problems and physical systems Understand several key algorithms that reflect computational thinking [for example, ones for sorting and searching]; use logical reasoning to compare the utility of alternative algorithms for the same problem. Use two or more programming languages to solve a variety of computational problems; make appropriate use of data structures [for example, lists, tables or arrays]. Understand a range of ways to use technology safely, respectfully, responsibly and securely, including protecting online identity and privacy.	(KS3 and KS4) Apply mathematical knowledge to the understanding of science, including collecting, presenting and analysing data. Pupils should: - apply mathematical concepts and calculate results; - present observations and data using appropriate methods, including tables and graphs; - interpret observations and data, including identifying patterns and using observations, measurements and data to draw conclusions; - present reasoned explanations, including explaining data in relation to predictions and hypotheses; and - evaluate data, showing awareness of potential sources of random and systematic error; identify further questions arising	KS3 Geography: develop greater competence in using geographical knowledge, approaches and concepts [such as models and theories] and geographical skills in analysing and interpreting different data sources. Use Geographical Information Systems (GIS) to view, analyse and interpret places and data. Use fieldwork in contrasting locations to collect, analyse and draw conclusions from geographical data, using multiple sources of increasingly complex information. No elements of data science are specified in History at KS3 PSHE is not statutory but DfE is consulting on this. The recommended curriculum at KS3 includes: The safe and responsible use of technology, including safe management of own and others' personal data.

KS3 Y7, Y8 & Y9 cont.			from their results. Under 'measurement' students should: - use and derive simple equations and carry out appropriate calculations; - undertake basic data analysis including simple statistical techniques.	Understand and appreciate the importance of following cybersecurity protocols in the workplace and the importance of the data protection act in the workplace.
KS4/GCSE YR10 & YR11	Infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling interpret and construct tables and line graphs for time series data Higher tier: {construct and interpret diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class intervals and cumulative frequency graphs, and know their appropriate use} Interpret, analyse and compare the distributions of data sets from univariate empirical distributions through: - appropriate graphical representation involving discrete, continuous and grouped data, {including box plots}	Develop and apply analytic, problem-solving, design, and computational thinking skills. Understand how changes in technology affect safety, including new ways to protect online privacy and identity. Beyond this, further Computing content is not compulsory at KS4. Optional content is set out in individual GCSE specifications and varies between exam boards. GCSE Computing Most specifications are divided into a theoretical paper focused on the fundamentals of computer systems and programming; a practical investigation (controlled assessment) on an area of	Develop and learn to apply observational, practical, modelling, enquiry, problem-solving skills and mathematical skills, both in the laboratory, in the field and in other environments. Apply the cycle of collecting, presenting and analysing data, including: - presenting observations and other data using appropriate methods - translating data from one form to another - carrying out and representing mathematical and statistical analysis - representing distributions of results and making estimations of uncertainty - interpreting observations and other data, including	History and Geography are not compulsory at KS4 but all pupils in maintained schools have a statutory entitlement to be able to study at least one humanity. Content for these subjects is set out in individual GCSE specifications and varies between exam boards. No elements of data science are specified in GCSE History GCSE Geography All geography specifications include the application of mathematics and analysis of data, setting out these requirements explicitly, mainly drawing from the statistics strand of KS3 and GCSE mathematics.

KS4/GCSE **YR10 & YR11 cont.**	- appropriate measures of central tendency (including modal class) and spread {including quartiles and inter-quartile range} Apply statistics to describe a population. Use and interpret scatter graphs of bivariate data; recognise correlation and know that it does not indicate causation; draw estimated lines of best fit; make predictions; interpolate and extrapolate apparent trends whilst knowing the dangers of so doing.	computing in depth; and a programming project (again, assessed through a controlled assessment). Content on data handling and analysis in all three GCSEs is limited. This generally focuses on low-level (e.g. machine-level) and technical (e.g. database management) aspects or repeating KS3 content on algorithms for sorting, searching etc.	identifying patterns and trends, making inferences and drawing conclusions - presenting reasoned explanations, including relating data to hypotheses - being objective, evaluating data in terms of accuracy, precision, repeatability and reproducibility and identifying potential sources of random and systematic error.	All specifications include in-depth case studies, setting out a range of topics which offer opportunities for students to draw on published data or statistics. <u>Psychology and Sociology</u> There is no entitlement to study Psychology or Sociology at GCSE, but these qualifications are popular. There is little evidence of elements of data science in <u>GCSE Sociology</u>, but the qualification includes the use of social statistics to explore sociological issues and there is potential to include analysis of basic of datasets. Data analysis features in <u>GCSE Psychology</u>, generally integrated into empirical/ experimental psychological research methods. Specifications set out the mathematics content applied in the analysis of research data, which is drawn from the statistics strand of mathematics at KS3 and GCSE.