Teach Computing

Report

Data and Information within the Computing Curriculum

June 2022

Teach Computing

Raspberry Pi Foundation

Summary

This report explores the Data and Information strand of the national curriculum. It explores the importance of both data science and data management and their relevance to learners as these concepts underpin a number of computer systems learners will experience.

The report aims to summarise the approach taken to data and information by the National Centre for Computing Education (NCCE). It presents a research-based view of data and information processes whilst acknowledging the role of data and information throughout computing and across the curriculum.

The report also highlights progression from the beginning of primary school education all the way through to when learners leave school. The report includes supporting resources and further reading for those looking for more information, courses, or classroom content.

The report consists of six sections:

- Section 1 discusses the role and importance of data and information within the curriculum. It considers the role of data in society and its prevalence in industry across all sectors, including up-and-coming areas such as artificial intelligence (AI) and machine learning. Beyond this, it discusses the areas of data literacy most pertinent to computing: data management and data science.
- Section 2 examines the current computing curriculum for England and the aims and objectives that relate to data and information. This includes data and information content that appears in other subjects within the national curriculum.
- Section 3 details the NCCE's implementation of the data and information aspects of the computing national curriculum. The structure of the Teach Computing Curriculum (TCC) is explored alongside an overview of the experiences provided across all five key stages.
- The final part of section 3 is all about progression; each unit relating to data and information is summarised and mapped against data life cycle processes. These key stage by key stage summaries are further collated and presented as a progression of skills and concepts from age 5 to 19.
- Pedagogy is the focus of section 4, in which pedagogy principles that relate to data and information are explored and examples of specific practices provided.
- Section 5 provides a summary of relevant professional development sessions and courses provided by NCCE programmes.
- To conclude, section 6 provides an overview of the key messages of this report.

Contents

1. Introduction

1.1. Data and Information

There is no question that data and information is playing an ever increasing role in the lives of individuals, organisations, and of course wider society. However, the concept of data is not new. The word 'data' itself has Latin etymology, and is the plural of the word 'datum', meaning 'thing'. The term became more widely used in the 1950s with the development of computer systems, where it described the information stored on computers.

The huge growth in computing has resulted in data becoming vital across society. The use of data is pervasive across all sectors and economies around the world, to analyse, evaluate, and inform decisions. The sheer volume and richness of data being collected and analysed, alongside the availability of high-performance, low-cost hardware, has meant that artificial intelligence (AI) and machine learning (ML) technologies are increasingly being applied to data to provide automated analysis and, in some cases, decision making. This increased use of data means that individuals with specialist data skills are in demand.

This relationship with data isn't limited to industry — it filters down and permeates almost every aspect of our lives. Many of our daily activities are influenced by data and information: from looking up a weather forecast so you can decide what to wear, to making important financial decisions, or simply choosing what to watch next on your chosen streaming service. To participate effectively in modern society, data skills and an awareness of how data is used are essential for all citizens; this creates an important goal for our education systems.

A particular challenge that stems from the ubiquitous use of data across a wide range of industries, disciplines, and contexts is that there are a variety of terms and definitions used to describe data skills.

To explain the relevance and importance of data, this report adopts a hierarchical model combining elements of literature from governments, organisations, and research. This model presents data from its widest application to its specific place in the computing curriculum.

Data literacy

There is a wide consensus that to be 'data literate' means you have the understanding and skills necessary to collect, analyse, and interpret data. In the UK, the Department of Digital, Culture, Media, and Sport recognised the importance of data literacy in their National Data Strategy[1] policy paper from December 2020:

> [...] data-literate individuals are more likely to benefit from and contribute to the increasingly data-rich environments they live and work in, while data-driven companies can deliver significant productivity benefits to their own business and the wider economy.

In its widest sense, being data literate is a skill that transcends almost all aspects of life in modern societies. It means possessing the skills and knowledge to recognise when and where data is influencing your life and how you can use data to your advantage.

[1] Department of Digital, Culture, Media, and Sport. *National Data Strategy*. UK Department of Digital, Culture, Media, and Sport. 2020. https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy [accessed 16 June 2022]

A recent Open Data Institute (ODI) report[2] recognised the importance of data literacy in business and the wider economy. Their Data Skills Framework[3] categorises data skills into two main areas:

- Technical and practical – foundation, engineering, and analysis
- Critical, contextual, and strategic – business, management, and leadership

In order for an economy to thrive in the digital age, high levels of data literacy in young people are essential so that they are equipped with the skills to succeed in a data-driven digital economy. Data literacy is identified by the ODI as being at the same level of importance as broader digital literacy, media literacy, and the emerging field of AI literacy.

Figure 1: ODI Data Skills Framework.

[2] Freeguard, G, Zimeta, M, Tarrant, D, & Mezeklieva, V. *Data literacy and the UK government*. Open Data Institute. 2022. http://theodi.org/wp-content/uploads/2022/04/OPEN_REPORT_Data-literacy-and-the-UK-government_ODI_2022-04-3.pdf

[3] Open Data Institute. *Data Skills Framework*. 2020. https://theodi.org/article/data-skills-framework/#1561999285194-8defa99a-39a52584-8cfe [accessed 16 June 2022]

Data and Information

Data and Information is a strand of computing[4,5] that focuses on how data is stored, organised, and used to represent the real world and provide meaningful insights.

It broadly maps to the “technical and practical” areas of the ODI model, specifically the foundation, engineering, and analysis categories. Whilst these categories are studied across several other subjects, particularly maths and science, these subjects focus on statistical concepts, experimentation, and interpretation of data. At almost any scale, most data collection, storage, and analysis will involve computers. As such, computing complements other subjects, with a practical focus on collection, storing, and analysing data at scale.

The Data and Information strand of the computing curriculum also provides a solid basis for any learner to understand the role of AI in analysing and interpreting data. The difference being that in traditional analysis a computer acts as a tool to aid the analysis with instruction from a human, whereas in AI analysis the computer can conduct its own analysis and draw conclusions.

In developing their data literacy competency model, Grillenberger and Romeike[6] reviewed research into skills and concepts within this strand of computing, although they apply the label “data literacy”. These skills and concepts came from two areas described as **Data Management**, which “focuses on rather static aspects related to data, in particular on how they are stored and accessed appropriately”, and **Data Science**, which, on the other hand, “sets its focus on the rather dynamic aspects, such as data analysis and visualisation”. Their research draws on past work outlining a data life cycle[7], which sequences the activities that are typical to a data science cycle.

From this life cycle, they propose a competency model that groups life cycle stages into four distinct process areas, each underpinned by content knowledge or concepts. This model serves as a useful lens through which to explore progression within the Data and Information strand of the computing curriculum in England.

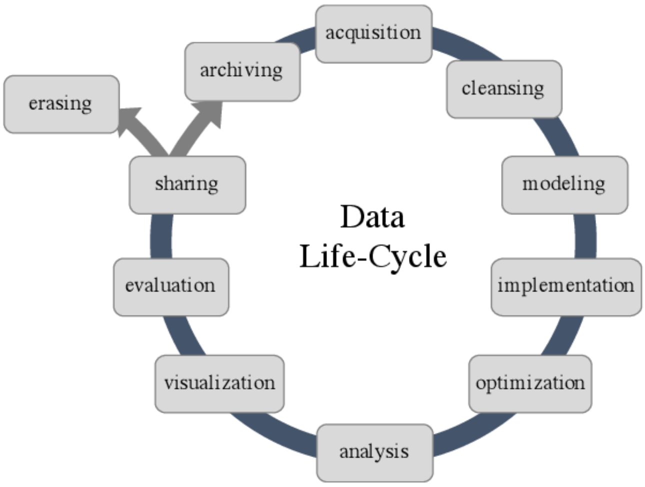


Figure 2: The data life cycle proposed by Grillenberger and Romeike[6].

[4] National Centre for Computing Education. *Categorising content across the National Centre Taxonomy*. 2020. https://blog.teachcomputing.org/categorising-national-centre-content/ [accessed 23 June 2022]

[5] Caspersen, M, Diethelm, I, Gal-Ezer, J, McGettrick, A, Nardelli, E, Passey, D, Rovan, B, Webb, M. *Informatics Reference Framework for School*. Informatics for All. 2022. https://www.informaticsforall.org/wp-content/uploads/2022/03/Informatics-Reference-Framework-for-School-release-February-2022.pdf

[6] Grillenberger, A, & Romeike, R. *Developing a theoretically founded data literacy competency model*. In *WiPSCE ‘18: Proceedings of the 13th Workshop in Primary and Secondary Computing Education*. New York: Association of Computer Machinery. 2018. Article no. 9. https://doi.org/10.1145/3265757.3265766

[7] Grillenberger, A, & Romeike, R. *Key Concepts of Data Management: An Empirical Approach*. In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*. New York: Association of Computer Machinery. 2017. pp. 30–39. https://doi.org/10.1145/3141880.3141886

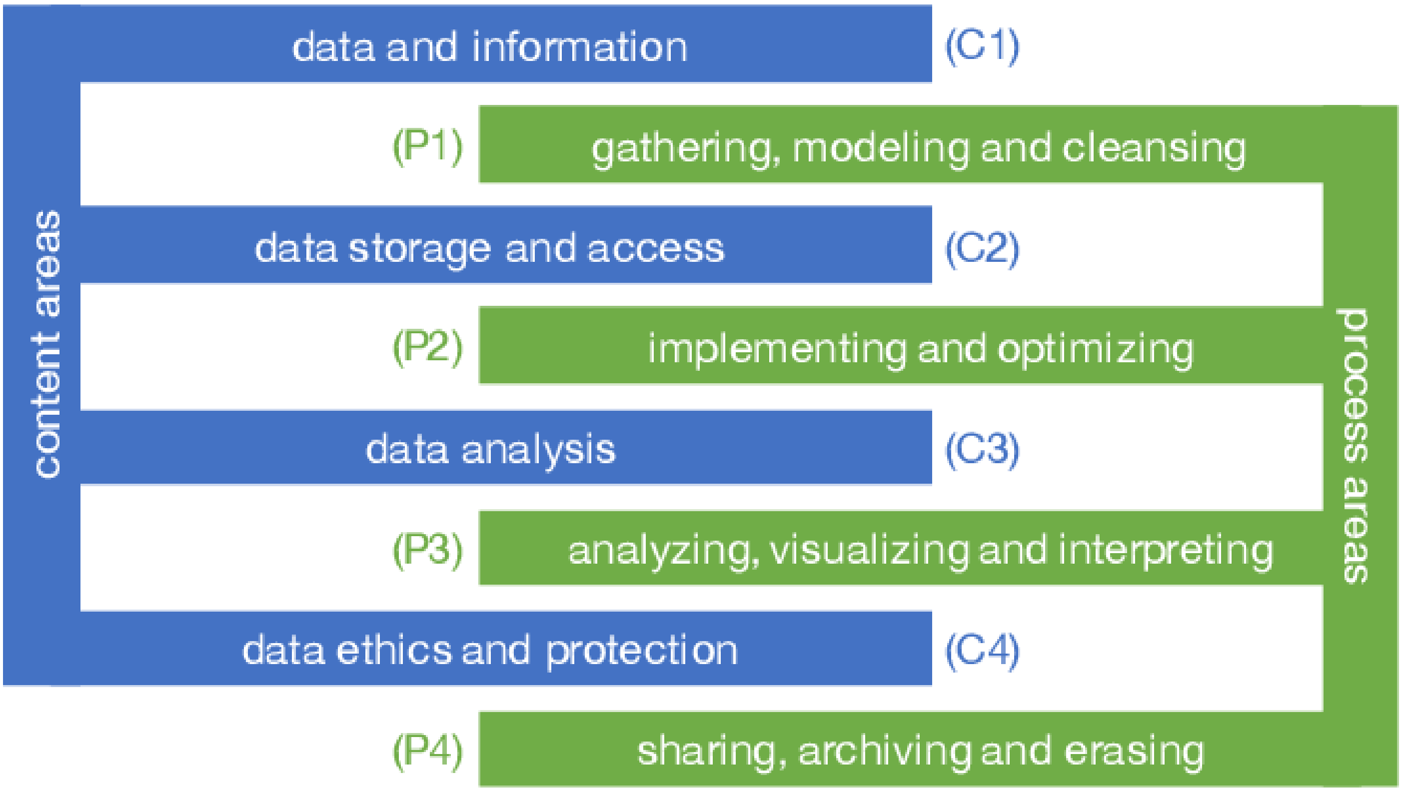


Figure 3: Data processes and content model developed by Grillenberger and Romeike[6].

1.2. The National Centre for Computing Education

The role of the National Centre for Computing Education (NCCE) is to support teachers to deliver the **entire computing curriculum**. A central part of this role has been the development of the Teach Computing Curriculum[8], which offers teaching resources for each stage of learning.

The NCCE taxonomy

Content from the NCCE is built upon a 'taxonomy', which helps organise the breadth of the subject of computing. This categorisation consists of ten strands that span the current national curriculum for computing in England. Each strand has a combination of skills and concepts that feature throughout the national curriculum. Six of these strands describe broad areas of study within computing. The other four strands describe cross-cutting concepts and skills that are taught and revisited through the first six (Figure 4).

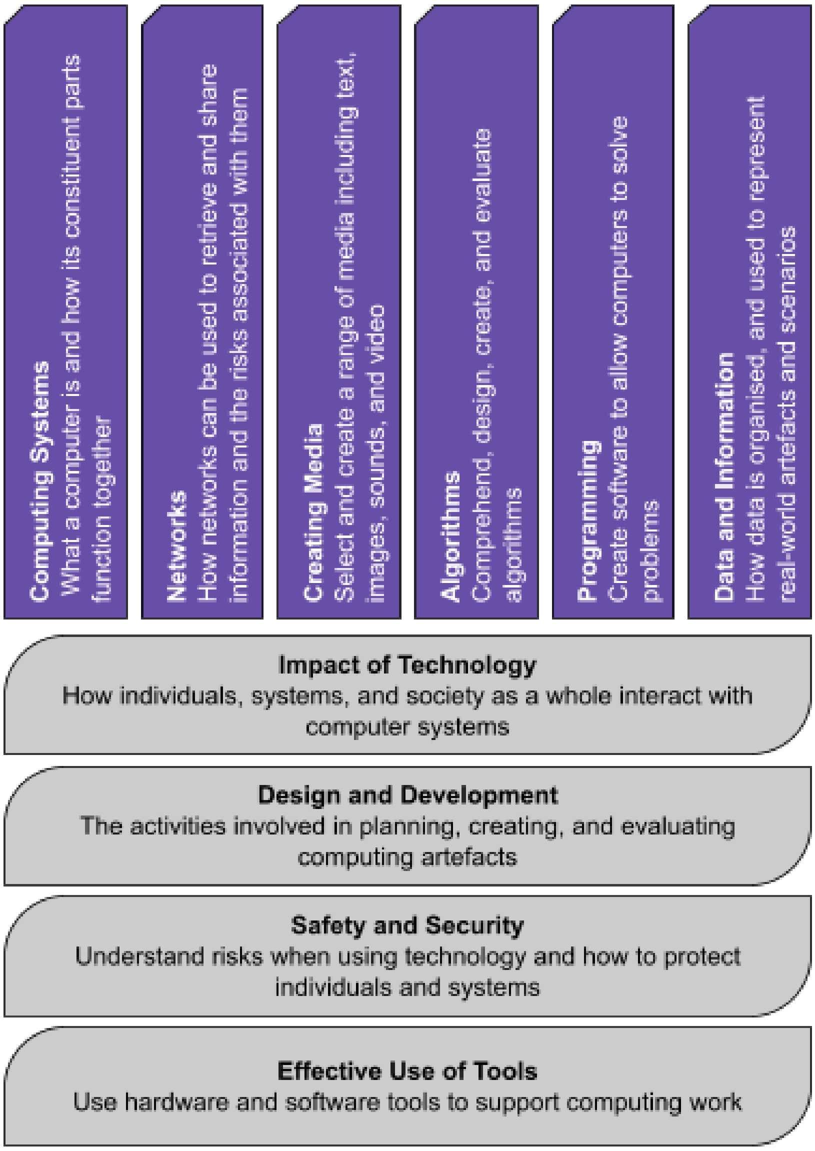


Figure 4: The ten computing strands and their relationship to each other.

[8] National Centre for Computing Education. *Teach Computing Curriculum*. https://teachcomputing.org/curriculum [accessed 1 June 2022]

Together, the Teach Computing Curriculum and Isaac Computer Science[9] cover the teaching of computing and computer science from key stage 1 to 5 (5- to 19-year-olds). Both have been expertly designed with progression in mind and exemplify our approach to sequencing concepts and skills.

Data and Information in the taxonomy

The **Data and Information** strand is where the majority of skills and concepts associated with either data science or data management can be found. There are aspects of data management in particular which relate to, or belong entirely in, other strands. One such area is the concepts and skills related to how different media are represented ultimately as binary digits. This knowledge helps inform the use and application of different media files and as such is part of the **Creating Media** strand. In learning how binary data is physically stored in memory on a storage medium, learners are exploring an area of **Computing Systems**. Similarly, transmission of data forms part of **Networks**.

There is also overlap with each of the four horizontal strands: learners will learn about the function and features of different tools (**Effective Use of Tools**) for manipulating data; **Design and Development** processes will apply when developing data storage solutions or designing a visualisation; and like every area of computing, learners must consider the **Impacts** associated with data and information as well as questions of **Safety and Security**.

This report, which is part of a series of NCCE reports, explores progression within the Data and Information strand. Its purpose is to outline the ways in which the NCCE can support you with all aspects of the teaching and learning of data and information. It has been written in relation to the curriculum in England, although you may also find it interesting if you are reading this from another context. The intended audience is all serving teachers, prospective teachers, and educators involved in teaching computing.
For more information on other themes not covered in this report and how they are addressed in the Teach Computing Curriculum, please refer to our previous reports[10] or teacher guides for your key stages[8]. In each guide, we describe how units are structured, the progression within and between units, as well as emphasising appropriate pedagogical approaches.

[9] National Centre for Computing Education. *Isaac Computer Science*. https://isaaccomputerscience.org/ [accessed 1 June 2022]

[10] National Centre for Computing Education. *Teach Computing*. https://teachcomputing.org/impact-and-evaluation [accessed 1 June 2022]

2. Data and information in the national curriculum and beyond

2.1. Data literacy in the curriculum

Data literacy is relevant across many subjects in the national curriculum, not just computing. The most prominent links are in mathematics and science.

Data literacy in the maths curriculum

In mathematics, collecting, presenting, and analysing data graphically, in the form of tables and pictograms, are introduced to learners aged 6–7 in the statistics part of the maths curriculum. From age 7–8, bar charts are introduced together with a greater emphasis on questioning. From ages 8–10, a wider range of analysis is taught. Other types of charts such as line and pie charts are introduced to learners aged 10–11.

From 11–14, learners extend their analysis skills further by distinguishing between different types of data (discrete, continuous, and grouped) and they present data in a broader range of formats such as frequency tables, histograms, scatter graphs, and more.

Beyond ages 11–14, learners can specialise in statistics through GCSE and A level qualifications. These explore the stages of the statistical enquiry cycle in detail: starting with a problem, planning a study, collecting data, conducting analysis, and finally drawing conclusions.

Data literacy in the science curriculum

The concept of data is first mentioned in the science curriculum at ages 5–7; it is mainly framed around asking and answering questions and classifying objects. By ages 7–9, children are expected to gather data in different ways (including the use of data loggers), present data in a variety of forms (in line with the mathematics curriculum), draw conclusions, and make predictions. Children aged from 9–11 record data with more accuracy, present it in more advanced forms, and across different platforms.

Secondary school aged learners continue to collect experimental data, perform analysis, identify trends and outliers, and draw conclusions. More time is spent focused on the importance of data for science, as well as evaluating results and their underlying data.

Data literacy in other areas of the curriculum

Beyond mathematics and science, data science also features in the geography and physical education curricula for learners aged 11–14, in relation to analysing geographical survey data and performance data, respectively.

The section below will analyse which of the skills described above are applicable to computing, and explore the breadth of the computing sections of the national curriculum.

2.2. Data and information

Whilst data literacy can be seen to be present across many areas of the curriculum, computing focuses on how computers are used to collect, analyse, and visualise data (data science) as well as how data is stored and organised (data management). We can see the importance of data, as well as the balance of data science versus data management, within the computing programme of study.

Key stage 1

In this early phase of education, the programme of study highlights learners being able to "organise, store, manipulate [...] digital content", which will include data. There is also a focus more generally on safety and privacy, including "keeping personal information private". Beyond this, learners of this age are encouraged to be curious about the world around them and "recognise common uses of information technology beyond school" — many of these common uses include data such as the tills in shops.

Key stage 2

By this stage, there is explicit reference to learners "collecting, analysing, evaluating, and presenting data and information" and combining a "variety of software" to achieve these goals. There is also a focus on considering the accuracy of information, where learners should "be discerning in evaluating digital content".

Key stage 3

As students begin secondary education, they will "model the state and behaviour of real-world problems", which will include data models. They begin to consider an aspect of data management by considering how "numbers can be represented in binary" and stored in memory. They are expected to "select and combine multiple applications" in varied projects that include "collecting and analysing data". Building on prior learning and with more online services available to them, they also consider how they can go about protecting their online identity and privacy.

Key stage 4

Although brief, the statutory guidance for computing at KS4 requires learners to be given the opportunity to "develop and apply their analytic [...] skills".
Beyond the expectations the national curriculum has for all learners, qualifications at GCSE and A level continue the progression of data and information, albeit with a more specialist focus.

GCSE computing

Despite some variation between examination boards, there are some key areas of focus at GCSE level, which predominantly focus on the data management aspects rather than data science. There is further exploration of how numbers, images, sounds, etc. are represented and stored as binary data as well as their representation as hexadecimal numbers, and how they can be converted or calculated in different number systems. The other area of data management that features to varying degrees is exploring how data is organised, stored, and retrieved using databases. Understanding how and where data is stored is also explored from a privacy and security perspective.

A level computing

The focus from GCSE level continues as learners are expected to work with and manipulate data represented in different number systems, including learning how to represent decimals and negative numbers. Databases are explored in greater depth, with a particular focus on their design, representation, and the efficiency of storage and access. Learners look beyond traditional databases to unstructured big data, data mining, predictive analysis, and the impact of data on society.

3. Data and Information within the Teach Computing Curriculum

3.1. Data and Information processes

Within the Data and Information strand, we adapted the competency model proposed by Grillenberger and Romeike[6] to ensure learners had exposure to the entire data life cycle. As described in section 1, this model comprises four process areas (P1–P4). Each process is dependent on a range of associated content areas that create a matrix of competencies. For the Teach Computing Curriculum, we took these processes and added a fifth (questioning), which we identified as being an important process that you need to make explicit with school-age learners. These process areas help us describe and sequence activities and learning. The five stages, labelled P0–P4, are described below.

P0 – Questioning

This process area is outside of those defined by Grillenberger and Romeike[6]; they did not include it as they considered it as the reason for the data analysis in the first place. For younger learners (age 5–11), we consider this to be an essential part of the progression in learning — children need to understand that the questions we ask determine the data that we collect. This links in with other areas of the curriculum, from maths investigations to science enquiry. Learning how to ask a good question in one area supports learning in another.

P1 – Gathering, modelling, and cleansing

The key skills we've associated with this area are deciding what data needs to be collected, how it should be collected, and then actually collecting it. For ages 5 to 11, the Teach Computing Curriculum introduces different strategies for collecting data, such as tally charts, putting an object in a jar to represent an answer, or using sensors connected to a computer to record live data. As learners move into secondary education, they begin to evaluate the quality and usefulness of the data they have collected. This can then lead to cleansing inaccurate or corrupt data. The aim is to develop learners' understanding of data collection processes, so that they become able to make decisions about the best method for a particular task.

P2 – Implementing, optimising, and structuring

This process area includes structuring and storing data so that it can be usefully worked with and analysed. For younger learners (age 5–11), this can include the headings and structure of a spreadsheet, appropriate labels for data in a database, and questions to build a branching database. Within secondary education, database structures progress to include relational databases and, at an advanced level, designing efficient database structures using nominalisation.

P3 – Analysing, visualising, and interpreting

The next area is where much of the computing element of data handling occurs: analysing, visualising, and interpreting. To help learners develop their data analysis skills, they are taught how to manipulate data to visualise it in different ways, such as creating a graph from data in a spreadsheet. Learners can also be asked questions about data in existing graphs, charts, and infographics, and should be encouraged to draw conclusions about data that they have collected and visualised.

P4 – Sharing, archiving, and erasing

The final category from the research is sharing, archiving, and erasing data. In this area, learners develop the skill of presenting the results of their data analysis work. This follows on from how the learners have visualised the data — while that involves thinking about how to show the data to enable it to be interpreted in the first place, in this stage learners think about how they can best communicate what they have found out with others. Learners can be provided with opportunities to share their work with peers, another class, or a wider audience via a class blog.

3.2. A data cycle

As shown in section 1, Grillenberger and Romeike[6] present their processes as part of a data cycle where one process leads to the next. Within the Teach Computing Curriculum, the same cycle has been used; however, it is presented as a more abstract version. This cycle is shown in Figure 5.

The stages in this cycle are synonymous with the processes outlined by Grillenberger and Romeike[6] (with the aforementioned inclusion of Questioning), however, a simplified model was easier to use across the wide age range that the Teach Computing Curriculum covers. With younger learners, each stage of the cycle is at least referred to in each unit — although there is often a predominant focus on one area of the cycle — whereas as learners get older, learning may be focused only on a singular stage of this cycle.

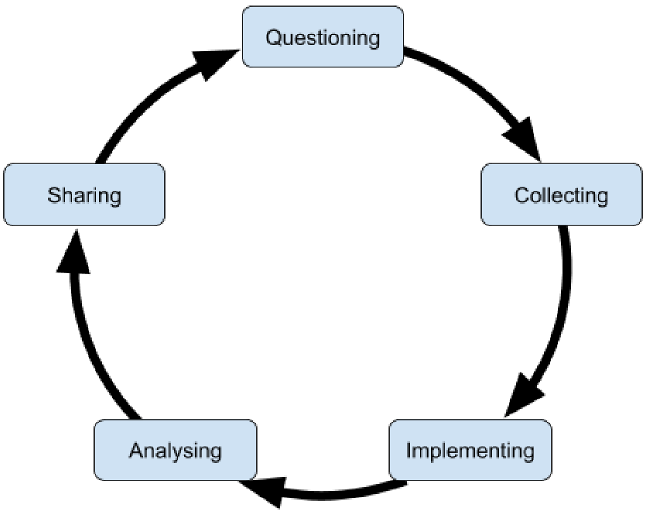


Figure 5: The data cycle used in the Teach Computing Curriculum.

3.3. Progression in Data and Information

To explore progression within the Data and Information strand, we consulted the objectives from the Teach Computing Curriculum for key stages 1 to 4 (ages 5 to 16), including the GCSE content. To examine progression further, we considered the curriculum areas covered by the Isaac Computer Science website, which represents all A level exam boards in England. Each objective is already categorised to the NCCE taxonomy.

Each of these objectives was then collaboratively categorised into one or more of the five process areas. This allowed us to gain a broad overview of the progression from each key stage to the next; in particular, we can observe that initially the focus is on introducing the full range of data handling skills, from collection through to presenting. From KS3 onwards, there is a gradual shift towards the more specialist computing processes, implementation in particular.

In the Teach Computing Curriculum, there are ten units of work, across four key stages, that focus specifically on data and information. Each one tackles a subset of the five process areas from the data handling cycle. For example, most primary units cover most of the processes, but not all. Generally, there is more emphasis on questioning and collection in KS1, with more implementation, analysis, and presentation in KS2.

Each key stage of the curriculum is different; some are longer stages than others, some are statutory while others are elective, and those resulting in a qualification generally involve many more teaching hours. As might be expected, each key stage includes a recap of concepts that have been encountered at an earlier stage. This is particularly noticeable at points of transition, such as when learners move from primary to secondary education or when learners choose to study for a GCSE or A level qualification in computing. Due to varying provision in different schools, prior knowledge cannot be assumed. This creates a degree of overlap between key stages. These differences and intersections make direct comparisons between key stages challenging. Instead, this report attempts to describe the focus and progression within each educational stage.

Key stage 1

At this stage, pupils begin to explore data handling processes. They begin by physically applying some structure to data by labelling and grouping it by attributes such as type, colour, or size. They then make basic comparisons between individual objects and groups of objects using language such as "more than" and "less than".

Pupils then move on to collecting and structuring data using digital tools with which they create tally charts and pictograms. In doing this, they develop their understanding of attributes and present data for the first time using a computer.

Table 1: Key skills and concepts at key stage 1.

Curriculum units	Key concepts and skills
Year 1: Data and information – Grouping data *Exploring object labels, then using them to sort and group objects by properties*	■ Questioning – asking and answering questions ■ Collecting – data collection and counting
Year 2: Data and information – Pictograms *Collecting data in tally charts and using attributes to organise and present data on a computer.*	■ Questioning – asking and answering questions ■ Collecting – structured data recording ■ Presenting – using a computer to present data in a pictogram

Key stage 2

At this key stage, learners cover four different tools for working with data and information to collect, structure, and analyse data. This includes using software to create branching databases and working with flat-file databases and spreadsheets. Learners also use data loggers to collect data automatically, which they later download to a computer to analyse.

Each unit of work has a different focus and therefore covers one or two of the processes mentioned in section 3.1. in more depth. The other processes will often be present within the unit, but may not be the main focus of learning.

A key difference with key stage 2 compared to other key stages is how frequently learners spend time working at each stage within the data life cycle. This helps learners develop a foundational understanding of each process, before specialising in particular stages as they get older.

Throughout this key stage, learners use a range of different tools and devices. Throughout the units, students ask and answer questions using existing data sets. They develop their own hypotheses and create questions they wish to investigate. As learners move through the data life cycle, they are encouraged to gain more independence in the types of questions that they can ask and the opportunities available to them.

Within key stage 2, learners move from small data sets, which they collect themselves over a short period of time, towards much larger data sets. This includes students using data loggers to collect data automatically over time. These data sets are then uploaded to computers for analysis by the learners. Beyond the use of data sets, learners also get to make use of a ‘real-world’ data set for the passengers on the Titanic and can begin to understand that data collection can involve making use of existing data sets.

As the volume of data increases, learners have to begin to consider how to implement structure in their chosen format. This ranges from considering what headings to make use of in a spreadsheet to using appropriate terminology in databases to identify which fields and records are required. Recognising how different solutions have been implemented within key stage 2 and discussing the choices that have been made is important for the extended focus on the Implementing process within key stages 3 and 4.

The final process is Sharing and this is where learners explore different visualisations to explain how they answered their questions and these can be incorporated into presentations as needed. They use tools to create graphs and charts, and use these visualisations to explain the conclusions they have drawn.

Table 2: Key skills and concepts at key stage 2.

Curriculum units	Key concepts and skills
Year 3: Branching databases *Building and using branching databases to group objects using yes/no questions.*	■ Implementing – choosing attributes that lead to an even split of objects ■ Analysing – using the tree structure to recognise the role of AND with each attribute ■ Presenting – representing objects in a branching database tool
Year 4: Data logging *Recognising how and why data is collected over time, before using data loggers to carry out an investigation.*	■ Questioning – developing hypotheses ■ Collecting – using an automated data logger ■ Analysing – reading data from a line graph
Year 5: Flat-file databases *Using a database to order data and create charts to answer questions.*	■ Questioning – asking and answering questions from a given data set ■ Implementing – understanding the structure of fields and records in a database ■ Analysing – using searches and filters (AND, OR, NOT) in a branching database ■ Presenting – creating charts and graphs in a database program
Year 6: Introduction to spread-sheets *Answering questions by using spreadsheets to organise and calculate data.*	■ Questioning – developing questions, including scenarios ■ Implementing – organising data in a structured way so that it can be represented in a spreadsheet and using formulae and functions ■ Presenting – creating charts and graphs and presenting them in other media

Key stage 3

For most pupils, moving into this key stage represents a moment of transition as they move to secondary school. To accommodate the wide variety of data skills and experiences that pupils from different schools bring, an initial focus of key stage 3 is to consolidate and reinforce the understanding of the core principles of data handling. It is at this stage that learners are also introduced to more data management skills through units focused on data representation.

The consolidation from KS2 begins with learners building on their knowledge of spreadsheets in year 7. Learners progress from creating basic formulae to applying more advanced formulae such as COUNTIF and SUMIF. This gives learners a conceptual understanding of some of the core features of spreadsheet applications, which they are also likely to apply in other areas of the curriculum.

Learners also explore data science in more depth. They learn how to use data to investigate problems in the world around them and attempt to identify useful insights or to suggest solutions. Learners are exposed to larger, real-world data sets and gain an understanding of how visualising data can help with the process of identifying patterns and trends.

Threaded throughout this stage are explorations of the impact of technology, in particular the benefits and challenges that come from sharing data. Learners also consider the importance of protecting their personal data and some approaches to doing so.

Table 3: Key skills and concepts at key stage 3.

Curriculum units	Key concepts and skills
Year 7: Modelling data – spreadsheets *Sorting and filtering data and using formulae and functions.*	■ Collection – explaining the difference between primary and secondary data ■ Implementation – structuring and organising data using a spreadsheet tool, making use of formulae to automate calculations ■ Analysis – selecting appropriate statistical measures (summation, averages, and spread) and associated formulae to inform analysis ■ Presentation – presenting insights from data by using charts and other visualisations
Year 9: Data science *Using data to investigate problems and make real-world changes.*	■ Collection – learners are exposed to global and local data sets ■ Analysis – how to use structured data to gain insights that can be used to solve problems in the real world ■ Visualisation – visualising data to help with the process of identifying patterns and trends

GCSE computer science and key stage 4

As learners progress into key stage 4, whether they opt to study computing as a GCSE or not, their focus on data becomes more specialised. Whilst still within the context of the data life cycle as a whole, the national curriculum and exam specifications largely emphasise the implementation phase.

All learners should continue to develop their skills at storing and organising data to help them solve problems and answer questions. The Teach Computing Curriculum units do this through a spreadsheet unit, which builds on past learning to include validation and conditional formatting techniques to enhance models.

Likewise, all learners continue to explore issues related to data protection and security across both computing and the wider curriculum.

Those learners studying GCSE computing will delve deeper and begin exploring relational database systems with particular focus on how they are structured. Whilst they may not create their own database at this stage, they will gain experience of Structure Query Language (SQL) to read and write data to a database.

Similarly, these learners will explore further ideas around data protection such as the responsibilities and rights of organisations and the individuals they store data about.

Table 4: Key skills and concepts at key stage 4.

Curriculum units	Key concepts and skills
Non-GCSE: Spreadsheets	■ Implementation – developing a data model for a specific scenario ■ Implementation – applying data validation and conditional formatting to make a model clear and robust ■ Analysis – creating charts and other visualisations ■ Presentation – communicating the output of a data model
GCSE: Databases and SQL	■ Implementation – describing the relationship between database tables, fields, records, and keys ■ Analysis – using SQL to interrogate and update an existing database

A level computer science

At this level, learners look even closer at how data is stored, organised, and analysed. Their knowledge of databases grows to include the design and representation of database structures, with some exploration of factors that affect a database's performance.

The use of SQL becomes a bigger focus and learners apply it to define, manipulate, and query more complex relational databases.

Learners consider even larger data sets, data that may be distributed across multiple systems, and the particular technical challenges that this presents. They will explore new approaches to structuring and querying "big data" and its suitability to be combined with functional programming approaches.

Table 5: Key skills and concepts within A level computer science.

Isaac Computer Science strand	Key concepts and skills
Database concepts	■ Implementation – distinguishing between flat-file and relational databases and describing the role of keys and indexes ■ Implementation – applying entity relationship modelling techniques to describe a scenario ■ Implementation – describing factors that affect the performance of a database system including normalisation, transaction processing, and distribution
SQL	■ Implementation – describing the categories of SQL command that are used to define, manipulate, and query databases ■ Implementation – identifying common definition (DDL) commands including CREATE, ALTER, and DROP ■ Implementation – identifying manipulation (DML) commands including INSERT, UPDATE, and DELETE ■ Analysis – identifying common query (DQL) commands including SELECT, GROUP BY, and COUNT

Table 5 (cont.): Key skills and concepts within A level computer science.

Big data	■ Implementation – distinguishing between structured and unstructured data ■ Analysis – describing analysis techniques including data mining and predictive analysis ■ Implementation – explaining the benefits of the fact-based model and use of functional programming

3.5. Progression across key stages

Having reviewed and analysed the role of data and information within the Teach Computing Curriculum and Isaac Computer Science for key stage 5, we are able to identify some broad themes of progression.

- The first theme is one of **specialism over time**, driven by the underlying curriculum and specifications. We can see that whilst learners begin studying the data handling cycle as a whole, their focus is narrowed and more specialised as they progress. This progression and specialisation may be self-evident given the more technical focus of the current GCSE and A level qualifications.

- Connected to this theme of specialisation is the recognition that the five data handling processes described above are not discrete stages, more a useful model to describe different activities within the data life cycle. This is particularly apparent in key stage 1 where distinguishing between each process in tasks undertaken by learners is extremely subtle. As learners progress, the **distinction between processes** becomes clearer.

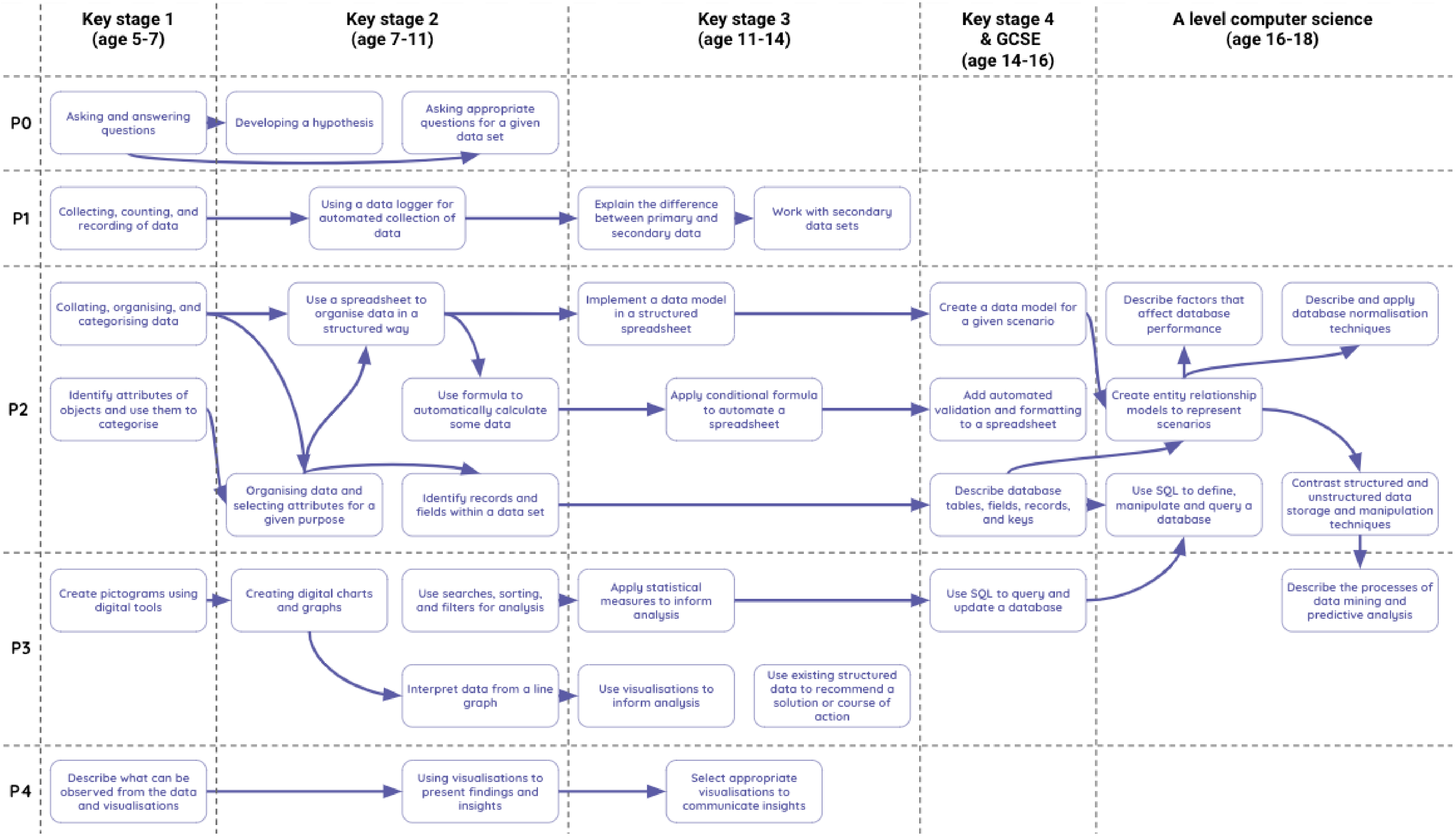
- Unsurprisingly, we can also observe a development in the **complexity of tools** that learners use to explore data. Their first conversations about data (in the Teach Computing Curriculum) focus on concrete objects, progressing to simply tallys and pictograms, before storing data in bespoke software tools, spreadsheets, and finally large scale database management systems.

- In parallel to a progression of tools, we see learners work with **data sets of increasing size and complexity**, starting with small-scale surveys and studies and ending with distributed "big data". Consequently, learners move from collecting their own data to having data provided for them or working with public data sets.

- The units and topics reviewed also suggest a change in the **role of humans and machines** in the data handling cycle. Early on, the learner performs all the substantive tasks from collecting through to analysing and presenting. As they progress, they encounter examples of where the machine can automate collection, storage and calculation, visualisation, and even analysis.

- In exploring questions relating to the impact of data as well as issues of safety and security, learners shift their focus **from individuals to organisations and wider society**. They learn about how data impacts them personally and how their data is useful before considering the responsibilities for organisations collecting data and the associated challenges, before thinking about the impact of data upon society as a whole.

Figure 6: Wider skills required by students of data and information.



4. Pedagogical strategies for teaching Data and Information

4.1. Pedagogy principles

The work of the National Centre for Computing Education is underpinned by 12 pedagogical principles that can be exemplified by a range of evidence, informed practices, and strategies. These principles apply across the teaching of computing; however, some are more applicable than others in each strand of the curriculum.

Data and Information is an area of computing education research where there has been less focus compared to areas such as programming or algorithms. However, our 12 principles draw upon general education theories as well as computing-specific research. Each of our principles can be explored in detail from a number of sources:

- The NCCE's collection of pedagogy *Quick Reads* provides short, digestible summaries on a range of topics[11]
- Hello World's *Big Book of Computing Pedagogy* provides similar summaries along with teacher stories and wider research insight[12]
- A comprehensive review of programming pedagogies can be found in a recent report from the Raspberry Pi Foundation[13]

Here, we present the most relevant of our 12 pedagogical principles for data and information as well as some illustrations of what they look like in practice.

Figure 7: NCCE's 12 pedagogy principles.

Lead with concepts

- The data life cycle model we employ incorporates five process areas, and each of these process areas incorporates a range of concepts and associated skills. These concepts are the starting point of all learning experiences and enable learners to understand, as well as practice, new data skills.

[11] National Centre for Computing Education. *Quick Reads*. Available at: https://blog.teachcomputing.org/tag/quickread/ [accessed 27 January 2022]

[12] Hello World. *Big Book of Computing Pedagogy*. Cambridge: Raspberry Pi Foundation. 2021.

[13] Waite, J, & Sentance, S. *Teaching programming in school: A review of approaches and strategies*. Raspberry Pi Foundation. 2021. https://www.raspberrypi.org/app/uploads/2021/11/Teaching-programming-in-schools-pedagogy-review-Raspberry-Pi-Foundation.pdf

- Data and information is deeply connected to other areas of computing and these connections between concepts should be explored. Concept maps, Frayer charts, and glossaries are useful approaches to recording shared understanding and associated vocabulary.

Model everything

As well as being a concept-rich part of the computing curriculum, data and information also involves many new skills that educators need to model.

- As each new tool is introduced, there are opportunities for teachers to model their use, highlighting familiar features and approaches.
- Part-filled templates or worked examples can provide learners with a scaffolded learning structure.
- Modelling good practices is essential to help learners avoid forming bad habits, which may lead to alternative conceptions and ultimately hinder future learning.
- Share your thought processes and any decisions you make whilst demonstrating, share the decision-making process as well as the steps taken.

Make concrete

Bring abstract concepts to life with real-world, contextual examples and a focus on interdependencies with other curriculum subjects.

- Choose contexts that surround your learners, particularly at home and school, but also those present in their everyday lives. Find examples where learners experience data, its collection, and analysis.
- Start with concrete manipulatives for younger learners, which will be familiar objects in the classroom, such as coloured 2D or 3D shapes. Use these objects to identify attributes, sort into groups, and find patterns.
- Where possible, adapt activities provided in the Teach Computing Curriculum to your learners' local community, experiences, and cultural background[14]. Use data to solve problems that matter to them.

Unplug, unpack, repack

To teach new concepts, first unpack complex terms and ideas, then explore these ideas in unplugged and familiar contexts, before you repack this new understanding into the original concept.

To help, teachers can apply a semantic wave[15] approach. In simple terms, this encourages educators to:

- Present learners with an abstract concept. For example, fields in a database.
- Unpack the meanings within the concept and relate it to a familiar concept. For example, discuss attributes that are familiar to learners such as height or age.
- Explore the concept within this familiar analogous context, perhaps using physical props to demonstrate. For example, using objects in the room, learners describe the attributes they recognise and then identify commonalities such as red, blue, and yellow being 'colour.'

[14] Raspberry Pi Foundation. *Quick Read: Culturally relevant pedagogy*. National Centre for Computing Education. 2021. https://blog.teachcomputing.org/quick-read-culturally-relevant-pedagogy/ [accessed 17 June 2022]

[15] Raspberry Pi Foundation. *Quick Read: Using semantic waves to improve explanations and learning activities in computing*. National Centre for Computing Education. 2020. https://blog.teachcomputing.org/quick-read-6-semantic-waves/ [accessed 17 June 2022]

- Repack the meanings of the original concept and highlight the similarities and differences between the analogy and the original computing context. In our example, the common attributes that learners have identified can now become the fields that they will use in their database.
- Finally, return the original concept in its own context. Here, learners may identify potential fields for objects that are not in front of them, for example, cars.

Challenge misconceptions

Regardless of how well a concept is taught, there is always space for alternate conceptions[16] (commonly known as misconceptions) to develop. In fact, sometimes we may knowingly introduce a misconception in order to simplify a concept or make it accessible. Recognising those misconceptions and knowing how to mitigate them is important, especially in an area of the curriculum that focuses on concepts.

- Teachers should make a conscious effort to seek out misconceptions and challenge them. Using regular formative assessment can help uncover misconceptions.
- Carefully written multiple choice questions can be used diagnostically[17] with distractors (wrong answers) that each result from a specific misconception.
- Concept mapping is another useful tool. If learners create their own maps, these should be a reflection of their internal understanding and can help identify the root of a misconception.
- Peer instruction[18] is a particularly effective technique that is based on a flipped learning approach. Learners complete a task before the lesson, in which they 'learn' new concepts. The lesson time is then used to answer diagnostic questions collaboratively and relies on peer discussion to build consensus around a concept. It not only helps identify misconceptions, but also helps address and correct them.

Create projects

Pupils need opportunities to apply the skills, knowledge, and understanding that they have developed, and project-based[19] activities can be a great way to facilitate this.

- Projects give pupils a goal, an audience, and a brief to fulfil, for which they need to make autonomous decisions about the skills, knowledge, and tools that they will use.
- Projects are a valuable context in which pupils can develop their questioning, collection, analysis, and presentation skills, as well as providing opportunities for collaboration.
- Repack the meanings of the original concept and highlight the similarities and differences between the analogy and the original computing context. In our example, the common attributes that learners have identified can now become the fields that they will use in their database.

---

[16] Raspberry Pi Foundation. *Quick Read: Addressing learners' alternate conceptions in computing*. National Centre for Computing Education. 2022. https://blog.teachcomputing.org/quick-read-addressing-learners-alternate-conceptions-in-computing/ [accessed 17 June 2022]

[17] Eedi. *Teach Computing NCCE*. https://eedi.com/projects/teach-computing [accessed 1 June 2022]

[18] Raspberry Pi Foundation. *Quick Read: Using peer instruction to discuss computing concepts*. National Centre for Computing Education. 2019. https://blog.teachcomputing.org/quick-read-4-peer-instruction/ [accessed 17 June 2022]

[19] Raspberry Pi Foundation. *Quick Read: Using project-based learning to apply programming knowledge to real-world scenarios*. National Centre for Computing Education. 2021. https://blog.teachcomputing.org/project-based-learning/ [accessed 17 June 2022]

5. Professional development for computing teachers

A core part of the NCCE's role is to help teachers develop their subject knowledge and pedagogy through continued professional development (CPD). There are a number of routes for teachers to participate in CPD to support their understanding of data and information.

Table 6 provides a sample of some of the courses that are available as part of the NCCE; these are designed to support teachers' development of their data and information subject knowledge. Many more courses, both online and face-to-face can be found on the Teach Computing website[20].

Teachers of A level computer science can find an additional range of bespoke courses[21] organised by Isaac Computer Science on a range of relevant topics. Beyond accessing formal courses, there are many opportunities for computing teachers to learn through networks, such as Computing at Schools (CAS). These local communities continue to meet regularly and share best practices and skills and are therefore a great source of inspiration and development for teachers. As well as local support and meetups, teachers can find many CPD focused events[22].

Table 6: Courses to support teachers' development of subject knowledge for the Data and Information strand.

Key Stage 1	Key Stage 2	Key Stage 3	Key Stage 4	Key Stage 5
Get Started Teaching Computing in Primary Schools		Teach Computing in Schools: Creating a Curriculum for Ages 11 to 16		Bespoke courses are available. Recent examples include: Databases and Programming a relational database using SQL
Teaching Data and Information to 5- to 11-year-olds		Introduction to Databases and SQL		

[20] Teach Computing. *Computing courses for teachers*. https://teachcomputing.org/courses [accessed 1 June 2022]

[21] Isaac Computer Science. *Events*. https://isaaccomputerscience.org/events [accessed 1 June 2022]

[22] Computing at School. *Upcoming events*. https://community.computingatschool.org.uk/events [accessed 1 June 2022]

6. Conclusion

This report has sought to explore a significant area of the computing curriculum: data and information plays an important role in everyone's lives and impacts each of us everyday in small and large ways. Existing research literature highlights the importance of all in society being equipped with essential data literacy skills to ensure they can make informed choices about their personal data and to interpret data presented to them with a critical eye.

Beyond general data literacy, there is widespread acknowledgement of the lack of skilled data professionals who will build the data architecture of the future or use data to solve complex problems and make decisions. Looking further ahead as artificial intelligence (AI) and machine learning become more advanced, and more commonplace, we will need society to understand the role of data in AI applications and the risks that come with using poor data.

Improving learners' data literacy skills is a truly cross-curricular challenge, but one in which computing has an important role. Learners will encounter data in almost every subject area, and particularly in maths and science. Given how crucial computers have become to data analysis, it is clear that understanding their use and applications is important.

To explore progression within the strand of Data and Information, this report outlines some current research into data literacy within computing and presents an adapted model used within the Teach Computing Curriculum. We have explored how this model relates to each of the ten units in the Teach Computing Curriculum as well as Isaac Computer Science content at key stage 5.

Some notable trends can be observed, including gradual specialisation as learners progress through the current curriculum from a broadly equal focus across the full data life cycle to a specialism at key stage 5 in the implementation of data storage and analysis systems.

We have also been able to summarise some high-level steps in the progression through Data and Information from key stage 1 to key stage 5.

We hope you find this report useful and we welcome feedback on it, via research@teachcomputing.org. In addition to this Data and Information report, we have previously published reports on *Programming and Algorithms*, *Computer Systems and Networking*, and *Digital Literacy within the Computing Curriculum*. This report is the final report in this series.

National Centre for Computing Education

The National Centre for Computing Education (NCCE) is funded by the Department for Education and marks a significant investment in improving the provision of computing education in England.

The NCCE is run by a consortium made up of STEM Learning, the Raspberry Pi Foundation, and BCS, The Chartered Institute for IT. Our vision is to achieve a world-leading computing education for every child in England.

The NCCE provides high-quality support for the teaching of computing in schools and colleges, from key stage 1 through to A level. Our extensive range of training, resources, and support covers elements of the curriculum at every key stage, catering for all levels of subject knowledge and experience.

For further information, visit: teachcomputing.org

This resource is available online at ncce.io/tcc.

Contributions: We would like to thank the many people who helped to create the Teach Computing Curriculum: our content writers, advisors, reviewers, pilot schools, and every teacher who has taken the time to send us feedback.