BCS, The Chartered Institute for IT - Consultation Response to:

DCMS National Data Strategy open call for evidence, July 2019

BCS
The Chartered Institute for IT
First Floor, Block D
North Star House
North Star Avenue
Swindon SN2 1FA

BCS, The Chartered Institute for IT

The purpose of BCS as defined by its Royal Charter is to promote and advance the education and practice of computing for the benefit of the public. We bring together industry, academics, practitioners and government to share knowledge, promote new thinking, inform the design of new curricula, shape public policy and inform the public. As the professional membership and accreditation body for IT, we serve over 60,000 members including practitioners, businesses, academics and students, in the UK and internationally. We also accredit the computing degree courses in ninety eight universities around the UK. As a leading IT qualification body, we offer a range of widely recognised professional and end-user qualifications.

This document is the BCS response to the DCMS open call for evidence[1] for a National Data Strategy.

National Data Strategy

BCS is one of the professional engineering institutes providing a collective response to the call for evidence through the new National Engineering Policy Centre, under the auspices of the Royal Academy of Engineering. This document provides separate evidence that is specific to BCS in its role as the Chartered Institute for IT and should be read in conjunction with the NEPC response.

The evidence included here is based on extensive consultations, workshops, roundtables and surveys we have conducted over 2018 and 2019. These include:

- In January 2018, BCS working with DCMS, surveyed[2] 6,500 schoolchildren through the BCS Computing At School network[3].
- A survey of BCS members in autumn 2018, which resulted in responses from over 4,000 members of all grades.
- Over the last six months BCS has conducted a series of roundtables in London, Edinburgh, Belfast and Cardiff with expert stakeholders from outside of the BCS membership, looking at data governance, lack of diversity and inclusivity in the IT profession, and the need to improve the level of digital skills in the general population.
- Working with the Office for AI, BCS has been involved in a series of consultation workshops over the last nine months that have included practitioners working for Amplify, ARM, BT, BAE Systems, Cambridge Consultants, Cisco, DeepMind, Deloitte, HSBC, IBM, Infosys, Lloyds, McKinsey Quantum Black, Microsoft, Nvidia, and Ocado.
- The BCS latest reports on diversity in the IT profession: gender[4], age[5], ethnicity[6].

[1] https://www.gov.uk/government/publications/national-data-strategy-open-call-for-evidence/national-data-strategy-open-call-for-evidence#questions
[2] https://www.bcs.org/upload/pdf/internet-safety-results.pdf
[3] https://www.computingatschool.org.uk/
[4] https://cdn.bcs.org/bcs-org-media/2946/insights-gender-2019.pdf
[5] https://www.bcs.org/upload/pdf/diversity-report-3-inclusive-it.pdf
[6] http://www.bcs.org/upload/pdf/diversity-report-2-ethnicity.pdf

People

Research area: opportunities and barriers to trust

1.1. How can organisations (private, public or third sector) demonstrate trustworthiness in their use of data?

The public can trust an organisation when they are confident that the organisation is demonstrably both competent and ethical. The evidence we include below shows that at present both the IT profession and the public at large are at best unsure of current ethical practices in organisations and in many cases believe organisations behave in ways that could be considered unethical. At the end of this section we include our expert opinion of some key prerequisites for improving trustworthiness in organisations.

Figure 1 shows the results from one of the questions from our survey of 6,500 schoolchildren in 2018. Roughly speaking about a third of schoolchildren across all ages do not believe companies think about the safety of children when developing websites and apps.

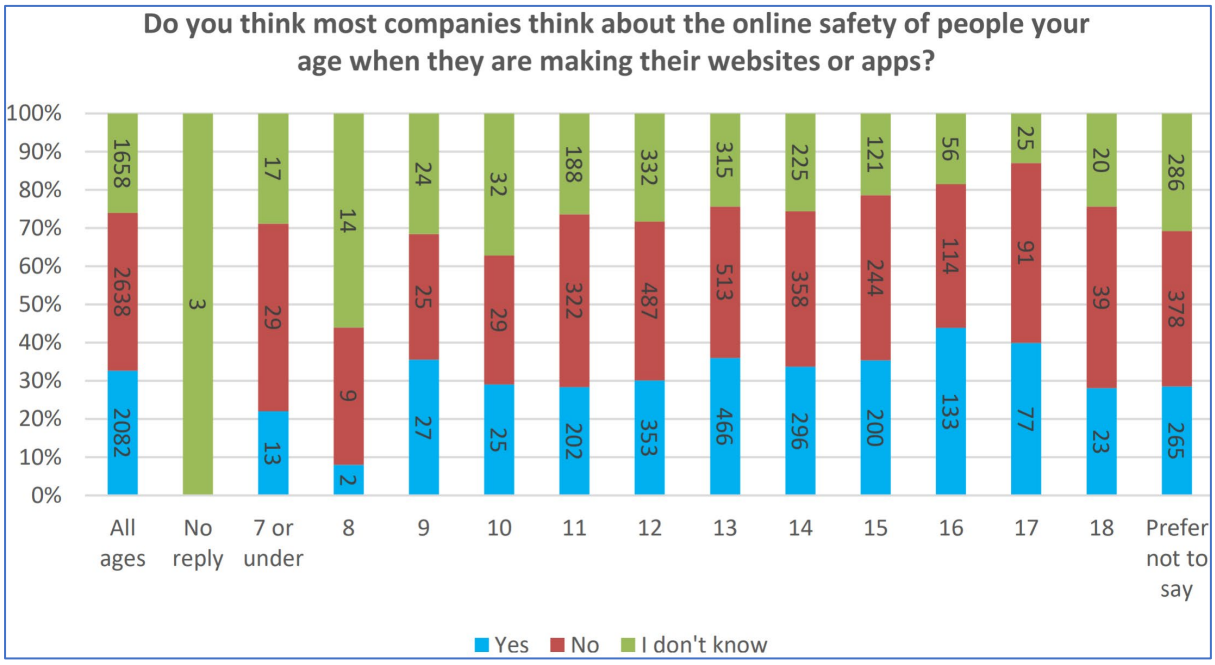

Figure 1

In late autumn of 2018 BCS surveyed its members on topics including ethical standards across the IT industry, receiving responses from over 4000 members, ranging from UK students in HE to Chartered members and senior Fellows. Figure 2 shows the result of one of the questions from the survey that asked members for their perception of ethical standards in the IT sector. The bottom line of the table shows the aggregation of the scores for the 'Neither high nor low', 'Quite low' and 'Very Low' categories. Collectively those three categories are termed 'not high'. This table shows that 47% of UK students chose 'not high'

categories for ethical standards, compared to over 64% of Chartered IT Professionals and Fellows who chose 'not high' categories.

Current State of Ethical Standard	Overall	Fellows	MBCS CITP	MBCS	AMBCS	Affiliate	Student UK	Student Non-UK
Neither high nor low	36%	35%	38%	38%	33%	33%	25%	33%
Quite high	28%	24%	26%	28%	34%	20%	29%	25%
Quite low	20%	28%	22%	20%	16%	20%	17%	15%
Don't know	8%	6%	7%	7%	10%	16%	12%	9%
Very high	4%	1%	3%	3%	5%	8%	12%	14%
Very low	4%	5%	4%	4%	1%	3%	5%	4%
Sum of not high categories	**60%**	**68%**	**64%**	**62%**	**50%**	**56%**	**47%**	**52%**

Figure 2

Overall across all categories of BCS members 60% chose a 'not high' category of ethical standard.

In contrast when we asked BCS members how important ethical practice was to them, 84% of them responded that ethical practice was very important, as shown in Figure 3. This suggests the underlying issues in organisational trust may have more to do with how well organisations institutionalise ethical practices, rather than any reluctance on the part of IT professionals to adopt ethical practices.

Importance of ethical practice	Overall	Fellows	MBCS CITP	MBCS	AMBCS	Affiliate	Student UK	Student Non-UK
Very important	84%	86%	86%	84%	83%	88%	80%	80%
Quite important	13%	9%	11%	13%	12%	10%	12%	19%
Not very important	2%	3%	2%	2%	3%	2%	3%	1%
No opinion	1%	1%	1%	1%	2%	0%	0%	0%
Not at all important	0%	1%	0%	0%	1%	0%	5%	0%

Figure 3

Based on this evidence BCS believes:

- Organisations need to demonstrably adhere to independent standards of professional and ethical practice giving IT professionals a mandate to uphold publicly visible ethical practices in their day to day work.
- The public and members of the IT profession should be able to appropriately influence professional and ethical standards as they evolve over time.
- Organisations and individuals must go beyond legislative compliance in developing innovative and sector leading practice and proactively collaborate with regulators in ways that are easily evidenced by the public.

One example of where this is beginning to happen is the NHS. FEDIP[7] is a professional register for NHS informaticians, founded by four professional associations and professional bodies including BCS. Its purpose is to improve modern health and care services through digital technologies demonstrated against publicly available competency standards, which are led by NHS professionals working with the public and the wider IT profession.

1.2. How easy is it for the public to find about how information provided to or inferred about them by an organisation is being used?

Our consultation and engagement indicates a largely reactive approach to personal data among the public who largely remain indifferent to what happens with their personal data until they experience harm as a result of data misuse. This reactive culture is a barrier to individual autonomy over personal data and leads to a focus on remedying harm once it occurs rather than proactively interrogating data governance processes and procedures to safeguard against future harm.

1.3. Are organisations (private, public or third sector) using personal data in ways that may damage trust?

The IT profession has significant concerns about the potential misuse of data, including:

- poor data governance that results in companies unable to effectively monitor how data is used, who is using the data, or where duplicates of data is stored, which may result in unethical practice going undetected
- lack of diversity in product development teams leading to hard-wired unconscious bias into new products or services
- using incomplete data to incorrectly infer personal characteristics,
- allowing data to be improperly shared within an organisation,
- improperly aggregating data from different sources to infer personal characteristics
- incorrectly cleaning data,
- incorrectly disambiguating data resulting in the wrong data being associated with an individual
- incorrectly merging different data pipelines from third parties
- not conducting proper due diligence to ensure correct provenance of data through the supply chain (which may well be offshored and distributed across different national jurisdictions)
- using data analysis methodologies that are not valid in a particular context
- applying analytical models as part of decision-making processes that are poorly tested (including for example inappropriate Machine Learning based neural networks)
- using invalid anonymisation techniques that do not provide sufficient protection against deanonymisation
- storing data insecurely so that it is at risk of being misappropriated by bad actors

[7] http://fedip.org/why-it-matters

Research area: concerns around trustworthiness

1.5. Do people know how information provided to, or inferred about them by, an organisation (private, public or third sector) is being used, stored and shared?

See our answer to 1.2 above.

1.6. To what extent are people concerned about how data about them is used, stored and shared? Are some groups more concerned than others? Are there particular categories of data that raise more concerns than others?

See our answer to 1.2 above.

1.7. What commercial practices or behaviours have affected trust in the use of personal data? Have targeted advertising and 'recommending' affected trust?

There is an inevitable tension between getting ethics right, achieving the best possible product performance, and minimising time to market. Minimising time to market may result in issues such as those listed in 1.3 above. Please see our joint response through the National Engineering Policy Centre for further details.

1.8. Have the General Data Protection Regulation (GDPR) and Data Protection Act 2018 made people more concerned about how personal data is managed? How has it influenced their behaviour?

Many employees have some level of access to personal customer data. Many of those people will have had GDPR compliance training provided by their employer, and may have a basic grasp of GDPR although this is not currently known and the Government may wish to commission further work in this area to test the efficacy of the UK's GDPR's implementation, public knowledge, competency and attitudes to risk. In our view that section of the population, having been made aware through their work of the risks around data, should be more aware about how they personally manage data.

At the same time since the introduction of GDPR, there have been more than 200,000 data breaches that raise questions for consumers about the competency and capability of companies holding personal information, which have been widely reported in the technical press[8]. That suggests those members of the public who read the technical press will be significantly more concerned with data use.

Significantly, a large section of the population do not deal with personal customer data through their work and are not familiar with the technical press, hence they are more likely to be indifferent to what happens to their personal data and be unaware of the possible risks they are exposed to.

[8] https://www.itproportal.com/features/one-year-on-how-gdpr-has-transformed-customer-experience/

1.9. How far do existing protections, such as in the Data Protection Act, go in promoting transparency and trust? What, if anything, should the government do to further build trust?

The most important aspect of the Data Protection Act in promoting trust is that it establishes in law that people have fundamental digital rights that companies must uphold. That provides the bedrock for ethical practice for companies, and it also makes clear that individuals can enforce their digital rights as and when they chose to. The very fact there are such fundamental universal digital rights is gradually having a transformative effect on industry, but this is still very much at an early stage and is complicated by the distributed, diverse nature of interconnected global supply chains and considerable variation in corporate values that underpins effective implementation of GDPR.

As stated in 1.1, our expert view is that trustworthy companies can demonstrate they are both competent and ethical. Our work with the Office for AI consulting widely with employers (see the top of the document for a list of companies involved in the consultations) highlighted that their most urgent business need is for **diverse interdisciplinary** teams who are able to **ethically**:

- transfer **scientific** knowledge into business contexts
- **engineer** systems that meet strategic business objectives
- **manage** the adoption of technology and maximise its value across strategic business units

In our view government should work with relevant stakeholders to ensure the UK's digital workforce, especially those in AI related roles, meet appropriate professional and ethical standards that meet the above stated needs of employers. Demonstrably meeting independent standards is an effective way of building trust with the public, and acts as a mechanism for continuously improving corporate behaviour as well as professional practice. Government should also consider how it can work with relevant stakeholders to substantially improve diversity and inclusivity in the digital workforce, which is vitally important to improving ethical practices in industry since diverse teams are better equipped to understand the societal consequences of technical decisions.

Objective 2. To ensure that everyone can effectively participate in an increasingly data-driven society

Research area: managing personal data

2.1. Are people aware of how to manage personal data about them? Do they know about tools to control access?

Please see our answer to 1.8.

Research area: inclusivity

2.2. How does effective participation and data use differ by location and demographic group across the UK? What does this look like in urban areas, rural areas and more remote parts of the country?

An important attribute of inclusivity in a data driven society is the extent to which the IT profession is diverse and inclusive. We listed web links to the 2018 BCS reports on gender, age and ethnic diversity at the start of this document. Some notable statistics from those reports are:

- There were 226,000 female IT specialists in the UK workforce during 2018 - 16% of the total at that time
- Female IT specialists were almost five times more likely to be working part-time than males (i.e. 14% versus 3%) – most often as they did not want full-time work.
- At £18 per hour, the median hourly earnings for female IT specialists in 2017 was 11% less than that recorded for males working in IT positions
- At 19%, BAME representation was higher amongst IT specialists than within the workforce as a whole (12%) in 2018 and in total there were 266,000 BAME IT specialists in the UK at that time.
- BAME representation amongst IT specialists varies significantly across the UK - from just 6% in the South West of England to 35% in London over the 2014-18 period.
- In 2018, BAME IT specialists (full-time employees) were earning 11% more than IT specialists as a whole, with median hourly rates of £22 and £20 per hour respectively.
- BAME IT specialists are less likely to be in 'positions of responsibility' than others in IT roles with 36% and 43% respectively stating that they were a manger/foreman or team leader in 2018.
- Almost nine in ten BAME IT specialists have an HE level qualification (86%) and 17% are IT graduates (with related under/post-graduate awards).

These selected highlights from the BCS reports suggest the largest diversity issue in the IT profession is gender diversity, which should be a priority area for concern.

2.3. How does effective participation and data use differ by sector?

See 2.4 for our response.

2.4. What barriers to participation do different groups face? How are marginalised and vulnerable groups affected?

ONS data[9] shows that 22% of disabled adults had never used the internet in 2017, compared to 9% in the general population. It seems reasonable to suppose this is in large part due to unacceptably poor accessibility of internet services, rather than a lack of desire by disabled adults to use the internet. This is a major concern as interaction with healthcare and social services will gradually move to being offered only online in the long-term. Lack of accessibility in websites is a major barrier for disabled adults being able to have full

[9] https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2017

participation in a digital society. The new government regulations[10] for public sector websites and apps are a welcome and positive move towards improving accessibility. Close monitoring of whether these result in a step change in overall website quality is recommended as well as monitoring technical compliance. In future, developers should consider the needs of our diverse communities and how best to improve the overall quality of the user experience, and not focus solely on minimal regulatory compliance.

Individuals less likely to declare culturally sensitive personal data to Government, companies or the public sector, may face significant barriers to participation. This may be the case for individuals from ethnic minority groups, individuals of faith, LGBT people, those with disabilities and this may largely be due to a cultural lack of trust of authority, corporations or data governance structures which exacerbates as their trust and confidence erodes through negative experiential evidence, hearsay and media portrayals. Internet scraping across different social media platforms can aggregate snippets of information that may collectively infer intimate details about someone and even where they live, leading to the potential for online harassment, Intimidation or discrimination. Peers or colleagues may inadvertently post information that can facilitate the connection of other data leading to breach of privacy through data aggregation. Given the different ways organisations could act incompetently or unethically with data (see 1.3) individuals attempting to keep culturally sensitive personal information offline may reasonably assume no aspect of their private life should be shared online, resulting in self-exclusion from online engagement, widening the digital participation gap for certain minorities and at-risk groups.

Recently the Guardian reported[11] a scam in which criminals obtain claimants' personal details, often by posing as DWP staff, to make bogus online applications for universal credit. This illustrates how some vulnerable individuals through lack of understanding of complex social security regulations, the risk structures in place in financial institutions and lack of access to trustworthy independent support can be criminally exploited.

2.5. What can government do to support those in marginalised and vulnerable groups? What elements of our digital society would most benefit these groups?

UK, Devolved and Local government are increasingly providing access to vital public services through online means. For the benefits of digital government to be fully realised, providers of public services must work to build public trust through both the delivery of a competent and ethical approach to data governance and protection and an iterative awareness campaign tailored to the needs of our diverse communities.

The first step in building such trust is for government to engage with our diverse communities and relevant stakeholders, including regulators and charities, to understand their concerns and develop an online user experience that genuinely delivers the best possible service for all. DCMS does publish public service standards[12] that if wholeheartedly

[10] https://gds.blog.gov.uk/2019/05/21/accessibility-update-sample-accessibility-statement-monitoring-and-enforcement/
[11] https://www.theguardian.com/society/2019/jul/09/universal-credit-scam-leaves-claimants-pounds-hundreds-out-of-pocket
[12] https://www.gov.uk/service-manual/service-standard

applied across all local and national government would also go a long way to resolving these concerns.

Increasingly AI technology is enabling cheap access to commercial personal medical diagnostic services that were previously unaffordable to many and until now only accessible to most through the public sector. The risk with such new unregulated services is that rushed and poorly developed healthcare products could make false claims undermining public trust in such technologies. Alarmingly, a recent article in the Lancet[13] raised significant questions about the medical competency of some healthcare apps and their potentially bogus marketing claims. The public should be able to trust medical expertise embedded in such apps. The rationale for this trust should be easily accessible and rooted in both rigorous research practice and properly safeguarded personal medical data. Government should review the regulatory framework around the ethical standards of healthcare technology. Although we have highlighted commercial healthcare apps as an area for concern, there are similar concerns around other commercial apps providing public services and information.

If private sector companies providing basic public services had to adhere to the DCMS public service standards mentioned above and comply with the DCMS data ethics framework[14] this would greatly improve matters.

Research area: data skills and employment

2.6. How important are basic data skills for employment in today's economy? What is the basic level of data skills needed and what kinds of skills are needed?

The following list describes a range of online basic skills relating specifically to data that apply to all adults irrespective of their employment. People with basic data skills should feel confident they can make the following statements:

- I have the skills necessary to find out my legal digital rights when I am shopping or accessing services online, and I know how to seek help if those rights are infringed.
- I have the skills necessary to find a range of approaches to deal with online harassment and trolling and where to get help if this happens
- I have the skills necessary to talk to my children about staying safe online and how best to support them safely engage in online activities.
- I have the skills necessary to investigate how different snippets of personal data can be aggregated from apps running on different devices that may reveal to a third party my sexual preferences, personal details about my mental health, my political views, where I live, and other highly personal data I am likely not to want to share with strangers.
- I have the skills necessary to access web browser controls to turn on or off those functions that I wish to use or not use, e.g. that turn on or off cookies and JavaScript,

[13] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)32819-8/fulltext
[14] DCMS Data_Ethics_Framework.pdf

that allow me to delete any personal information that may be present, to control access to credit card details, or how to disable audio and video on my computer to stay safe.
- I have the skills necessary to investigate whether web-browsers and software are flawed and contain errors that mean I can never assume they are 100% safe.
- I have the skills necessary to investigate whether third party apps on my phone should be treated as insecure and not safe if they contain advertising from arbitrary sources, I understand how to check what data on my phone apps want to access, and that using them can lead to a breach of my privacy.
- I understand that passwords alone are not sufficiently safe for highly sensitive accounts, such as my bank account, and I understand how to use two-step authentication to make them safer.

There are a great many other basic data skills, which are not listed here. For example, according to the DfE statutory school computing curriculum[15], all children by the age of 12 years old (i.e. at the end of Key Stage 2) should be able to:

- 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.

If all 12 year olds are now expected to have these skills it seems reasonable to assert that all adults should also have them, and since they are part of the primary school curriculum it is safe to assume they are basic skills.

2.7. In which professions are data skills most important?

BCS awards Chartered IT Professional status based on levels of competency evidenced against the employer led Skills Framework for the Information Age (SFIA[16]). Virtually all of the roles in the SFIA framework require proficiency with data analysis related skills in some shape or form. SFIA skills range over the following categories of job roles (see the SFIA website for specific skills listed against each of these categories):

- Information strategy
- Advice and guidance
- Business strategy and planning
- Technical strategy and planning
- Business change implementation
- Business change management
- Systems development
- User experience
- Installation and integration

[15] https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study/national-curriculum-in-england-computing-programmes-of-study
[16] https://www.sfia-online.org/en/framework/sfia-7/skills-home

- Service design
- Service transition
- Service operation
- Skill management
- People management
- Quality and conformance
- Stakeholder management
- Sales and marketing

We argue that data skills in some form are vital for a professional job that falls into any of the above categories. This view is consistent with Nesta[17] research into digital skills that are likely to be needed in the future, published in 2018.

People in the above types of professional roles can be found in the NHS, the financial sector, the armed forces, the civil service, as well as the retail sector, energy sector, construction sector and of course the IT sector. It is also the case that many employers are looking for people in roles related to those above who in addition to the existing SFIA skills related to those roles also understand how to apply machine learning techniques and methodologies within those roles to create business value across strategic business units. At the present time such people are very rare.

2.8. Are the relevant skills available and supported where they're needed?

The answer is a partial yes in some contexts, but with caveats. For example, the following table shows the top ten degree apprenticeships in 2017/18, published by the Office for Students[18]. This shows that the Digital and Technology Solutions Professional apprenticeship (a degree level apprenticeship) was the second highest taken degree level apprenticeship for that year.

Framework or standard	Level	Numbers
Chartered manager	6 (degree)	2315
Digital and technology solutions professional	6 (degree)	1310
Chartered surveyor	6 (degree)	815
Registered nurse	6 (degree)	305
Civil engineer	6 (degree)	160
Healthcare science practitioner	6 (degree)	110
Manufacturing engineer	6 (degree)	105
Product design and development engineer	6 (degree)	100
Embedded electronic systems design and development engineer	6 (degree)	95
Aerospace engineer	6 (degree)	85

Figure 4

[17] https://www.nesta.org.uk/report/which-digital-skills-do-you-really-need/
[18] https://www.officeforstudents.org.uk/publications/degree-apprenticeships-a-viable-alternative/

DfE data for new apprenticeship registrations[19] for May 2019 shows there were 68,644 for all ICT practitioner apprenticeships started across all levels, of which 12,558 were at level 4 or above, and of which 2079 were Data Analyst apprenticeships.

Although this number is significant and growing substantially year on year, many employers have been reporting shortages of people with the necessary digital skills every year in CBI and other regular business surveys. Of note, we have found in our consultations across industry as part of our work to support the Office for AI that employers are experiencing a severe shortage of people with AI and Machine Learning related skills.

Government has recognised there is employer demand for at least 3000 AI MSc graduates every year based on its AI review[20]. The BCS independent report (summarised here[21]) commissioned by the Office for AI found that:

- There is extensive capacity throughout the UK to provide high quality Machine Learning and AI MSc courses.
- There is a growing sense of a shared Machine Learning body of knowledge across universities that is highly relevant to the needs of employers.
- Risk: There aren't enough MSc programmes that develop professional skills through extensive work-related experience, but there are MSc courses which are exemplars of how this can be achieved.

2.9. Is industry able to provide the relevant skills or is further skilling needed through the education system?

See our answer to 2.8 above. In our view developing appropriate skills needs to be viewed holistically, as a collaboration of education providers and employers through both apprenticeships, including degree apprenticeships, as well as ensuring that university undergraduates on traditional academic degrees are given extensive opportunities to develop professional skills through work-related experience evidenced against appropriate professional standards. Industry on its own is therefore not able to develop all the relevant skills that will meet the nation's needs, but industry is essential to developing and supporting different academic and technical education pathways that lead to a diverse, substantial pipeline of ethical and competent professionals.

Economy

Objective 3. To ensure that all businesses and non-profit organisations can effectively operate in an increasingly data-driven economy

Research area: technological developments

[19] https://www.gov.uk/government/statistics/apprenticeship-and-levy-statistics-may-2019
[20] https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk
[21] https://cdn.bcs.org/bcs-org-media/3047/ethical-ai.pdf

3.5. Are small and medium businesses and non-profit organisations sufficiently benefiting from new emerging technologies such as AI and the Internet of Things (IoT)?

From our consultations with employers it was clear that at present there is a severe shortage of skilled professionals who can mange the deployment of AI enabled technologies across strategic business units. This is true for all companies, SMEs and large corporates. Both SMEs and large corporations reported in our consultations that they believe diverse interdisciplinary teams are both more effective and tend to have higher standards of ethical data practices than non-diverse teams. The difficulty SMEs have in recruiting diverse teams therefore means they are less able to benefit from AI and IoT.

A related concern for SMEs is data security. SMEs accounted for 61% of data breaches in 2018[22]. The cost of providing adequate data security may act as a barrier to adopting new AI and IoT technologies for SMEs. Compliance with regulations, such as for example GDPR, also represents a larger overhead for SMEs than international corporations. If SMEs choose to adopt cloud based services that manage their data in a secure environment as well as provide AI enhanced data analysis services this could mitigate security risks and reduce costs, but this could also result in an AI technology oligarchy that entrenches the market dominance of a few global corporations and holds back future innovation. This possibility would appear to be an issue where carefully thought out and well-informed government policy will be critically important to ensure markets are competitive and fair.

3.6. How do businesses envisage that future technological developments will change how they use data?

Data is only useful when it can be generated, gathered, structured, curated, processed and analysed to create actionable information relevant to a company's business model. Being able to automate all the above functions almost instantaneously makes innovative new data engineering and analysis services possible, due to the ability to deal efficiently with incredibly large amounts of data that is generated concurrently in many different places.

Data will gradually evolve to be much more highly structured and meaningful as technology makes it possible to automatically generate and gather much richer data, which is standardised across multiple data pipelines connected to heterogeneous IT systems. As data becomes more structured and accessible it becomes more valuable, and that means exchanging data between companies can lead to potentially attractive business opportunities offering innovative data engineering or data analysis services. For example, population scale personal data is already effectively a commodity for global social media platforms, which has resulted in previously unimaginable digital marketing services offered by third party companies that can access this data.

Objective 4. To improve growth and productivity through the effective use of data across the economy

[22] https://www.businessnewsdaily.com/8231-small-business-cybersecurity-guide.html

Research area: productivity

4.1. How is the effective use of data driving business productivity through increased efficiency?

The availability of data can make it possible to automate and outsource some business functions, hence improving efficiency. Similarly having access to timely data makes it possible to control supply chains more efficiently and have better 'just in time' capability. Comprehensive market data, which may be the result of real time data generated from internet connected devices, allows companies to be more efficient in adapting to or possibly even predicating changes in their markets. In other words, quantitative methods that are fed by real time meaningful data can be used to automate management decisions and business processes that was not previously possible, and what's more to make them more accurate and dependable.

4.2. What are the barriers to the potential productivity gains from more effective data use?

Companies need the infrastructure and capability necessary to automate generating, gathering, structing, curating, processing and analysing data. The inability to automate any of these functions is a significant barrier as the whole is far greater than the sum of the parts.

4.3. Are there best practices in particular sectors that others can learn from?

Exemplars of applications that provide both good and not so good practice are FinTech, social media companies, and healthcare.

4.4. How do firms develop expertise in their use of data?

To develop expertise, companies need access to networks of companies, universities, professional bodies, trade associations, public organisations such as the Alan Turing Institute, the Digital Catapult, and InnovateUK for example, who can facilitate the sharing of good practice and provide advice and guidance on adoption of new technologies. They need low cost, low risk, opportunities to experiment with innovative data engineering and data analysis within their business models. They need effective CPD for existing staff as well as a ready supply of talented people to recruit. They need access to funding that will allow them to invest in new technologies, and they need access to new markets to grow their businesses. Finally, they need sound regulation of data use to know how to operate legally and give confidence they will be competing in fair markets, and that regulation whilst providing comprehensive rights and obligations will as much as possible not hinder innovation.

Research area: societal and environmental benefits of better data use

We do not have anything further to add beyond our joint response through the National Engineering Policy Centre, please see that response for further details.

Government

Objective 5. To improve public services and government operations through the effective collection, sharing and use of data

Research area: current use of data

5.1. How effectively are government and the wider public sector collecting, sharing, analysing and storing the data it holds? What does good practice look like? What does bad practice look like?

In our view the GDS blog[23] last year gives a comprehensive answer to this question. Where data is being created through new IT systems connected to new data pipelines then it is possible to engineer them to be effective at collecting, sharing, analysing and storing data. When government departments have to rely on opaque legacy systems that were not intended to share data then it becomes incredibly hard to reverse engineer functionality onto a monolithic, undocumented, inflexible technology.

DfE provides a nice example of data sharing, where each year detailed GCSE results for all subjects and all schools are made available as properly structured data sets (which in practice people download as spreadsheets). Unfortunately, the attributes within the tables are not always consistent from one year to the next, which makes year on year comparisons only possible through extensive manual manipulation. This example illustrates the significant benefit of providing open access to structured machine readable data, as well as the frustration caused by departments not appreciating the need for consistency over time in order for external organisations to maximise benefit from access to data.

The NHS provide examples of good practice as well as less good practice. NHS England have a website showcasing some examples of good data sharing practice[24]. Ideally it should be the case that NHS Trusts can exchange structured data that enables one Trust to benchmark itself against other local trusts for a wide range of healthcare services. For example, a report last year by the Healthcare Financial Managers Association (HFMA[25]) explained how local NHS trusts can adopt good practice to improve confidence and accuracy in internal costing data, with the intention to help reduce overspends in acute services. Not all NHS Trusts have yet followed such good practice, which means as a whole the NHS lacks the capability to systematically use data to reduce costs and provide better value for money. In this instance the good intentions of the government to devolve healthcare and introduce market forces has inadvertently allowed poor data practices to evolve through the lack of effective data governance at national level.

5.2. What are the main barriers to more effective data use within government? Are there barriers in cases where government works with the private sector?

[23] https://technology.blog.gov.uk/2018/05/24/understanding-legacy-technology-in-government/
[24] https://www.england.nhs.uk/ourwork/tsd/data-info/open-data/examples/
[25] https://www.hfma.org.uk/publications/details/using-costing-data-to-improve-efficiency-in-an-acute-hospital

See our answer to 5.1. A potentially significant risk for government working with the private sector is vendor lock-in. Once a particular vendor's technology services are embedded at the heart of an organisation it can be extremely challenging to migrate to another vendor's services. For example, Amazon, arguably one of the most technically advanced, competent and well resourced organisations on the planet, needed over a year[26] to migrate away from Oracle's propriety systems. It is worth noting that Amazon is likely to have far more capacity to take on such a task than the UK government might.

That example suggests engineering practices, systems architectures, technology stacks, user interfaces, and data structures etc. that underpin commercial services to government must be based on industry recognised standards that are openly available to all companies and that partnerships with the private sector must be designed from the outset in ways that allow government to gracefully migrate data and systems to other providers whenever necessary.

On May 21st 2018 large parts of the national rail network suffered widespread delays and cancellations as railway companies tried to implement new timetables[27]. At the time it was suggested government lacked sufficient visibility of those companies plans for implementing new timetables and had been unable to validate whether those plans were credible. This highlights another risk in that if too much operational data oversight is outsourced to private companies then government may effectively outsource its governance capability altogether, leaving it unable to validate whether services are fit for purpose.

A secondary risk for government is the danger of undermining public trust if there could be a perception of private sector capture of government functions. As a thought experiment it is interesting to speculate how the public might view, for example, a company such as say Ubber winning a hypothetical contract to manage the ambulance services across the UK. What would it take for the public to trust Ubber with their medical details, which would likely be necessary to share with Ubber in order to prioritise when and where to send an ambulance? What about NHS Hospital Trusts, how would they view such a move? How would it be possible to verify that when the hypothetical contract ended all public data that may have been stored on Ubber systems was destroyed? If Ubber offered to run the service for free provided they could monetize data, would that have consequences for the public's trust in government? What if they wanted exclusive access to ambulance data for a limited time in exchange for providing their service for free?

This hypothetical example is not that farfetched. On July 10th the government announced[28] Amazon's Alexa will be partnering with the NHS to allow people to ask Alexa to find health related advice from NHS sources. This example is seen by some as controversial although the intention is entirely to ensure people have access to trusted advice from an authoritative source. In part it is seen as controversial because although Amazon are a highly reputable and an extremely competent company, the public will have no way to verify whether or not Amazon do act ethically. Members of the public might have doubts

[26] https://www.theregister.co.uk/2019/04/02/amazon_fulfilment_oracle_database/
[27] https://www.theguardian.com/uk-news/2018/may/21/cancellations-and-delays-as-new-rail-timetables-introduced
[28] https://www.bbc.co.uk/news/health-48925345

about how much they should rely on promises of confidentiality considering the recent discovery[29] that some Google workers in different parts of the world listen to around 0.2% of all voice recordings from its smart speakers. Presumably Google are just as reputable and competent as Amazon? It has been reported that a Google worker in the Netherlands was found to have leaked some of those recordings.

These examples raise the question of whether the public may want to have strong evidence that confidentiality and data safeguarding are hard-wired into technologies they use to access essential public services from the private sector. Ensuring such evidence is possible and readily available could be a barrier to government working with the private sector in some instances.

If private sector companies providing basic public services had to provide robust, publicly available evidence they adhere to the DCMS public service standards and comply with the DCMS data ethics framework mentioned above this would greatly improve matters (see our answer to 2.5).

[29] https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-home-recordings-listen-privacy-amazon-alexa-hack-a9002096.html