The Digitalisation of Science, Technology and Innovation

KEY DEVELOPMENTS AND POLICIES

The Digitalisation of Science, Technology and Innovation

KEY DEVELOPMENTS AND POLICIES

BETTER POLICIES FOR BETTER LIVES

Please cite this publication as:
OECD (2020), *The Digitalisation of Science, Technology and Innovation: Key Developments and Policies*, OECD Publishing, Paris, *https://doi.org/10.1787/b9e4a2c0-en*.

ISBN 978-92-64-97802-7 (print)
ISBN 978-92-64-50177-5 (pdf)

Photo credits: Cover : © Adobe Stock.

Corrigenda to publications may be found on line at: *www.oecd.org/about/publishing/corrigenda.htm*.

Foreword

The OECD's Committee for Scientific and Technological Policy (CSTP) brings together representatives from OECD countries, and a number of partner economies, to examine major aspects of public policy relevant to science, technology and innovation (STI). By guiding the OECD's empirical research and data gathering, and promoting peer-based learning, the Committee works to improve understanding of these policies and, ultimately, to advance policymaking itself.

The digital revolution and its implications have been central to the OECD's, and CSTP's, work for many years. Recently – during 2017 and 2018 – the OECD's Going Digital project comprehensively examined digital technology's economic and social impacts. The resulting report, *Going Digital: Shaping Policies, Improving Lives*, provides a roadmap for policy making in the digital age.

In 2015, in their joint declaration, ministers from OECD countries and partner economies, at the OECD Ministerial Meeting in Daejeon (Korea), recognised that digital technologies are revolutionising STI. Ministers highlighted that the rapid development of digital technologies is changing the way scientists work, collaborate and publish; increasing the importance of access to scientific data and publications; opening new ways for the public to engage and participate in science and innovation; facilitating research co-operation between businesses and the public sector; contributing to the transformation of how innovation occurs; and, driving the next production revolution. The ministers asked the OECD to monitor this ongoing transformation.

This publication examines digitalisation's effects on STI and the associated consequences for policy. It draws mainly on work performed under the aegis of CSTP during 2017 and 2018. Some of the topics addressed are longstanding themes in CSTP's work – from access to publicly funded research data, to the measurement of digital science and innovation. Other topics are newer and emerging, from the role of artificial intelligence in production, to how digital technology could help utilise the collective intelligence of the scientific community, to recent advances in the digitalisation of biotechnology.

Certain aspects of the digital revolution are still relatively new, even if their effects are already profound. It is evident that, owing to the general-purpose character of digital technology, its future development will also have far-reaching consequences. As digital technology and its many ramifications evolve, CSTP will continue to serve as a unique international and inter-governmental focal point for policy analysis and guidance in the field of STI.

This book was declassified by CSTP on 12 August 2019 by written procedure and prepared for publication by the OECD Secretariat.

Acknowledgements

Alistair Nolan edited this publication. Alistair Nolan also wrote Chapter 1 (“An overview of key developments and policies”) as well as Chapter 5 (“Artificial intelligence, digital technology and advanced production”). Mr. Nolan, and all the authors of this publication, work in the OECD’s Directorate for Science, Technology and Innovation.

Fernando Galindo-Rueda wrote Chapter 2 (“How are science, technology and innovation going digital? The statistical evidence”).

Carthage Smith wrote Chapter 3 (“Digital technology, the changing practice of science and implications for policy”).

Chapter 4 (“Digital innovation: Cross-sectoral examples and policy implications”) was written by Dominique Guellec, Caroline Paunov and Sandra Planes-Satorra.

James Philp wrote Chapter 6 (“Digitalisation in the bioeconomy: Convergence for the bio-based industries”).

Chapter 7 (“The digitalisation of science and innovation policy”), was written by Michael Keenan, Dmitry Plekhanov, Fernando Galindo-Rueda and Daniel Ker.

Dominique Guellec, Head of the OECD’s Division for Science and Technology Policy, provided detailed comments throughout. Valuable comments were provided by Andy Wyckoff and Dirk Pilat, Director and Deputy-Director, respectively, of the OECD’s Directorate for Science, Technology and Innovation.

Mark Foss provided editorial support. Serve Blandine gave statistical assistance. Florence Hourtouat and Beatrice Jeffries formatted and checked the text of successive drafts. And Angela Gosmann oversaw all aspects of the publication process. Their contributions are gratefully acknowledged.

Thanks are likewise due to the delegates to the OECD’s Committee for Scientific and Technological Policy who provided comments on chapter drafts.

Thanks are similarly due Professor Morten Irgens, of Oslo Metropolitan University (OsloMet), who arranged sponsorship for, and gave intellectual support to, a joint workshop between OsloMet and the OECD in Oslo on 1 and 2 November 2018. The workshop, titled “Digital Technology for Science and Innovation – Emerging Topics for Policy and Assessment”, explored emerging issues at the intersection between digital technologies and policies for science and innovation. Many of the ideas examined at that event appear in this publication.

Table of contents

Tables

Figures

Acronyms, abbreviations and units of measure

3D	Three-dimensional
ACE	Angiotensin converting enzyme
AI	Artificial intelligence
API	Application programming interface
ATBP	Advanced technology and business practice
BDA	Bio-design automation
BRICS	Biological Robustness in Complex Settings
CAD	Computer-aided design
CAMD	Computer-aided molecular design
CASRAI	Consortia Advancing Standards in Research Administration Information
CERIF	Common European Research Information Format
CO_2	Carbon dioxide
CPU	Central processing unit
CRIS	Current Research Information System
CSTP	Committee for Scientific and Technological Policy (OECD)
DARPA	Defense Advanced Research Projects Agency (United States)
DLT	Distributed ledger technology
DNA	Deoxyribonucleic acid
DOI	Digital object identifier
DPIA	Data Protection Impact Assessment
DSA	Data-sharing agreement
DSIP	Digital science and innovation policy
EBP	Earth BioGenome Project
EGF	Edinburgh Genome Foundry
ETIS	Estonian Research Information System

EU	European Union
EUR	Euro
FAIR	Findability, accessibility, interoperability and reusability
FRIS	Flanders Research Information Space (Belgium)
GBP	British pound
GDPR	General Data Protection Regulation (European Union)
GPU	Graphic-processing unit
HEI	Higher education institution
HPC	High-performance computing
HT	High throughput
ICT	Information and communication technology
IoT	Internet of Things
IP	Intellectual property
IPR	Intellectual property right
IRB	Institutional review board
ISSA	International Survey of Scientific Authors (OECD)
IT	Information technology
IVOA	International Virtual Observatory Alliance
MAM	Metals-based additive manufacturing
MB	Megabyte
ML	Machine learning
MOOC	Massive open online course
MTA	Materials transfer agreement
NA	National account
NARCIS	National Academic Research and Collaborations Information System (Netherlands)
NESTI	National Experts on Science and Technology Indicators (Working Party) (OECD)
NHS	National Health Service (United Kingdom)
NIH	National Institutes of Health (United States)
NLM	National Library of Medicine (United States)
nm	Nanometre
NNMI	National Network for Manufacturing Innovation (United States)
NSF	National Science Foundation (United States)
NZRIS	New Zealand Research Information System
OA	Open access
OECD	Organisation for Economic Co-operation and Development

OEM	Original equipment manufacturer
OMTA	Open Materials Transfer Agreement
ORCID	Open Researcher and Contributor ID
PB	Petabyte
R&D	Research and Development
RD&I	Research, Development and Innovation
RI	Research infrastructure
ROI	Return on investment
SMEs	Small and medium-sized enterprises
SOFA	Self-Organized Funding Allocation
SPIAS	SciREX Policymaking Intelligent Assistance System
STC	Statistics Canada
STEM	Science, technology, engineering and mathematics
STI	Science, technology and innovation
UK	United Kingdom
UPPI	Unique, persistent and pervasive identifier
URI	Uniform resource identifier
US	United States
USD	United States dollar
VR	Virtual reality

Executive Summary

This report examines digitalisation's effects on science, technology and innovation and the associated consequences for policy. Digitalisation today is the most significant vector of innovation in firms, science and governments. If properly harnessed, digital technologies could advance science, raise living standards, help protect the natural environment and improve policymaking itself.

Digitalisation and science

Digitalisation is bringing change to all parts of science, from agenda setting, to experimentation, knowledge sharing and public engagement. To achieve the promise of open science research budgets need to account for the increasing costs of managing data. Greater policy coherence and trust between research data communities are needed to increase sharing of public research data across borders. Co-operation is required to build and provide access to cyber-infrastructure internationally. And open access (OA) publication requires incentives for OA that match mandates coming from research funders.

Governments should also support platform technologies for science, such as distributed research and development networks, and storage for digital/genetic data. Room exists to better exploit advanced digital technologies in science. Artificial intelligence (AI) can increase productivity in science, at a time when research productivity may be falling. But policies are needed on high-performance computing, skills, and access to data (such as standardisation for machine readability of scientific datasets). AI in science also raises novel policy issues: for instance, will intellectual property systems need adjustment as invention by machines expands?

Realising the untapped potential of digital technology in policy

Digital technology could support policymaking for science and innovation in novel ways. Few governments have experimented with the opportunities available. Examples include: self-organised funding allocation; using collective intelligence through digitally enabled prediction markets and machine-crowd combinations; developing blockchain applications in science; and, using social media to help spread innovation.

Digitalisation and innovation in firms

As businesses innovate with data, new policy issues are likely to arise. For instance, restricting cross-border data flows can raise firms' costs of doing business, especially for small and medium-sized enterprises (SMEs). Decisions may soon be required on as yet unanswered policy questions: for example, should data transmitted in value chains be protected from sale to third parties?

AI is finding applications in most industrial activities. But firms with large volumes of data may not have the in-house skills to analyse it fully. Governments can work with stakeholders to develop voluntary model agreements and programmes for trusted data sharing. For more general AI applications, governments can also promote open data initiatives and data trusts and ensure that public data exist in machine-readable formats.

Effective sectoral support is also needed, for instance through roadmaps or sectoral plans, prepared with industry and social partners. Policy should also facilitate collaboration for innovation, for instance, by digitally enabled crowdsourcing and open challenges.

Even in the most advanced economies, the diffusion of advanced digital technologies needs to accelerate. Institutions for technology diffusion – such as applied technology centres – can be effective, and should be empowered to take longer-term perspectives, rather than prioritising short-term revenue generation. To help diffuse digital technology to SMEs governments can: systematise key information for SMEs; develop information on the expected return on investments in new technologies, and on complementary process changes; provide signposts to reliable sources of SME-specific expertise, along with facilities where SMEs can test varieties and novel combinations of equipment.

Developing digital skills

Occupational titles like "industrial data scientist" and "bioinformatics scientist" are recent and reflect a pace of technological change that is contributing to shortages of digital skills. Entirely new fields of tuition are needed, such as dedicated programmes for the autonomous vehicle industry. Existing curricula may also need to change. Too few students learn the fundamental role of logic in AI. Many schools barely teach data analysis, and more multidisciplinary education is needed.

Measures are required to address the fact that in many countries, in some subjects, such as AI, male students far outnumber female students. Digital technologies such as virtual reality could also facilitate skills development, as is happening in industry.

Committing to public sector research

Publically financed basic research has often been critical to advances in digital technology. A recent levelling off – and in certain cases decline – in government support for research in some major economies is a concern. The complexity of some emerging digitally based technologies exceeds the research capacities of even the largest individual firms. This necessitates a spectrum of public-private research partnerships. Interdisciplinary research is also essential. Policies on hiring, promotion and tenure, and funding systems that privilege traditional disciplines, may impede interdisciplinary research. Scientists working at the interface between disciplines need to know that opportunities for tenure are not jeopardised by doing so.

Building expertise in government

Without governments fully understanding technologies and sectors, opportunities to benefit from digital technologies might be lost. Calls to regulate AI highlight the need for expertise in government, such that any regulation of this fast-evolving technology does more good than harm. Technical expertise in government will also help to avoid unrealistic expectations about new technologies. As a wide array of critical systems become more complex, mediated and interlinked by code, governments also need improved understanding of complex systems. And as innovation agendas quickly evolve, governments also need to be flexible and alert to change. They must likewise ensure the availability of key infrastructures. For instance, broadband networks – especially fibre-optic connectivity – are essential to Industry 4.0.

To use digital science and innovation policy (DSIP) systems to help formulate and deliver science and innovation policy, governments must: ensure the interoperability of the data sets involved; prevent misuses of DSIP systems in research assessments; and, manage the roles of non-government actors, particularly the private sector, in DSIP systems.

1 An overview of key developments and policies

Alistair Nolan

Directorate for Science, Technology and Innovation, OECD

Chapter 1 summarises the main themes and policy lessons examined in the rest of the report. It provides background to the broader policy concerns facing OECD countries. It also introduces topics not considered elsewhere in the report, particularly in connection with artificial intelligence in science; using digital technology to deliver skills in science, technology, engineering and mathematics; possible targets for public research; and blockchain in science. The chapter also discusses potential uses of digital technology for policy making and implementation, mainly linked to various forms of collective intelligence. These essentially untapped opportunities – such as self-organised systems for funding allocation, and prediction markets – might have significant benefits for science, technology and innovation. They invite further study and, possibly, pilot testing.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

In 2015, in their joint declaration (OECD, 2015), ministers from OECD countries and partner economies, at the OECD Ministerial Meeting in Daejeon (Korea), recognised that digital technologies are revolutionising science, technology and innovation (STI). The ministers asked the OECD to monitor this transformation.

During 2017 and 2018, the OECD's "Going Digital" project comprehensively examined digital technology's economic and social impacts (OECD, 2019). The resulting report, *Going Digital: Shaping Policies, Improving Lives*, presents a strategy for policy making in the digital age. Complementing that report, this publication examines digitalisation's effects on STI and the associated consequences for policy. It draws mainly on work performed under the aegis of the OECD's Committee for Scientific and Technological Policy.

Apart from this overview, the publication has six other chapters:

Chapter 2 ("How are science, technology and innovation going digital? The statistical evidence") presents recent statistical evidence of key developments in the digitalisation of STI. It also reviews current and future measurement priorities.

Chapter 3 ("Digital technology, the changing practice of science and implications for policy") focuses on digitalisation and open science, and the associated policy consequences.

Chapter 4 ("Digital innovation: Cross-sectoral examples and policy implications") explores the many ways that digital technology is affecting innovation in firms, and the priorities for innovation policy in the digital age.

Chapter 5 ("Artificial intelligence, digital technology and advanced production") discusses digital technology in advanced manufacturing.

Chapter 6 ("Digitalisation in the bioeconomy: Convergence for the bio-based industries") explains the fast-evolving applications of digital technology in bio-based science and industry, and the priorities for government action.

Chapter 7 ("The digitalisation of science and innovation policy") reviews developments in digital information systems that support policy for STI, what these systems could look like in future, and what policy makers should do to maximise their potential.

Why does digitalisation matter?

The importance of digitalisation in STI is hard to overstate. Today, it is usual to view the future of STI through the lens of digitalisation's projected impacts. Carlos Moedas, the EU Commissioner for Research, Science and Innovation, recently announced that the Ninth EU Framework Programme for Research and Innovation will focus on digitalisation, beginning in 2021 (Zubașcu, 2017). Digitalisation also makes the present moment unique in the history of technology. As the technology commentator Kevin Kelly observed, "This is the first and only time the planet will get wired up into a global network" (Kelly, 2013). Furthermore, digitalisation's impacts are just beginning. Around a century passed before the full effects of earlier technology revolutions, linked to steam and electricity, became clear. By those standards, the digital revolution has generations to go.

Digitalisation is ubiquitous in STI in part because its effects are both microscopic and macroscopic. At the microscopic level, for example, researchers recently stored 200 megabytes (MB) of high-definition video and books in deoxyribonucleic acid (DNA) (see Chapter 6). At the macroscopic level, new digital technology means that a standard 10-pound satellite can capture better images of any point on Earth than a 900-pound satellite 20 years ago (Metz, 2019).

If anything, this publication illustrates that digitalisation's effects are deeper than most media reports reflect. Areas of research not traditionally associated with digitalisation, and on which advanced economies depend, from materials science to biology, are increasingly digital in character. At the same time, digital technology is changing the processes of science and enlarging its scope.

In STI, the pace of change brought by digitalisation is also striking. In all likelihood, no one foresaw in 2007 that ten years later more than a million people would be working in companies labelling and annotating data and images for machine-learning systems (Hutson, 2007). A decade ago, few anticipated how far artificial intelligence (AI) would progress in generating scientific hypotheses, scanning scientific literature and orchestrating experiments performed by robots. Similarly, until recently, only a few devotees understood distributed ledger technologies (DLTs), much less the possibility of combining AI and DLTs such that each amplifies the other (Corea, 2017).

Digitalisation is also facilitating convergence among technologies, a hallmark of innovation. There are several reasons for this convergence. Digital technologies can be combined – more easily than many other technologies – because of the shared numerical basis of different digital devices. Moreover, as it progresses, science can represent more of the natural world in the form of digital information. For example, as Chapter 5 shows, materials science is advancing in a transformational way because of the growing ability to observe, represent in computer models and then simulate the properties of a material's microstructure.

Convergence between the digital and biological worlds also reflects the relatively new understanding that life itself is informational and algorithmic (Valiant, 2013). Miniaturisation, which digital technology propels, likewise facilitates convergence. For instance, millimetre-sized computers could become common in the next decade (Biles, 28 September 2018). Such devices are likely to converge with medical technologies, for example in monitoring disease processes from inside the body.

Recent achievements in STI opened by digital technologies are extremely diverse, which reflects the technology's general-purpose character. In 2014, for example, Japan introduced the first trillion-frame-a-second camera, which gives scientists new ways to explore complex ultrafast phenomena. Supercomputers partition the globe into tens of thousands of digital units to simulate local weather, improving the accuracy of weather prediction. Indeed, a seven-day weather forecast in 2018 is as accurate as a two-day forecast 50 years ago (Fischer, 2018). The firm Lex Machina blends AI and data analytics to assist patent litigation (Harbert, 2013). Using digital tools, and in a break from previous norms, consumers now innovate in significant ways in many industries. Furthermore, digitalisation is making science more collaborative and networked. In 2015, for instance, researchers working on the Large Hadron Collider published a paper with a record-breaking 5 154 authors.

The broader context in which science, technology and innovation are digitalising

The digitalisation of STI is directly relevant to many important short- and long-term policy challenges. Over recent decades, for example, labour productivity growth has declined in many OECD countries. Developing and adopting efficiency-enhancing digital production technologies, along with organisational changes, are necessary to counter this decline. Rapid population ageing means that raising labour productivity is ever more urgent; the dependency ratio in OECD countries is set to double over the next 35 years. Digital technology contributes to productivity in part by making the mixing and recombining of ideas easier, which facilitates innovation. Some evidence even suggests that innovation increasingly occurs by combining existing ideas rather than by forming new ones (Youn et al., 2015).

Demographic change is likely to exert long-term downward pressure on discretionary public spending in OECD countries. Relative to national incomes, this pressure could entail static or even reduced levels of public support for science and innovation (OECD, 2018a). A protracted period of slow growth could have a similar effect. Such scenarios raise the question of whether, and by how much, digital technology could increase the efficiency of policy.

A related and worrying possibility is that the productivity of science might be falling. Some scholars claim that science is becoming less productive. They argue, variously, that the low-hanging fruits of knowledge have now been picked, that experiments are becoming more costly, and that science must increasingly be done across complex boundaries between a growing number of disciplines.

Scientists are also flooded with data and information. The average scientist reads about 250 papers a year, but more than 26 million peer-reviewed papers exist in biomedical science alone.[1] In addition, the overall quality of scientific output may be declining. Freedman (2015) estimated that around USD 28 billion per year is wasted on unreproducible preclinical research in the United States alone.

Not everyone agrees that research productivity is faltering (Worstall, 2016). However, any slowdown would have serious implications for growth. Increased funding would be needed to maintain discovery at previous levels and to seed the innovations and productivity improvements necessary to cope with demographic change and public spending constraints. Any boost to research productivity spurred by digital technology, from open science to the wider use of AI, could be of structural importance.

If deployed effectively, digitalisation could also help accelerate science and technology's ability to resolve global challenges. Environmental challenges include a warming atmosphere, loss of biodiversity, depleted topsoil and water scarcity. Health challenges include threats of disease – from multidrug-resistant bacteria to new pandemics. Demographic challenges include the consequences of ageing populations and the pressing need to treat neurodegenerative diseases. Breakthroughs in science and technology are necessary to address such challenges, and to do so cost-effectively.

While this report describes many ways in which digitalisation can strengthen STI, it also examines policy challenges created by digital technology. For example, owing to digitalisation, technology choice may be becoming more complex, even for large firms. One eminent venture capitalist recently wrote:

> *"Many of my friends at big companies tell me that 'what is 5G ?' floats around a lot of corporate headquarters almost as much as 'what is machine learning?'" (Evans, 2019).*

Digitalisation might also widen capability gaps in science across countries, owing to the uneven distribution of complementary assets such as computational resources, human capital and data access. In addition, the complex digital systems that underpin vital infrastructures, from transport networks to financial markets, might become more difficult to manage safely. Issues such as how to cope with so-called "predatory" online science journals (see Chapter 3), and how to keep personal research data anonymous, illustrate that new (and useful) applications of digital technology can generate new policy concerns.

Digitalisation also creates the need for new thinking about institutions and norms, both public and private. For example, in the public sector, governments in a number of countries are considering whether commissions for AI and robotics might be necessary. Similarly, in the private sector, as AI voice assistants become increasingly lifelike, firms must decide if customers should have the right to know that they are talking with machines (Ransbotham, 21 May 2018). The rapid pace of developments in digital technology may also require that regulatory processes become more anticipatory.

Digitalisation also raises other more far-reaching challenges, which this report does not tackle. What, for instance, should policy makers do about corrosive social and psychological effects that stem from the seepage of digital technology into much of everyday life?

Measuring the digitalisation of science and innovation

Chapter 2 provides a statistical context for the rest of the publication. It addresses measurement challenges and reports statistics on some key trends in the digitalisation of science and innovation. To that end, it draws principally on work under the OECD's Working Party of National Experts on Science and Technology Indicators.

The chapter examines four broad dimensions of the digital transformation of science: i) adoption of facilitating digital practices and tools; ii) access to digitised scientific outputs, especially publications, data and computer code; iii) use and further development of advanced digital procedures to make research more data-driven; and iv) communication of scientists' work and how this influences the way scientists are rewarded.

Overall, while digital activity in science is pervasive, there is considerable room to better exploit the potential of digital technology, particularly advanced tools. Findings in this chapter include the following:

- Digital technology facilitates sharing of scientific knowledge. However, OECD analysis reveals that, one year later, 60% to 80% of content published in 2016 was only available to readers via subscription or payment of a fee.
- Less than half of respondents in all science fields deliver data or code to a journal or publisher to support their publications.
- One-third of research and development (R&D) performed and funded by companies in the United States is software-related. OECD research suggests that for companies using advanced digital technologies, the odds of reporting innovations are doubled. A positive relationship also exists between the *development* of technologies and innovation, especially product innovation.
- From 2006 to 2016, the annual volume of AI-related publications grew by 150%, compared to 50% for indexed scientific publications overall. The People's Republic of China (hereafter "China") is now the largest producer of AI-related science, in terms of publications. The country is also fast improving the quality of its scientific output in this area.
- Public funding of science relating to AI is growing significantly, with a spate of recent policy and funding announcements. However, comparisons across countries are difficult because AI does not fit into pre-established taxonomies of R&D funding. Indeed, available data systems are ill equipped to address queries about subject areas supported by publicly funded research. Addressing this shortcoming is an OECD priority (through the "Fundstat" pilot project). The OECD has also begun to map trends in research funding for AI using institutional case studies, as Chapter 2 illustrates with two examples from the United States.
- At both doctorate and master's levels, many more men than women graduate in information and communication technology (ICT). ICT doctorate holders are especially likely to have been born abroad, exposing this population to policies that change residential or nationality requirements. Holders of doctorates in ICT are also more mobile across jobs than their counterparts. For example, in the United States, 30% of ICT doctorate holders changed jobs in the last year compared to 15% on average across other fields.
- Data from the OECD International Survey of Scientific Authors show that younger scientists are more likely to engage in all dimensions of digital behaviour.

Digitalisation, science and science policy

Chapter 3 shows that digitalisation is bringing change to all parts of science, from agenda setting, to experimentation, knowledge sharing and public engagement. Digital technology is facilitating a new paradigm of open science, a term referring to efforts to make scientific processes more open and inclusive. Open Science has three main pillars: open access (OA) to scientific publications and information; enhanced access to research data; and broader engagement with stakeholders. Together, the three pillars could increase the efficiency and effectiveness of science and speed the translation of research findings into innovation and socio-economic benefits. However, transitioning to open science requires the management of policy tensions associated with each pillar.

In his book *Imagined Worlds*, the physicist Freeman Dyson observed that there have been seven concept-driven revolutions in science during the past 500 years (Dyson, 1998). These revolutions are associated with the names of Copernicus, Newton, Darwin, Maxwell, Freud, Einstein and Heisenberg. During roughly the same period there were around 20 tool-driven revolutions, from the telescope in astronomy to X-ray diffraction in biology. Today, ICT is an evolving tool creating revolutionary change in science.

Many of the processes and outputs of science also improve digital technology. For example, the Laser Interferometer Gravitational-Wave Observatory, which detected cosmic gravitational waves, yielded new algorithms for detecting small signals in noisy data. And physicists designing the Large Hadron Collider federated computing systems at hundreds of sites to analyse petabytes of data, further developing grid computing.

Accessing scientific information

Emerging OA publishing models and pre-print servers, mega-journals, institutional repositories and online information aggregators are simplifying access to scientific information. However, the new era brings challenges compared to traditional specialised journals that published scientific research after peer review. It is less clear how editorial and peer review processes will work and how the academic record will be maintained and updated over time. There is considerable concern about the number of "predatory" online journals that charge authors for publication but carry out little or no quality control. It is important to identify predatory journals publicly and revise any funding mandates or other incentives that inadvertently encourage publication in such journals.

Digital tools can support the publication of scientific papers in several ways. Stimulated by a growing global scientific community, and by academic pressure to publish, the volume of scientific papers is vast and growing. ICT can help organise, share and analyse this growing volume of scientific information. At the same time, online open lab notebooks such as Jupyter provide access to primary experimental data and other information. Researchers are also employing AI to scrutinise suspicious scientific research and identify falsified data (Sankaran, 2018). Such tools depend on the broad adoption of standards and unique digital identifiers, which policy can facilitate.

Many science funders mandate OA publication, but academic careers, and in some cases institutional funding, are largely determined by publishing in high-impact, pay-for-access journals. Incentives and changes to evaluation systems need to match funders' mandates in order to transition faster to OA publication. A stronger focus on article-based metrics rather than journal impact factors is one way forward. New indicators and measures will also be required to incentivise data sharing.

A tiered publication process might emerge to address the challenges of using digital tools. Sharing and commenting on scientific information could occur earlier, with only some findings eventually published in journals. Some fields of research are testing open post-publication peer review, whereby the wider scientific community can discuss a manuscript. Such a process has strengths: transparent public discussion among peers gives incentives for sound argumentation, for instance. But it could also have weaknesses if, for example, reviewers making false or erroneous comments capture the process. However, with proper safeguards, post-publication peer review could bolster the quality and rigour of the scientific record.

Enhancing access to research data

Policy responses are needed to enhance access to research data. The OECD first advocated for greater access to data from publicly funded research in 2006. Since then, tools to enable greater access have improved, and guidelines and principles have been widely adopted. Nevertheless, as the following points illustrate, obstacles still limit access to scientific data:

- **The costs of data management are increasing, straining research budgets**. Science funders should treat data repositories as part of research infrastructure (which itself requires clear business models).
- **A lack of policy coherence and trust between communities hinders data sharing across borders**. The sharing of public research data requires common legal and ethical frameworks. Through such fora as the Research Data Alliance, funders should co-ordinate support for data infrastructure. New standards and processes, such as safe havens for work on sensitive data, could also strengthen trust, as might new technology such as blockchain.

- **Science must adapt its governance and review mechanisms to account for changing privacy and ethical concerns**. For example, to use human subject data in research requires informed consent and anonymisation. However, anonymising personal data from any given source might be impossible if new ICTs can link it to other personal data used in research. Transparent, accountable, expert and suitably empowered governance mechanisms, such as institutional review boards and/or research ethics committees, should oversee research conducted with new forms of personal data.
- **Strategic planning and co-operation are required to build and provide access to cyber-infrastructure internationally**. Global bodies such as the aforementioned Research Data Alliance can help develop community standards, technical solutions and networks of experts.
- **The skills needed to gather, curate and analyse data are scarce**. New career structures and professions – such as "data stewards" – need to be developed for data management and analysis.

Broadening engagement with science

Engagement with a broader spectrum of stakeholders could make scientific research more relevant. Digitalisation is opening science to a variety of societal actors, including patient groups, non-governmental organisations, industry, policy makers and others. Such opening aims to improve the quality and relevance of science and its translation into practice. Societal engagement can enhance the entire research process, from agenda setting to co-production of research and dissemination of scientific information. Perhaps the most critical area of enlarged engagement is in setting priorities for research. If well designed, a more inclusive process of agenda setting could make research more relevant and might even generate entirely new research questions.

Recent years have seen the expansion of "citizen science", whereby scientific research is conducted or supported through ICT-enabled open collaborative projects. ICT is helping science elicit input from the networked public to label, generate and classify raw data, and draw links between data sets. ICT is also creating opportunities for the networked public to take part in novel forms of discovery. For instance, by playing a video game – Eyewire – over 265 000 people have helped neuroscientists develop thousands of uniquely detailed neuronal maps, colour-coding over 10 million cell sections and generating data on neuron function (Princeton University, 2018). Whether, and how best, to expand citizen science requires answers to a number of questions. These include how to break complex research projects into parallel subtasks that do not depend on understanding the entire project. Crowdfunding of science is also emerging. It appears to provide opportunities for small-scale but meaningful funding for young scholars with risky research projects.

Digital technology could benefit science by levering collective input in other ways. For example, recent research suggests that digital technology could help draw on the collective insight of the entire scientific community to improve allocation of public research funds (Box 1.1).

Artificial intelligence for science

AI might increase productivity in science at a time when – as discussed earlier – some evidence suggests research productivity may be falling (Bloom et al. 2017). AI is being used in all phases of the scientific process, from automated extraction of information in scientific literature, to experimentation (the pharmaceutical industry commonly uses automated high-throughput platforms for drug design), large-scale data collection, and optimised experimental design. AI has predicted the behaviour of chaotic systems to distant time horizons, tackled complex computational problems in genetics, improved the quality of astronomical imaging, and helped discover the rules of chemical synthesis (Musib et al., 2017). Today, AI is regularly the subject of papers published in the most prestigious scientific journals.

Recent drivers of AI in science

AI in various forms has assisted research for some time. In the 1960s, the AI program DENDRAL helped identify chemical structures. In the 1970s, an AI known as Automated Mathematician helped perform mathematical proofs. Several key developments explain the recent rise of AI and ML. These include vast improvements in computer and AI software, much greater data availability and scientists' access to open-source AI code (King and Roberts, 2018).

Box 1.1. Collective intelligence to help allocate science funding

Bollen et al. (2014) and Bollen (2018) examine a new class of Self-Organized Funding Allocation (SOFA) systems to address issues associated with peer review. Peer review is the dominant approach to assessing the scientific value of proposals for research funding. However, critique of peer review is mounting. A major concern is the opportunity cost of scientists' time. For example, one study in Australia found that 400 years of researchers' time was spent preparing unfunded grant proposals for support from a single health research fund (Herbert, Barnett and Graves, 2013). Peer review has other drawbacks, too. The expertise in review panels is not interchangeable: many successful grant applications would be rejected if panel membership changed randomly (Graves, Barnett and Clarke, 2011). Some studies also show that peer review is less favourable to minorities, women and unconventional ideas.

To lower administrative overheads and improve funding allocation, Bollen et al. (2014) propose a SOFA system that would work like this: funding agencies would give all qualified scientists an unconditional and equal baseline amount of money each year. Scientists would then distribute a fixed percentage of their funding to peers who they think would make best use of the money. Every year, all scientists would therefore receive a fixed grant from their funding agency and an amount passed on by peers. Scientists could log on to their funding agency's website and simply select the names of scientists to whom they wish to donate, and indicate the amount for each.

As funding circulates between scientists, it would come to reflect the funding preferences of the entire scientific community, not small review panels. Widely esteemed scientists, who also distribute a fixed share of the money they receive, would end up with greater influence on how funding is allocated overall. At the same time, because all scientists receive an unconditional yearly grant, they would have greater stability and autonomy for discovery. Funding levels would adjust as the collective perception of scientific merit and priorities evolve. Scientists would also have incentives to share research because if colleagues were positively impressed, more funding could follow. In addition, funding people rather than projects might provide scientists with more freedom to explore new research paths.

Individual distributions would be anonymous (to avoid personal influence) and subject to conflict of interest restrictions. For example, scientists might be prohibited from donating to themselves, advisees, colleagues at their own institution, etc. By tuning distribution parameters, funding agencies and governments could still target research in ways that promote policy goals, such as funding under-represented communities. Existing funding systems could also link to a SOFA to complement peer review and maintain societal accountability.

Using millions of Web of Science records, simulation of a SOFA yielded a distribution of funding similar to that produced by grant review, but without a single proposal being prepared (Bollen et al., 2014). SOFAs merit further study and pilot testing. In 2018, the Dutch Parliament mandated the Netherlands Organisation for Scientific Research to explore a pilot study.

AI can also combine with robot systems to perform scientific research

Laboratory-automation systems can exploit techniques from AI to execute cycles of scientific experimentation. For instance, one system uses AI to analyse molecular models with desirable properties. A robot then tests the predictions by physically combining chemical samples and analysing the results. These results become inputs to continue improving the system's predictions (Knight, 2018). AI-enabled automation in science, especially in disciplines that require intensive experimentation, such as molecular biology and chemical engineering, has several potential benefits (King and Roberts, 2018):

- **Faster discovery**. Automated systems can generate and test thousands of hypotheses in parallel.
- **Cheaper experimentation**. AI systems can select more cost-effective experiments.
- **Improved knowledge/data sharing and scientific reproducibility**. Robots can automatically record experimental procedures and results, along with the associated metadata, at no additional cost (recording the data, metadata and procedures adds up to 15% to the total costs of experimentation by humans).

Challenges still exist in using AI and ML in science. Scientific models developed by ML are not always explainable. This is partly because ML poses general challenges of interpretability. It is also because laws that underlie an AI/ML-derived model might depend on knowledge that scientists do not yet possess. Furthermore, some scientific laws might be so complex that, if discovered by an AI/ML system, experts would still struggle to understand them (Butler et al., 2018).

As AI plays a greater role in science, certain policies will grow in importance. These include policies that affect access to high-performance computing (HPC) (the computational resources essential to some leading-edge fields of research, including in AI, can be extremely expensive), skills (discussed later in this chapter), and access to data (such as standardisation for machine readability of scientific datasets). Policies on access to data not only matter for training AI systems, and for the scope of scientific problems on which AI can operate, they also matter for reproducibility. Without access to underlying data, the validity of conclusions arrived at by complex algorithms – some of which may already have a "black box" character – will be open to question. AI in science also raises new and so far unanswered questions: for instance, Should machines be included in academic citations? Will intellectual property (IP) systems need adjustment in a world in which machines can invent?

Digitalisation and innovation in firms

Digitalisation is also shaping innovation throughout the economy, generating new digital products and services and enhancing traditional ones with digital features. Chapter 4 shows that four trends characterise innovation in the digital age: data are a key innovation input; digital technologies enable services innovation; innovation cycles are speeding up; and, digital technology is making innovation more collaborative. The following paragraphs describe these four trends.

Innovation processes increasingly rely on data. They use data to explore product and services development, and gain insight on market trends; to understand the behaviour of competitors; to optimise development, production and distribution processes; and to tailor products and services to specific or fluctuating demand.

More diverse and voluminous types of data have driven the development of new business models. Such models include peer-to-peer accommodation (e.g. Airbnb) and on-demand mobility services (e.g. Uber). Other examples are platforms to search, compare and book accommodation and transportation options (e.g. Booking.com), digitalised invoice discounting (e.g. Due Course) and digital co-operatives (the latter described in Scholz and Schneider, 2019). All these new business models are enabled by the availability and capacity to analyse (large volumes of) real-time data.

Digital technologies also facilitate services innovation. Examples include new digitally enabled services, such as predictive maintenance services using the Internet of Things (IoT) and web-based business services. Manufacturers increasingly offer services enabled by digital technology to complement the goods they produce, and service providers increasingly invest in digital technology to improve their activities. Large retailers, for instance, invest intensively in the IoT to improve inventory management.

Digital innovations such as generative design software and three-dimensional (3D) printing speed innovation cycles by accelerating product design, prototyping and testing. ICTs also enable the market launch of product beta versions that can be updated to incorporate consumer feedback. For example, GE Appliances' FastWorks system involves consumers early in the development of new products such as refrigerators.

Digital technology is also making innovation ecosystems more open and diverse. Firms increasingly interact with research institutions and other firms for three reasons. First, they gain access and exposure to complementary expertise and skills. Second, collaboration helps share the costs and risks of uncertain investments in digital innovation. Third, reduced costs of communication allow greater interaction, regardless of location. One example of a collaboration using digital technology is the SmartDeviceLink Consortium, an open-source platform for smartphone app development for vehicles created by Ford and Toyota.

Does innovation policy need to be adapted for the digital age?

Innovation increasingly involves the creation of digital products and processes. Consequently, policies for innovation need to align with generic features of digital technology. In this connection, Chapter 4 proposes overarching considerations for policy design. These considerations include access to data for innovation; providing suitably designed support and incentives for innovation and entrepreneurship; ensuring that innovation ecosystems support competition; and supporting collaboration for innovation. The following paragraphs further describe these considerations.

Ensuring access to data for innovation

To favour competition and innovation, data access policies should aim to ensure the broadest possible access to data and knowledge (incentivising sharing and reuse). At the same time, they must respect constraints regarding data privacy, ethics, intellectual property rights (IPRs), and economic costs and benefits (i.e. incentives to produce data). To foster data-driven innovation, some governments provide access to data generated by public services, such as urban transportation. Policy can also facilitate the emergence of markets for data.

Restricting cross-border data flows could be harmful. Manufacturing, for instance, creates more data than any other sector of the economy, and cross-border data flows are set to grow faster than growth in world trade (Chapter 5). Research suggests that restricting such flows, or making them more expensive, for instance by obliging companies to process customer data locally, can raise firms' costs and increase the complexity of doing business. This is especially the case for small and medium-sized enterprises (SMEs).

As businesses innovate with data, new policy issues are likely to arise. One such issue is whether firms should have legal data portability rights. Companies such as Siemens and GE are vying for leadership in online platforms for the IoT. Such platforms will become repositories of important business data. If companies had portability rights for non-personal data, competition among platforms could grow, and switching costs for firms could fall. Another incipient policy issue concerns the treatment of non-personal sensor data. Individual machines can contain multiple components made by different manufacturers, each with sensors that capture, compute and transmit data. This raises legal issues. For example, which legal entities should have rights to own machine-generated data and under what conditions? Who owns rights to data if a business becomes insolvent? More broadly, are provisions needed to protect data transmitted in value chains – say, between contractors and sub-contractors – from sale to or use by third parties?

Providing the right support and incentives for innovation and entrepreneurship

Government needs to be flexible and alert to change as innovation agendas quickly evolve. One approach to achieving policy responsiveness is the deployment and monitoring of small policy experiments, after which policies might be scaled up or down. In a context of rapid change, application procedures for innovation support instruments also need to be streamlined. For example, the Pass French Tech programme offers fast-growing start-ups simplified and rapid access to services (e.g. in financing, innovation and business development).

Policies should also address services innovation. Relevant measures might include projects to develop entirely new services using digital technologies such as the Smart and Digital Services Initiative in Austria. Other potential measures include policies to help manufacturing SMEs to develop new services related to their products (e.g. service design vouchers for manufacturing SMEs in the Netherlands).

Ensuring that innovation ecosystems support competition

Markets in which digital innovation is important are subject to rapid innovation (a source of competition) and scale economies (a source of persistent concentration). Competition authorities and innovation policy makers should work together to ensure the contestability of these markets. They should also address the role of data as a source of market power.

Supporting collaboration for innovation

Digital technology permits new ways for firms and institutions to collaborate for innovation. These new mechanisms include crowdsourcing, open challenges and so-called living labs. The latter typically involve concurrent research and innovation processes within a public-private-people partnership. New research and innovation centres, often public-private partnerships, help multidisciplinary teams of public researchers and businesses work together to address technology challenges. Such centres often have innovative organisational structures. Examples include Data61 in Australia and Smart Industry Fieldlabs in the Netherlands.

Digitalisation and the next production revolution

Digital technologies are at the heart of advanced production (Chapter 5). The widely used term “Industry 4.0” refers to a new paradigm in which all stages of manufacturing are controlled and/or connected by digital technology. These stages range from product design, fabrication and assembly to process control, supply-chain integration, industrial research and product use. Industry 4.0 technologies can raise productivity in many ways, from reducing machine downtime when intelligent systems predict maintenance needs, to performing work faster, more precisely and consistently with increasingly autonomous, interactive and inexpensive robots. The digital production technologies in question are evolving rapidly. For instance, recent innovations permit 3D printing with novel materials such as glass, printing strands of DNA, and even, most recently, printing on gels using light (OECD, 2017; Castelvecchi, 2019).

AI in production

With the advent of deep learning using artificial neural networks – the main source of recent progress in AI – AI is finding applications in most industrial activities. Such uses range from optimising multi-machine systems to enhancing industrial research. Beyond production, AI is also supporting functions such as logistics, data and information retrieval, and expense management.

Several types of policy affect the development and diffusion of AI in industry. These include policies for education and training; access to expertise and advice; research support, policies on digital security, and liability rules (which particularly affect diffusion). In addition, while AI entrepreneurs might have the knowledge and financial resources to develop a proof-of-concept for a business, they sometimes lack the hardware and hardware expertise to build an AI company. As Chapter 5 describes, governments can help resolve such constraints.

Without large volumes of training data, many AI/ML models are inaccurate. Often, training data must be refreshed monthly or even daily. Data can also be scarce because many industrial applications are new or bespoke. Research may find ways to make AI/ML systems less data-hungry (and in some cases artificially created data can be helpful). For now, however, training data must be cultivated for most real-world applications. But many industrial companies do not have the in-house capabilities to exploit the value in their data, and are understandably reluctant to let others access their data. As Chapter 5 describes, some public programmes exist to bridge between company data and external analytic expertise. In addition, to help develop and share training data, governments can work with stakeholders to develop voluntary model agreements and programmes for trusted data sharing. More generally, governments can promote open-data initiatives and data trusts, and ensure that public data exist in machine-readable formats. While such actions are not usually aimed at industry, they can be helpful to industrial firms in incidental ways (for instance in research, or in demand forecasting that draws on economic data, etc.).

New materials and nanotechnology

Advances in scientific instrumentation, such as atomic-force microscopes, and progress in computational simulation, have allowed scientists to study materials in more detail than ever before. Powerful computer modelling can help build desired properties such as corrosion resistance into new materials. It can also indicate how to use materials in products.

Professional societies are working hard to develop a materials-information infrastructure to support materials discovery. This includes databases of materials' behaviour, digital representations of materials' microstructures and predicted structure-property relations, and associated data standards. Policy co-ordination at national and international levels could enhance efficiency and avoid duplicating such infrastructures.

Closely related to new materials, nanotechnology involves the ability to work with phenomena and processes occurring at a scale of 1 to 100 billionths of 1 metre. The sophistication, expense and specialisation of tools needed for research in nanotechnology – some research buildings must even be purpose-built – make inter-institutional collaboration desirable. Publicly funded R&D programmes on nanotechnology could also allow collaboration with academia and industry from other countries. The Global Collaboration initiative under the European Union's Horizon 2020 programme is an example of this approach.

Developing digital skills

Digitalisation raises demand for digital skills. For example, rapid improvements in AI systems have led to an overall scarcity of AI skills. Occupations like "industrial data scientist" and "bioinformatics scientists" are recent, reflecting a rate of technological change that is generating skills shortages. A dearth of data specialists is impeding the use of data analytics in business. Some countries also have too few teachers of computer programming (Stoet, 2016). A shortage of cybersecurity experts has led at least one university to recruit students to protect itself against hackers (Winick, 2018). Furthermore, the general-purpose nature of digital technology means that skills required to be a good scientist are also increasingly attractive in industry, adding to competition for talent (Somers, 2018).

Rising demand for digital skills has implications for income distribution and economic productivity. In terms of income distribution, for instance, lack of ICT skills in low-skilled adult populations in semi-skilled occupations places this demographic group at high risk of losing jobs to automation. In terms of productivity, the ability of education and training systems to respond to changing skills demand affects the pace of technology adoption.

Education and training systems must draw on information from all social partners

Skills forecasting is prone to error. Just a few years ago, few could have foreseen that smartphones would so quickly disrupt, and in some cases end, a wide variety of products and industries, from notebook computers and personal organisers to niche industries making musical metronomes and hand-held magnifying glasses (functions now available through mobile applications). Because foresight is inherently uncertain, education and training systems should draw on information about skills needs from businesses, trade unions, educational institutions and learners. Students, parents and employers also need access to data with which to judge the performance of educational institutions. In turn, resources in educational and training systems must flow efficiently to courses and institutions that best match skills demand. Institutions that play such roles include Sweden's job security councils and SkillsFutureSingapore.

New courses and curricula may be needed

New courses and curricula may be needed to keep pace with rapid changes brought on by digitalisation. Advances in digital technology may require entirely new fields of tuition, such as dedicated programmes for the autonomous vehicle industry. Existing curricula may also need to change. For example, software engineers are effectively becoming social engineers. Society might benefit if they were to learn civics and philosophy, subjects rarely taught in science, technology, engineering and mathematics programmes (Susskind, 2018).

In many countries, schools do not teach logic, and universities rarely teach logic outside of specialised courses. As a result, too few students learn the fundamental role of logic in AI. Many schools barely teach data analysis (King and Roberts, 2018). Various parts of this report also emphasise the need for greater multidisciplinary education. For instance, the bioeconomy increasingly requires degree programmes that combine biology, engineering and programming (Chapter 6). In addition, in many countries, male students far outnumber female students in some subjects, including AI. One recent survey of 23 countries found that, on average, 88% of researchers were male (Mantha and Hudson, 2018).[2]

Lifelong learning must be an integral part of work

In a context of significant technological change, lifelong learning must be an integral part of work. Achieving this demands greater collaboration between government and social partners to develop and/or fund appropriate programmes. Strong and widespread literacy, numeracy and problem-solving skills are critical, because these foundation skills provide the basis for subsequent acquisition of technical skills, whatever they may be in the future. Working with social partners, governments can help develop entirely new training programmes, such as conversion courses in AI for those already in work, and ensure effective systems of certification. Beyond technical know-how, workforce education can help impart other important skills, such as the ability to work well in teams and in complex social contexts, to be creative and exercise autonomy.

Many countries have far-reaching programmes to develop digital technology skills. Using online tuition, Finland aims to teach every citizen the basics of AI. All Finnish students in initial education learn coding. Estonia is using public-private partnerships to teach coding and robotics. And the United Kingdom's government recently committed up to GBP 115 million (EUR 134 million) for 1 000 students to complete doctoral degrees in AI. Digital technology is also creating novel ways to deliver skills (Box 1.2).

Box 1.2. Using digital technology to deliver skills

Digital technologies are beginning to facilitate skills development in new ways. In 2014, for example, Professor Ashok Joel and graduate students at Georgia Tech University created an AI teaching assistant – Jill Watson – to respond to online student questions. For months, students were unaware that the responses were non-human. iTalk2Learn, a European Union project, aims to develop an open-source intelligent platform for mathematics tutoring in primary schools. Researchers at Stanford University are developing systems to train crowdworkers using machine-curated material generated by other crowdworkers. In France, on an experimental basis, haptic technology – which allows a remote sense of touch – has shortened the time required to train surgeons, and promises many other applications.

Augmented reality (AR) uses computer vision to overlay objects in the user's field of view with data and annotations (such as service manual instructions). Tesla has applied for a patent for an "Augmented Reality Application for Manufacturing", built into safety glasses. With AR, skills such as those needed to repair breakdowns in complex machine environments will effectively become downloadable.

Virtual reality (VR) environments could improve the speed and retention of learning, as is beginning in industry. Using VR, Bell Helicopter reports reducing a typical six-year aircraft design process to six months. Furthermore, Walmart has put 17 000 VR headsets in its US stores for training. VR could also permit safe and costless "learning by doing" for beginners in fields where this is otherwise too dangerous or expensive.

The declining cost of VR and AR, and the integration of AR into mobile devices, should lower barriers to public participation in education, training and research. Elon Musk, for example, promises a high-definition VR live-stream of a future SpaceX moon mission (Craig, 2018).

Facilitating the diffusion of digital technologies and tools

Most countries, regions and companies are primarily technology users, rather than technology producers. For them, technology diffusion and adoption should be priorities. Technology diffusion helps raise labour productivity growth and may also lower inequality in wage growth rates. Policy makers tend to acknowledge the importance of technology diffusion, but to underemphasise it in the overall allocation of resources.

Even in the most advanced economies, diffusion of technology can be slow or partial. For example, a survey of 4 500 German businesses in 2015 found that only 4% had implemented digitalised and networked production processes or had plans to do so (Chapter 5). One recent study examined 60 manufacturers in the United States with annual turnovers of between USD 500 million and USD 10 billion. The study found that just 5% had mapped where AI opportunities lie within their company and were developing a strategy for sourcing the data AI requires, while 56% had no plans to do so (Atkinson and Ezell, 2019).

New digital technologies may make diffusion more difficult

Certain features of new digital technologies could hinder diffusion. As technology becomes more complex, potential users must often sift through burgeoning amounts of information on rapidly changing technologies and knowledge requirements. Once the technology is chosen, deployment can pose difficulties as well. Even the initial step of collecting sensor data can be daunting. A typical industrial plant, for example, might contain machinery of many vintages from different manufacturers. This machinery may have control and automation systems from different vendors, all operating with different communication standards. To deploy AI, firms must often invest in costly information technology upgrades to merge data from disparate record-keeping systems (consumer and supply-chain transactions are often separate, for instance). Firms also have unique challenges – from proprietary data types to specific compliance requirements. These conditions may require further research and customisation (Agrafioti, 2018). Difficulties in determining the rate of return on some AI investments may also hinder adoption. Furthermore, to understand how an AI system works, staff may need to take time away from other critical tasks (Bergstein, 2019). In addition, the expertise required for all of the above is scarce.

Institutions for diffusion can be effective, if well designed

As Chapters 4 and 5 discuss, various micro-economic and institutional settings facilitate diffusion. These range from supportive conditions for new-firm entry and growth, to economic and regulatory frameworks for efficient resource allocation. In addition to enabling framework conditions, effective institutions for technology diffusion also matter. Institutions for diffusion range from applied technology centres such as the Fraunhofer Institutes in Germany to open technology mechanisms such as the Bio-Bricks Registry of Standard Biological Parts.

New diffusion initiatives are emerging, often involving partnership-based approaches. An example is the US National Network for Manufacturing Innovation (NNMI). The NNMI uses private non-profit organisations as the hub of a network of company and university organisations to develop standards and prototypes in areas such as 3D printing and digital manufacturing. Some initiatives aim to facilitate the testing of new digital technology applications, such as by creating test beds, regulatory sandboxes, and state-of-the-art facilities as well as providing expertise. As Chapter 4 describes, the Industry Platform 4 FVG, in the Italian region of Friuli Venezia Giulia, is an example of an institution that offers access to testing equipment, prototyping tools and demonstration labs.

To strengthen science, and the interface between science and industry, governments should also support platform technologies. These could include biofoundries, distributed R&D networks, data curation and digital/ genetic data storage. This is a public role because the associated investment risks are too high for the private sector. Moreover, for the private sector such investments may not provide a clear route to market.

Technology diffusion institutions need realistic goals and time horizons

Effective diffusion is more likely under two conditions. First, technology diffusion institutions must be empowered and resourced to take longer-term perspectives. Second, evaluation metrics must emphasise longer-run capability development rather than incremental outcomes and revenue generation. Introducing new ways to diffuse technology also takes experimentation. Yet many governments want quick and riskless results (Shapira and Youtie, 2017).

Diffusion in SMEs involves particular challenges. In Europe, for example, as Chapter 5 describes, 36% of surveyed companies with 50 to 249 employees use industrial robots compared to 74% of companies with 1 000 or more employees. Such discrepant patterns of technology use reflect, among other reasons, the more limited availability of digital skills in SMEs. For instance, only around 15% of European SMEs employ ICT specialists compared to 75% of large firms (Box 1.3). As Chapter 4 discusses, traditional instruments to foster technology adoption by SMEs – such as innovation vouchers and training – have been redesigned to meet specific challenges of the digital age, and often use digital tools themselves (for example, the SME 4.0 Competence Centres in Germany).

Box 1.3. Diffusing digital technology to SME : Some key considerations

Various measures can help diffuse digital technology to SMEs, including:

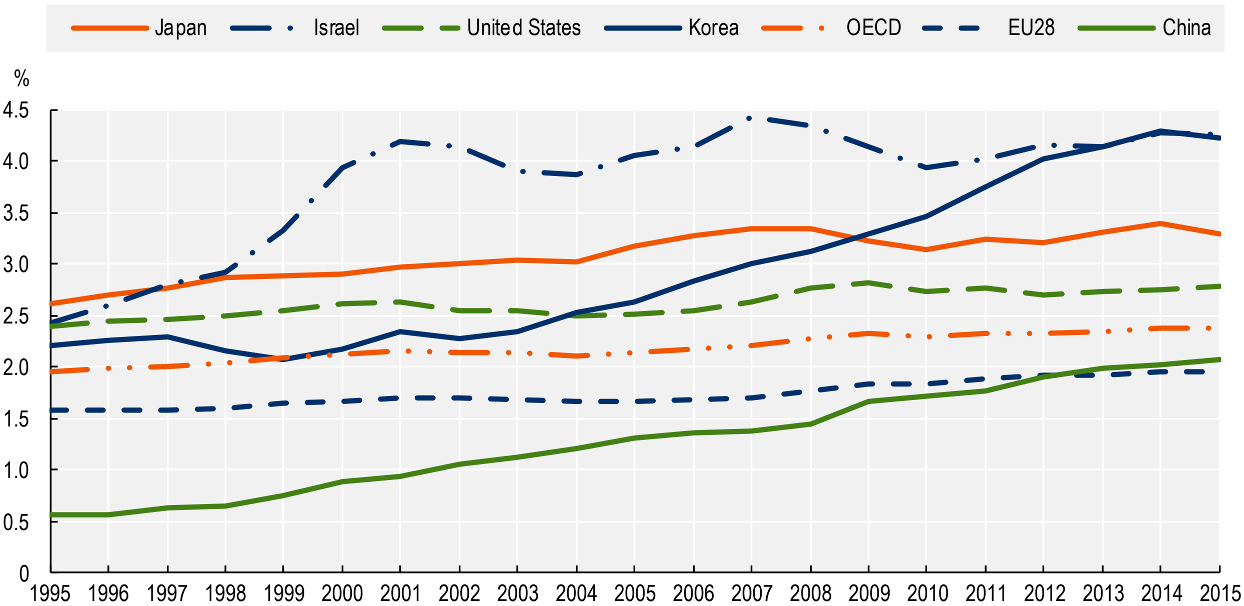
- Systematising key information for SMEs. For example, Germany's Industry 4.0 initiative has documented over 300 uses cases of applications of digital industrial technologies, along with contacts to experts (www.plattform-i40.de).
- Providing information on the expected return on investments in new technologies, as well as information on essential complementary organisational and process changes.
- Providing signposts to reliable sources of SME-specific expertise, because the skills to absorb information are scarce in many SMEs. For example, as part of its "SMEs Go Digital Programme", Singapore's TechDepot provides a list of pre-approved digital technology and service solutions suited to SMEs. And Tooling U-SME, an American non-profit organisation owned by the Society of Manufacturing Engineers, provides online industrial manufacturing training and apprenticeship programmes.
- Providing facilities where SMEs can test varieties and novel combinations of equipment to help de-risk prospective investments.

Committing to public sector research

The technologies discussed in this publication have arisen because of advances in scientific knowledge and instrumentation. Publicly financed basic research has often been critical. For decades, for example, public funding supported progress in AI, including during periods of unproductive research. Today, AI attracts huge private investment. In this context, a recent hiatus – and in certain cases decline – in government support for research in some major economies is a concern (Figure 1.1).

Figure 1.1. Trends in total R&D performance, OECD countries and selected economies, 1995-2015

As a percentage of GDP



Note: R&D = research and development; GDP = gross domestic product.

Source: OECD (2017), *OECD Science, Technology and Industry Scoreboard 2017: The Digital Transformation*, http://dx.doi.org/10.1787/9789264268821-en.

StatLink https://doi.org/10.1787/888934075678

Multidisciplinary research

Various chapters in this publication stress the importance of multidisciplinary research. The importance of understanding the interplay between disciplines reflects the need to address complex and cross-cutting problems, the fact that new disciplines are born as knowledge expands, and the increased complexity of scientific equipment. It also reflects the frequent need to bring together different digital technologies. For example, developing the potential of haptic technologies – not least for uses in education and training – requires the combination of electrical engineering (communications, networking), computer science (AI, data science) and mechanical engineering (kinaesthetic robots) (Dohler, 2017).

Policies on hiring, promotion and tenure, and funding systems that privilege traditional disciplines, may impede interdisciplinary research. Scientists need to know that working at the interface between disciplines will not jeopardise opportunities for tenure. Institutions that demonstrably support multidisciplinary research can provide useful lessons. Such cases include the United Kingdom's Interdisciplinary Research Collaborations, networks in Germany to support biomedical nanotechnology, and individual institutions such as Harvard's Wyss Institute for Biologically Inspired Engineering.

Public-private research partnerships

The complexity of some emerging digitally based technologies exceeds the research capacities of even the largest individual firms. This necessitates a spectrum of public-private research partnerships. For example, materials science relies on computational modelling, enormous databases of materials' properties and expensive research tools. It is almost impossible to gather an all-encompassing materials science R&D infrastructure in any single company or institute.

Many possible targets exist for government R&D and commercialisation efforts to continue progress in the digital revolution. These range from quantum computing to new mathematics for big data. Box 1.4 presents a small selection of these ideas.

Box 1.4. Public research goals relevant to the digital transformation of STI

Responding to the end of Moore's Law. In many digital devices, processing speeds, memory capacity, sensor density and accuracy, and even numbers of pixels are linked to Moore's Law (the law asserts that the number of transistors on a microchip doubles about every two years). Atomic-level phenomena now limit the shrinkage of transistors on integrated circuits. Some experts believe a new computing paradigm is needed. The current computing paradigm is based on von Neumann's design of the electronic computer. This architecture involves a channel for instructions that pass through one or more central processing units (CPUs) that retrieve data, compute and store results. This architecture, in which CPUs are a bottleneck, has not changed since 1948 (Damer, 2018). Hopes for significant advances in computing rest on research breakthroughs in optical computing (using photons instead of electrons), biological computing (storing data in and calculating using segments of DNA) and quantum computing.

Advancing the development of quantum computing, communication and information. Quantum technology has mostly been a theoretical possibility until recently, but Google, IBM and others are now trialling practical applications. In 2017, Biogen worked with Accenture and quantum software company 1QBit on a quantum-enabled application to accelerate drug discovery. Quantum technologies, if successful, could revolutionise certain types of computing. This would have strategic consequences for secure communication. Quantum computing still involves major research and technical challenges. For example, most of today's quantum devices require operating temperatures near to absolute zero, as well as the development of new materials. Quantum computing, communication and information is becoming a priority for a number of governments. China plans to open a National Laboratory for Quantum Information Sciences in 2020, with USD 10 billion of investment.

Creating more capable AI. Brooks (15 July 2018) observes that AI does not yet possess the object recognition of a two-year-old, the language understanding of a four-year-old, the manual dexterity of a six-year-old or the social understanding of an eight-year-old. While businesses far outspend governments on R&D for AI, much of this R&D focuses on application rather than breakthroughs in knowledge. Furthermore, Jordan (2018) observes that much research on human-like AI is not directly relevant to the major challenges involved in building safe intelligent infrastructures such as in medical or transport systems. Unlike human-imitative AI, such critical systems must have the ability to cope with "cloud-edge interactions in making timely, distributed decisions and they must deal with long-tail phenomena whereby there is lots of data on some individuals and little data on most individuals. They must address the difficulties of sharing data across administrative and competitive boundaries."

Many research challenges are important for public policy. These range from the explainability of AI, to the robustness of AI systems (image-recognition systems can easily be misled), to how much *a priori* knowledge AI might need to perform difficult tasks. Jordan (2018) also describes a number of major open research questions in classical human-imitative AI research. These include the need to bring meaning and reasoning into systems that perform natural language processing and the need to infer and represent causality.

Developing technology- and sector-specific capabilities in government

Understanding major technologies is particularly important when these evolve quickly. For instance, one leading authority argues that converging developments in several technologies are about to yield a "Cambrian explosion" in robot diversity and use (Pratt, 2015). Without governments fully understanding technologies and sectors, strategic opportunities to benefit from digital technologies might be lost.

Chapter 5 describes an example of this challenge. Technical and sector experts in the United States understand that a strategic opportunity exists to use metals-based 3D printing in commercial aviation. However, as an immature technology, metals-based 3D printing does not meet the stringent tolerances and high reliability needed in aviation. Targeted policy could change this, with measures ranging from funding and curating databases on materials' properties, to brokering essential data-sharing agreements (DSAs) across government laboratories, academia and users of metals-based 3D printing. Perceiving and successfully acting on such opportunities require technical and sectoral expertise.

Regulation, when used, also needs deep technology and industry-specific understanding. The effects of regulation on innovation can be complex, of uncertain duration and ambiguous, making them difficult to predict. Calls to regulate AI highlight the need for expertise in government, so that any regulation of this fast-evolving technology does more good than harm. Developments in fast-changing technologies such as AI may also require that regulatory processes become more anticipatory and innovative. As Chapter 4 describes, three policy domains require a sectoral approach for designing new initiatives: data access policies, given the diversity of data types in different sectors; digital technology adoption and diffusion policies; and policies supporting the development of sectoral applications of digital technologies.

Technical expertise in government will help avoid unrealistic expectations about new technologies, especially those emerging from science (such as quantum computing). New discoveries and technologies often attract hyperbole. No more than 6 years ago, for example, massive open online courses (MOOCs) were widely held to represent a democratising transformation in postsecondary education. However, recent research shows that less than 12% of MOOC students return for a second year, and most students come from affluent families in rich countries (Reich and Ruipérez-Valiente, 2019).

Similarly, many hailed Bitcoin as the democratisation of money. Indeed, a 2013 article in WIRED called Bitcoin "the great equalizer" (Hernandez, 2013). However, by 2017 just 1 000 users owned 40% of Bitcoin (Kharif, 2017). Public discussion of AI also involves wildly varying accounts of its likely impacts. AI-related hyperbole may even have particular psychological roots: experiments show that subjects unconsciously anthropomorphise AI and robots (Fussell et al., 2008).

Effective sectoral support requires, as a first step, mechanisms to strengthen policy intelligence. As Chapter 4 discusses, these mechanisms include roadmaps or sectoral plans prepared with industry and social partners. One example is the Sector Competitiveness Plans developed by Industry Growth Centres in Australia. Developing a shared vision for the future, with industry and social partners, is also useful.

Ensuring access to complementary infrastructures

Certain types of infrastructure help to utilise digital technology. These include HPC, cloud computing and fibre-optic connectivity. HPC is increasingly important for firms in industries ranging from construction and pharmaceuticals to the automotive sector and aerospace. In manufacturing, the use of HPC is going beyond applications such as design and simulation to encompass real-time control of complex production processes. However, like other digital technologies, manufacturing's use of HPC falls short of potential. A number of possible ways forward exist. SMEs could receive low-cost, or free, limited experimental use of HPC, while online software libraries/clearing houses could help disseminate innovative HPC software to a wider industrial base.

Industry 4.0 requires increased data sharing across production sites and company boundaries (Chapter 5). For example, BMW aims to know the real-time status of production equipment at every company that produces key components for its vehicles. Increasingly, machine data and data analytics, and even monitoring and control systems, will operate in the cloud. The cloud will also allow independent AI projects to start small, and scale up or down as required. Indeed, Google's chief AI scientist, Fei-Fei Li, argues that cloud computing will democratise AI.[3] Cloud computing will also increasingly help data sharing and analysis in science: Amazon Web Services, for instance, participates in the 1 000 Genomes Project, helping researchers to access and analyse vast amounts of cloud-based genetic data. However, cloud use varies greatly between small and large firms, and across countries. For example, only 20% of Austrian manufacturers used cloud computing in 2016. By comparison, in Finland, the country with the highest incidence of cloud use in manufacturing in the OECD, the rate was 69% (OECD, 2018b).

Broadband networks – especially fibre-optic connectivity – are also essential to Industry 4.0. Policy priorities here include overhauling laws governing the speed and coverage of communication services. Policies to promote competition and private investment, as well as independent and evidence-based regulation, have also helped extend coverage. In addition, new technology could expand services in underserved areas. A case in point is the delivery of broadband through "White Spaces", the gaps in radio spectrum between digital terrestrial television channels.

Improving digital security

Among other issues, digital technology is creating wholly new sources of risk. For example, as Chapter 5 observes, with respect to new materials, a novel risk could arise because, in a medium-term future, materials development processes based on computer simulations could be hackable. Chapter 6 notes that bio-production relies heavily on data, IP and research, all of which need protection from cyber-attack. Companies in the bioeconomy are elevating cybersecurity to a strategic imperative, but at a pace that lags behind their desire to adopt digital technologies. Enhancing trust in digital services is also critical to data sharing and, in some countries, uptake of cloud services.

While challenging to measure, digital security incidents appear to be increasing in terms of sophistication, frequency and influence. New digital security solutions are emerging, such as homomorphic encryption, through which data remains encrypted even when being computed on in the cloud. The technological race between hackers and their targets is nevertheless unrelenting. Government awareness-raising initiatives are important. SMEs, in particular, need to introduce or improve their digital security risk management practices.

Chapter 6 suggests that governments could encourage timely sharing of information on digital security threats. Public sector actors could also run cyber-attack simulations and share the lessons learned. Voluntary standards, regulations, industry programmes and information-sharing networks could draw attention to digital security enhancements. In addition, in public-private research partnerships, individual facilities could be encouraged to develop and validate methods for staff or external service providers to strengthen digital security. OECD (2019) includes detailed recommendations on digital security. These focus on managing rather than eliminating digital security risk – among individuals, firms and governments – because some degree of risk is inevitable.

Examining intellectual property systems in light of digitalisation

New digital technologies are raising new challenges for IP systems. 3D printing, for example, might create complications in connection with patent eligibility. For instance, if 3D-printed human tissue improves upon natural human tissue, it may be eligible for patenting, even though natural human tissue is not. Ensuring legal clarity around IPRs is also important for 3D printing of spare parts (when printed by anyone other than the original equipment manufacturer).

More fundamentally, a world in which machines can invent could require new patenting frameworks. For example, AI systems that automatically – and unpredictably – learn from many publicly available sources of information could complicate the task of identifying deliberate infringements of patent laws. In another example, a licensor might hold IP rights on an AI system and license this. The licensee might run the AI system using data on which it too has IP rights (as certain jurisdictions permit protection of data ownership). This might lead to an improvement in the AI system. A conflict might thereby arise with respect to ownership of the improved AI. Current IP law is also silent on the issue of whether AI can itself acquire IP rights.

All the chapters in this report address different types of standards. For instance, Chapter 5 shows that Industry 4.0 currently involves more than 100 standards initiatives. Chapter 6 likewise explains that in the bioeconomy, standards for product and process interoperability directly affect issues of IP.

Countries and firms that play primary roles in setting international standards can enjoy advantages if new standards align with their own national standards and/or features of their productive base. The public sector's role should be to encourage industry, including firms of different sizes, to participate at early stages in international (and in some cases national) standards setting. Dedicated support could aim to include under-represented groups of firms in standards development processes. Relevant public agencies should also pursue standards development in the research system.

Optimising digital systems to strengthen science and innovation policies

Chapter 7 examines digital science and innovation policy (DSIP) systems. DSIP systems use digital procedures and infrastructures to help formulate and deliver science and innovation policy. They are used to monitor policy interventions, develop new STI indicators, assess funding gaps, strengthen technology foresight, and identify leading experts and organisations. Data are mainly sourced from funding agencies (e.g. databases of grant awards), R&D-performing organisations, proprietary bibliometric and patent databases, and the web.

There are various types of DSIP systems. Databases of public funders are one type, of which Belgium's Flanders Research Information Space (FRIS) is an example. The FRIS portal, launched in 2011, aims to accelerate innovation, support science and innovation policy making, share information on publicly funded research with citizens, and reduce the administrative burden of research reporting.

A second type of DSIP infrastructure is a Current Research Information System. Through the Estonian Research Information System (ETIS), for example, Estonian higher education institutions (HEIs) manage research information and showcase research. Public funders use ETIS to evaluate and process grant applications. National research assessments and evaluations also draw on ETIS.

A third type of DSIP infrastructure is what might be termed an "intelligent system". For example, to examine the socio-economic impacts of research, Japan's SciREX Policymaking Intelligent Assistance System (SPIAS) uses big data and semantic technologies (which aim to extract meaning from data). They help to process data on Japan's research outputs and impacts, funding, R&D-performing organisations and research projects.

Chapter 7 discusses three main challenges facing DSIP systems: ensuring the interoperability of diverse data sets; preventing misuses of DSIP systems in research assessments; and managing the roles of non-government actors, particularly the private sector, in developing and operating parts of DSIP systems. The following subsections briefly describe these three themes.

Ensuring interoperability in DSIP systems

DSIP systems pull data from multiple sources, linking them to gain policy insights that are otherwise impossible to achieve. But linking data is highly problematic, chiefly on account of different data standards. Recent years have seen attempts to establish international standards and vocabularies to improve data sharing and interoperability in science and research management. These include unique, persistent and pervasive identifiers, which assign a standardised code unique to each research entity, persistent over time and pervasive across datasets. Many DSIP infrastructures have adopted such standards to link data from universities, funding bodies and publication databases, thereby relating research inputs to research outputs.

Using DSIP systems in research assessment

Many metrics aim to quantify scientific quality, impact and prestige. More than half of the DSIP systems identified in OECD work play a role in research assessment. The growing digital footprint of academic and research activities suggests that, in future, most relevant dimensions of research activity might be represented digitally. In this connection, the altmetrics movement promotes metrics generated from social media as a type of evidence of research impact that is broader and timelier than academic citations. However, as with traditional metrics, questions remain over the extent to which altmetrics afford valid signals of research impact.

The roles of the business sector in DSIP

Non-government actors are emerging as a main force in DSIP systems. The large academic publishers, Elsevier and Holtzbrinck Publishing Group, together with the analytics firm, Clarivate Analytics, are particularly active in developing products and services into platforms that mimic fully fledged DSIP systems. Multinational corporations like Alphabet Inc. and Microsoft Inc., and national technology companies such as Baidu Inc. (China) and Naver Inc. (Korea), have also designed platforms to search academic outputs. In the future, these platforms could become key elements in national DSIP systems.

Harnessing these private sector developments in public DSIP systems has many potential benefits. Solutions can be implemented quickly and at an agreed cost, sparing the public sector the need to develop in-house skills beforehand. Private companies can promote interoperability through their standards and products, which can expand the scope and scale of data used in a DSIP system. However, outsourcing data management activities to the private sector may bring risks. These could include loss of control over the future development of DSIP systems, discriminatory access to data and even the emergence of private platforms that become dominant because of hard-to-contest network effects.

The outlook for DSIP systems

Governments need to shape DSIP ecosystems to fit their needs. This will require interagency co-ordination, sharing of resources (such as standard digital identifiers) and coherent policy frameworks for data sharing and reuse in the public sector. Since several government ministries and agencies formulate science and innovation policy, DSIP systems should involve co-design, co-creation and co-governance. In a desirable future, DSIP infrastructures will provide multiple actors in STI with up-to-date linked microdata. Policy frameworks will have resolved privacy and security concerns, and national and international co-operation on metadata standards will have addressed interoperability issues.

Digitalisation in science and innovation: Possible “dark sides”

The thrust of this report is that digitalisation offers many positive opportunities for STI, so long as complementary policies receive proper attention. This subsection considers the possibility of unwelcome outcomes from digitalisation in STI. These include widening capability gaps across countries and subnational regions, negative effects on science processes, excessive complexity in machine ecosystems, and risks that are diffuse, hard to foresee and primarily social. Evidence on the likelihood or scale of these undesirable outcomes is scant. A conclusion from this subsection, therefore, is the need for greater awareness and further study. Public concerns about automation, jobs and inequality, where the literature is vast, are not discussed.

Distributional effects and digitalisation of STI

Aspects of digitalisation could widen gaps in STI capability and income across countries and regions. Three possibilities are considered here:

Centralisation effects in science. Science increasingly occurs within data (Hey, Tansley and Tolle, 2009). Developed countries have a comparative advantage in capital-intensive scientific tools that generate data. It is an open question whether these conditions might affect the broad geography of scientific activity. In one scenario, with suitable data access, developing-country researchers might be able to do science without making the sorts of capital investments made by developed countries. In another, researchers in developed countries might strengthen their existing advantages in leading-edge science. As a narrower but possibly related issue, laboratory automation is now essential to many areas of science and technology, but is expensive and difficult to use. Consequently, laboratory automation is most economical in large central sites, and companies and universities are increasingly concentrating their laboratory automation. The most advanced example of this trend is cloud automation in biological science. Biological samples are sent to a single site and scientists design their experiments using application programming interfaces (King and Roberts, 2018). The effect of such cloud-based possibilities on the overall dispersion or concentration of scientific work is unclear.

Effects on subnational geographies. The digital economy may exacerbate geographic disparities in income, as it amplifies the economic and social effects of initial skills endowments (Moretti, 2012). In many OECD countries, income convergence across subnational regions has either halted, or reversed, in recent decades (Ganong and Shoag, 2015). Among remedial policies, investments in skills and technology are most important (because investments in infrastructure and transport, while often beneficial, also have diminishing returns (Filippetti and Peyrache, 2013).

Effects from supercomputing. Today, some supercomputers are designed specifically for AI. Previously, supercomputers were used mostly for modelling, such as in climate and nuclear science. Many tech companies are orienting towards supercomputing (Knight, 2017). Worldwide, however, only 27 countries possess a supercomputer listed among the top 500 most powerful. China, notably, has made major strides in building supercomputers with domestically produced components. China also boasts large numbers of supercomputers, along with abundant data to train AI algorithms. Might capabilities across countries diverge because of increasing synergy between supercomputing and AI? Will the value of owning/building increasingly powerful supercomputers change relative to using cloud-based computing?

Complex systems and unmanageable machine ecologies

Governments need improved understanding of complex systems (Nesse, 2014). As a wide array of critical systems becomes more complex, mediated and interlinked by code, the risk and consequences of vulnerabilities could increase. As code controls a growing number of connected systems, errors can cascade, with effects that become more extensive than in the past. For instance, owing to software faults, the United States recently experienced the first national – rather than local – 911 outages (Somers, 2017). Critical ICT systems might behave in unpredictable and even emergent ways, and the ability to anticipate failures in technology could diminish (Arbesman, 2016). A widely publicised case was the unexpected interaction of algorithms that contributed to the "Flash Crash" of May 2010, when more than 1 trillion dollars in value was lost from global stock markets in minutes. However, many more examples exist of software errors that caused system failures. In 1996, for instance, the European Space Agency's Arianne 5 rocket exploded on launch owing to a software glitch.[4]

AI and other measures will help to automate and improve software verification. Nevertheless, as the physicist Max Tegmark observes "the very task of verification will get more difficult as software moves into robots and new environments, and as traditional pre-programmed software gets replaced by AI systems that keep learning, thereby changing their behaviour…" (Tegmark, 2017).

An inbuilt feature of technology is that it deepens complexity: systems accumulate parts over time, and more connections develop between those parts. Technologies that become more complex can end up depending on antiquated legacy systems. This is especially so for code. For example, in the lead up to 1 January 2000, amid Y2K concerns, the US Federal Aviation Administration examined computers used for air traffic control. One type of machine required fixing, an IBM 3083 that had been installed in the 1980s. However, only two persons at IBM knew the machine's software, and both had retired (Arbesman, 2016).

Negative impacts on science from digitalisation

This chapter has already described a number of challenges that digitalisation raises for science – from coping with predatory online science journals to keeping personal research data anonymous. Chapter 2 – on measurement – reports that a sizeable number of scientists think digitalisation will have at least some negative impacts on science. These potential impacts include the growth of hypothesis-free research in data-driven science, and divides in research between those who possess advanced digital competences and those who don't. Digitalisation could also encourage a celebrity culture in science, lead to premature diffusion of research findings and expose individuals to pressure groups. Other concerns are the use of readily available but inappropriate indicators for monitoring and incentivising research, and the potential concentration of workflows and data in the hands of a few companies providing digital tools.

Another potentially problematic issue is the misapplication of AI in science and society. The design and use of effective AI systems requires expertise which is scarce. Moreover, stricter requirements on performance, robustness, predictability and safety will increase the need for expertise. This is especially true for deep learning techniques that are now central to AI research and applications.

With expertise bottlenecks and, sometimes, unrealistic expectations about what AI can achieve, non-experts are increasingly deploying AI. Such systems often suffer from deficiencies in performance, robustness, predictability and safety, outcomes that even AI experts can struggle to achieve (Hoos, 2018). Hoos and others propose building a next generation of AI systems known as Automated AI as one way to alleviate the AI complexity problem. This could help develop and deploy accurate and reliable AI without the need for deep and highly specialised AI expertise. Automated AI builds on work on automated algorithm design, and automated ML, which is developing rapidly in academia and industry (Hoos, 2012).

Wider risks linked to digital technology

Like all technology, digital technologies can help and harm. AI, for instance, can increase digital security by predicting where threats originate, but it can decrease digital security by adding intelligence to malware. Synthetic biology can help cure disease, but it can also make pathogens more virulent. Some risks of digital technology reflect complex interactions with social systems and as such may be impossible to foresee.

Today, one risk is the fragmentation of public discourse by social media. The future might also see a loss of trust in accredited information owing to high-fidelity audio and video fakes. In addition, the diminished economic viability of journalism and literary writing, a development attributed to digital technology, could have unwanted social and political effects (de León, 2019).

Harari (2018) even suggests the future of computing could shape the future of democracy. Autocracy, he notes, has generally failed in advanced economies, partly because information processing could not be centralised sufficiently. Decentralised information processing gives democracies an efficiency advantage. However, if AI comes to encompass ever more of the digital economy, it may have a centralising tendency. AI will also become more effective as data are concentrated. Harari (2018) suggests that finding ways to keep distributed data processing more efficient than centralised data processing could ultimately help safeguard democracy.

Policy makers can take additional steps to mitigate emerging risks brought on by the dual-use nature of technology. Past episodes in the history of science might provide useful lessons. The case of Paul Berg, the Nobel laureate who helped create recombinant DNA, is one example. Aware of the ramifications of his discovery, Berg convened the Asilomar Conference. This led to a moratorium on the most dangerous experiments until the science improved.

Policy makers can mitigate technological risk in several ways. They can earmark part of research budgets to study the broader implications of science. Engaging the public in debate, while avoiding hyperbole about technology, is useful. In addition, they can ensure that science advice is trustworthy. Investments in research and innovations that reduce risk (such as in cyber-security) might also help.

The untapped potential of digital technology for STI policy

This section explores new ideas for how digital technology might support policy for science and innovation. Earlier, Box 1.1 described new thinking on collective intelligence and the allocation of public research funds. Other examples considered here are prediction markets, various applications of blockchain, and using social media to increase exposure to innovation in a selective way. Some of these ideas have yet to receive significant attention, and few governments have experimented with the opportunities available.

Prediction markets for STI policy

Prediction markets, which involve trading bets on whether some specific outcome will occur, could inform STI policy. Prediction markets have outperformed the judgement of experts in forecasting outcomes in fields as diverse as sporting tournaments and political elections. They aggregate decentralised private information, which is captured in the changing price of the next bet on the outcome in question (in a similar way to a futures market). Prediction markets incentivise participants to find or generate new information (from which profit could derive). Recent experiments (see Dreber et al. [2019], Munafo et al. [2015], Dreber et al. [2015] and Almenberg, Kittlitz and Pfeiffer [2009]) show that prediction markets might accomplish the following:

- Predict the results of otherwise expensive research evaluations (e.g. of HEIs).
- Quickly and inexpensively identify research findings that are unlikely to replicate.
- Help optimally allocate limited resources for replications.
- Help institutions assess whether strategic actions to improve research quality are achieving their goals.
- Test scientific hypotheses.
- Help understand specific scientific processes. For instance, a research project could be examined alongside a history of the project's market prices, to show when hypotheses had strengthened or weakened (Dreber et al., 2015).

Specialised digital platforms make it easier to implement prediction markets. On the Augur platform, for example, with an initial commitment of less than a dollar, anyone can ask a question and create a market based on a predicted outcome. Using prediction markets in STI appears more constrained by tradition than by technical infeasibility.

Prediction using human-machine combinations

Human intelligence (of individuals or crowds) and machine intelligence could be combined for prediction and research. For instance, researchers at Stanford University and Unanimous AI, a California-based company, connected small groups of radiologists over the Internet using AI, and tested their ability to diagnose chest X-rays. Radiologists and algorithms together were more accurate than the unaided group. They were even more accurate than individual radiologists, and 22% more accurate than state-of-the-art AI alone (Rosenberg et al., 2018).

Accurate foresight is particularly elusive when technological change is radical. One complication is that it often takes considerable time before the main applications of radical innovations emerge. After Gutenberg, for example, it took nearly a century of technical and conceptual improvements to arrive at the modern book (Somers, 2018).

Indeed, even the most knowledgeable experts frequently misjudge technological timelines. In the digital sphere, one example of such misjudgement is the 1955 proposal for the Dartmouth Summer Research Project on Artificial Intelligence, a seminal event in the history of AI. The proposal stated that a significant advance in AI could be made "…if a carefully selected group of scientists work on it for a summer". Whether using prediction markets, a human-machine approach or other methods, harnessing collective intelligence might strengthen policy foresight.

Blockchain for science, technology and innovation

One leading commentator has described blockchain as follows: "blockchain technology facilitates peer-to-peer transactions without any intermediary such as a bank or governing body…the blockchain validates and keeps a permanent public record of all transactions. This means that personal information is private and secure, while all activity is transparent and incorruptible – reconciled by mass collaboration and stored in code on a digital ledger" (Tapscott, 2015). As Chapter 5 discusses, while blockchain applications in production are still incipient, companies such as Microsoft, IBM and others now offer commercial blockchain services. Proposals to use blockchain in STI are flourishing (Box 1.5).

Box 1.5. Possible applications of blockchain in science and innovation

Recent proposals for how blockchain might benefit STI include the following:

Establishing a cryptocurrency for science. Using a cryptocurrency, publishers of scientific works could receive micro-payments as content is consumed. A science cryptocurrency could also facilitate a system of rewards for sometimes under-rewarded activities such as statistical support, exchange of lab equipment, data hosting and curation, and peer review (van Rossum, 2018). Science Matters – an OA publishing platform – will soon implement a crowdsourced peer-review process using the Ethereum blockchain. Ideally, researchers and publishers will quickly see metrics that can help expedite publication. Furthermore, for their time, reviewers will also receive cryptocurrency linked to the platform (Heaven, 2019).

Storing and sharing research data. Databases that encompass large parts of the research ecosystem are technically possible. However, the need for centralised management and ownership complicates their implementation. Data security and ease of access are just some of the concerns. In principle, the blockchain could make scalable, safe and decentralised data stores more practical. It could also enhance the reproducibility of science by automatically tracking and recording work such as statistical analysis, while reducing the risk of data fraud. In addition, metrics could be developed for activities that are not well recognised, such as data development, because they could be clearly attributed (van Rossum, 2018).

Enabling data use. Data sharing can be difficult for several reasons, including institutional and technical issues, as well as regulations. Institutional obstacles include bureaucratic processes that hinder permission to share data. Even when a DSA is reached, data holders still worry about inappropriate use of their data, or about accidental sharing of client data. Furthermore, on a technical level, some datasets are just too big to share easily. For instance, 100 human genomes could consume 30 000 000 MB. Uncertainty about the provenance of data can also hinder data sharing or purchase. In addition, regulators might increasingly require that AI systems demonstrate auditable data use. In this environment, efforts are underway to link blockchain and AI in a system that gives data holders the benefits of data collaboration, but with full control and verifiable audit. Ocean Protocol, an open-source not-for-profit foundation, is pioneering such a system. Under one use case, data are neither shared nor copied. Instead, algorithms go to the data for training purposes, with all work on the data recorded in a distributed ledger (Chhabra, 2018).

Making ownership of creative material transparent. Commercial services now offer secure attribution of ownership of creative works by providing a blockchain-verified cryptographic ID (Stankovic, 2018). Launched in 2018, Artifacts is a platform for publishing any material that researchers consider worth sharing. This ranges from data sets to single observations, hypotheses and negative research results, all logged to the blockchain. Artifacts aims to disseminate more scientific information, securely and in citable ways, more quickly than occurs with peer-reviewed written articles (Heaven, 2019).

Broadening access to supercomputing. Golem aims to create a global supercomputer, accessible to anyone, using processing power from idle computers and data centres around the world. Users would rent processing time from each other, and rely on blockchain to track computations and payments, and to keep data secure (Golem, n.d.).

Technical and policy challenges such as interoperability must be resolved before blockchain in STI can be widely deployed. Without consensus on the protocols in blockchain and other DLTs, use will be limited. One effort towards consensus, IBM's Hyperledger, seeks an interoperable architecture for DLTs. Technical limits also exist on the volume of transactions that blockchain networks can process. However, the scalability challenge is less severe for so-called permissioned blockchain applications – where participation in the network is controlled. Permissioned blockchain networks are the most likely in STI, because they will generally be used to help a particular professional community achieve some policy-relevant outcome. Mechanisms to ensure the veracity of information in a blockchain registry are lacking (although efforts are underway to establish the veracity of the identity of those feeding information into the blockchain).[5] Agreement is also lacking on how to terminate a so-called smart contract – a contract that executes itself, enabled by blockchain – and how to treat smart contracts that contain errors or illegal instructions. The tamper-proof design of the blockchain could also be problematic if the system prevents corrections, even when necessary (Stankovic, 2018).

Using social media to spread innovation

People's propensity to innovate involves an element of imitation. Research shows that children who grow up in areas with more inventors are more likely to become inventors. Greater exposure to innovation among minorities and children from low-income families might increase the prevalence of innovation. Among other measures, social media could provide a channel for targeted interventions (Bell et al., 2019).

Conclusion

Scientific progress cannot be taken for granted. There are many areas of science – fundamental to human well-being – where knowledge is still surprisingly limited. For example, the process by which *E. coli* (a bacterium) consumes sugar for energy is one of the most basic biological functions, and also important for industry. But how the process operates has not been fully established, even though research on the subject was first published over 70 years ago. Uncertainty also exists on many critical questions in climate science. To name a few, what is the tipping point for the inversion of the flows of cold and hot oceanic waters? When could changes become irreversible (e.g. melting of West Antarctic or Greenland ice-shelves)? What is the quantitative role of plants and microbes in the carbon cycle?

Progress in STI is also necessary because, despite striking advances in technology, the pace of innovation is insufficient in some crucial fields. For instance, today's leading energy generation technologies were mostly developed or demonstrated over a century ago. The combustion turbine was invented in 1791, the fuel cell in 1842, the hydro-electric turbine in 1878 and the solar photo-voltaic cell in 1883. Even the first nuclear power plant began operating over 60 years ago. The performance of all these technologies has, of course, improved. But truly disruptive breakthroughs have not occurred (Webber et al., 2013). Indeed, some high-profile commentators from academia and industry have gone further, claiming (debatably) that a more general innovation plateau has been reached.

Furthermore, efficient and effective policies for STI are ever more important in countries where rapid population ageing is likely to constrain discretionary public spending over the long run.

For these and other reasons examined in this publication, utilising the full potential of digital technology in STI is important.

References

Agrafioti, F. (2018), "How to set up an AI R&D lab", *Harvard Business Review,* 20 November, https://hbr.org/2018/11/how-to-set-up-an-ai-rd-lab.

Almenberg, J., K. Kittlitz and T. Pfeiffer (2009), "An experiment on prediction markets in science", PLOS One, 30 December, PLOS, San Francisco, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008500.

Arbesman, S. (2016), *Overcomplicated: Technology at the Limits of Comprehension*, Penguin Random House, New York.

Atkinson, R.D. and S. Ezell (2019), "The Manufacturing Evolution: How AI will Transform Manufacturing and the Workforce of the Future", Information Technology and Innovation Foundation, Washington DC, https://itif.org/publications/2019/08/06/manufacturing-evolution-how-ai-will-transform-manufacturing-and-workforce.

Bell, A.M. et al. (2019), "Who becomes an inventor in America? The importance of exposure to innovation", *NBER Working Paper* No. 24062, issued in November 2017, revised in January 2019, National Bureau of Economic Research, Cambridge, Massachusetts, www.nber.org/papers/w24062.

Bergstein, B. (2019), "This is why AI has yet to reshape most businesses", *MIT Technology Review*, 13 February, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/612897/this-is-why-ai-has-yet-to-reshape-most-businesses/.

Biles, S. (28 December 2018), "Tiny computers could transform our lives", Scientific American Observations blog, https://blogs.scientificamerican.com/observations/tiny-computers-could-transform-our-lives/.

Bloom, N. et al. (2017), "Are ideas getting harder to find?", *NBER Working Paper* No. 23782, September, National Bureau of Economic Research, Cambridge, Massachusetts, https://www.nber.org/papers/w23782.

Bollen, J. (2018), "Who would you share your funding with?", *Nature*, Vol. 560, Nature Research, Springer,p.143, www.nature.com/articles/d41586-018-05887-3.

Bollen, J. et al. (2014), "From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review", *EMBO Reports*, Vol. 15/1, Wiley Online Library, pp. 131-133, https://onlinelibrary.wiley.com/doi/full/10.1002/embr.201338068.

Brooks, R. (15 July 2018), "Steps towards superintelligence IV: Things to work on now", Robots, AI and other stuff blog, https://rodneybrooks.com/forai-steps-toward-super-intelligence-iv-things-to-work-on-now/.

Butler, K.T. et al. (2018), "Machine learning for molecular and materials science", *Nature*, Vol. 559/7715, Nature Research, Springer, pp. 547-555, https://doi.org/10.1038/s41586-018-0337-2.

Castelvecchi, D. (2019), "Forget everything you know about 3-D printing – the 'replicator' is here", *Nature,* Vol. 566, Nature Research, Springer, p. 17, http://doi.org/10.1038/d41586-018-07798-9.

Chhabra, C.S (2018), "New data and the AI economy", presentation at the workshop on digital technology for science and innovation, Oslo, 5-6 November, www.oecd.org/sti/inno/digital-technology-for-science-and-innovation-emerging-topics.htm.

Corea, F. (2017), "The convergence of AI and blockchain: What's the deal?", *Medium,* 1 December, https://medium.com/@Francesco_AI/the-convergence-of-ai-and-blockchain-whats-the-deal-60c618e3accc.

Craig, E. (2018), "An overview of the possibilities and challenges in delivering STEM education using AR and VR", presentation at the workshop on digital technology for science and innovation, Oslo, 5-6 November, www.oecd.org/sti/inno/digital-technology-for-science-and-innovation-emerging-topics.htm.

Damer, B. (2018), "The origin of life and the engine of emergence", presentation at the workshop on digital technology for science and innovation, Oslo, 5-6 November, oe.cd/2vU.

de León, C. (2019), "Does it pay to be a writer?", *The New York Times,* 5 January, www.nytimes.com/2019/01/05/books/authors-pay-writer.html.

Dohler, M. (2017), "Global reach: Will the tactile Internet globalize your skill set?", *ComSoc Technology News*, 23 January, www.comsoc.org/publications/ctn/global-reach-will-tactile-internet-globalize-your-skill-set.

Dreber, A. et al. (2019), "Using prediction markets to predict the outcomes in DARPA's next generation social science program", 11 February, https://osf.io/k6xcq.

Dreber, A. et al. (2015), "Using prediction markets to estimate the reproducibility of scientific research", *Proceedings of the National Academy of Sciences,* 15 December, Vol. 112/50, United States National Academy of Sciences, Washington, DC, pp. 15343-15347, https://doi.org/10.1073/pnas.1516179112.

Dyson, F.J. (1998), *Imagined Worlds*, Harvard University Press, Cambridge, Massachusetts.

Evans, B. (2019), "5G: If you build it, we will fit it", benevans.com, 16 January, www.ben-evans.com/benedictevans/2019/1/16/5g-if-you-build-it-we-will-fill-it.

Filippetti, A. and A. Peyrache (2013), "Labour productivity and technology gap in European regions: A conditional frontier approach", *Regional Studies,* Vol. 49/4, 14 June, Regional Studies Association, Brighton, pp. 532-554, http://dx.doi.org/10.1080/00343404.2013.799768.

Fischer, E. (2018), "How far we've come: 70 years of weather technology advances", CBS Boston, 22 May, https://boston.cbslocal.com/2018/05/22/wbz-weather-eric-fisher-severe-weather-alerts-ibm-radar-advancements/.

Freedman, L.P., I.M. Cockburn and T.S. Simcoe (2015), "The economics of reproducibility in preclinical research", *PLOS Biology*, 9 June, PLOS, San Francisco, https://doi.org/10.1371/journal.pbio.1002165.

Fussell, S.R. et al. (2008), "How people anthropomorphize robots", presentation at the international conference on human robot interaction, 12-15 March, Amsterdam, https://doi10.1145/1349822.1349842.

Ganong, P. and D. Shoag (2015), "Why has regional income convergence in the US declined?", http://scholar.harvard.edu/files/shoag/files/ganongshoagjan2015.pdf.

Golem (n.d.), Golem website, https://golem.network/ (accessed 20 June 2019).

Graves, N., A.G. Barnett and P. Clarke (2011), "Funding grant proposals for scientific research: Retrospective analysis of scores by members of grant review panel", *The BMJ* 2011; 343:d4797, London, https://www.bmj.com/content/343/bmj.d4797.

Harari, Y. (2018), "Why technology favors tyranny", *The Atlantic*, October, www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/.

Harbert, T. (2013), "Supercharging patent lawyers with AI: How Silicon Valley's Lex Machina is blending AI and data analytics to radically alter patent litigation", *IEEE Spectrum*, 30 October, Institute of Electrical and Electronics Engineers, New York, https://spectrum.ieee.org/geek-life/profiles/supercharging-patent-lawyers-with-ai.

Heaven, D. (2019), "Bitcoin for the biological literature", Nature, Vol. 566, Nature Research, Springer, pp. 141-142, www.nature.com/articles/d41586-019-00447-9.

Herbert, D.L, A.G. Barnett and N. Graves (2013), "Australia's grant system wastes time", *Nature*, Vol. 495, 21 March, Nature Research, Springer, pp. 314, www.nature.com/articles/495314d.

Hernandez, D. (2013), "Homeless, Unemployed and Surviving on Bitcoins", *WIRED*, 20 September, https://www.wired.com/2013/09/bitcoin-homeless/.

Hey, T., S. Tansley and K. Tolle (2009), "The fourth paradigm: Data-intensive scientific discovery", Microsoft Research, Redmond, United States.

Hicks, M. (2017), *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing,* MIT Press, Cambridge, United States.

Hoos, H.H. (2018), "Democratisation of AI by automating the creation of AI systems", presentation at digital technology for science and innovation workshop, Oslo, 5-6 November.

Hoos, H.H. (2012), "Programming by optimization", *Communications of the ACM,* Vol. 55/2, Association for Computing Machinery, New York, pp. 70-80.

Hutson, M. (2017), "The future of AI depends on a huge workforce of human teachers", *Bloomberg Businessweek*, 7 September, www.bloomberg.com/news/articles/2017-09-07/the-future-of-ai-depends-on-a-huge-workforce-of-human-teachers.

Jordan, M. (2018), "Artificial intelligence – The revolution hasn't happened yet", *Medium,* 19 April, https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7.

Kelly, K. (2013), "The post-productive economy", The Technium, https://kk.org/thetechnium/the-post-produc/.

Kharif, O. (2017), "The bitcoin whales: 1,000 people who own 40 percent of the market", *Bloomberg Businessweek*, 8 December, www.bloomberg.com/news/articles/2017-12-08/the-bitcoin-whales-1-000-people-who-own-40-percent-of-the-market.

King, R.D. and S. Roberts (2018), "Artificial intelligence and machine learning in science", in *OECD Science, Technology and Innovation Outlook 2018*, OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-10-en.

Knight, W. (2018), "A robot scientist will dream up new materials to advance computing and fight pollution", *MIT Technology Review*, 7 November, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/612388/a-robot-scientist-will-dream-up-new-materials-to-advance-computing-and-fight-pollution/.

Knight, W. (2017), "Google reveals a powerful new AI chip and computer," *MIT Technology Review*, 17 May, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/607879/google-reveals-a-powerful-new-ai-chip-and-supercomputer/.

Mantha, Y. and S. Hudson (2018), "Estimating the gender ratio of AI researchers around the world", *Medium,* 17 August, https://medium.com/element-ai-research-lab/estimating-the-gender-ratio-of-ai-researchers-around-the-world-81d2b8dbe9c3.

Metz, C. (2019), "'Businesses will not be able to hide': Spy satellites will give edge from above", 24 January, *The New York Times*, https://www.nytimes.com/2019/01/24/technology/satellites-artificial-intelligence.html.

Moretti, E. (2012), *The New Geography of Jobs*, Houghton, Mifflin, Harcourt Publishing, New York.

Munafo, M.R. et al. (2015), "Using prediction markets to forecast research evaluations", Royal Society Open Science, Vol 2/10, Royal Society, London, http://dx.doi.org/10.1098/rsos.150287.

Musib, M. et al. (2017), "Artificial intelligence in research", *Science,* Vol. 357/6346, 7 July, American Association for the Advancement of Science, Washington, DC, pp. 28-30, https://doi.org/10.1126/science.357.6346.28.

Nesse, R. (2014), "The Fragility of Complex Systems", in Brockman, J. (ed.), *What Should We Be Worried About?*, Harper Perrenial, New York, London, Toronto, Sydney, New Delhi, Auckland.

OECD (2019), *Going Digital: Shaping Policies, Improving Lives,* OECD Publishing, Paris, https://doi.org/10.1787/9789264312012-en.

OECD (2018a), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption,* OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-en.

OECD (2018b), *OECD Reviews of Innovation Policy: Austria 2018*, OECD Reviews of Innovation Policy, OECD Publishing, Paris, https://doi.org/10.1787/9789264309470-en.

OECD (2017), *OECD Science, Technology and Industry Scoreboard 2017: The Digital Transformation*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264268821-en.

OECD (2015), "Daejeon Declaration on Science, Technology, and Innovation Policies for the Global and Digital Age", webpage, www.oecd.org/sti/daejeon-declaration-2015.htm (accessed 20 June 2019).

Pratt, G.A. (2015), "Is a Cambrian explosion coming for robotics?", *Journal of Economic Perspectives*, Vol. 29/3, Summer, American Economic Association, Pittsburgh, pp. 51-60.

Princeton University (2018), "Researchers crowdsource brain mapping with gamers, discover six new neuron types", *Medical Express*, 17 May, https://medicalxpress.com/news/2018-05-crowdsource-brain-gamers-neuron.html.

Ransbotham, S. (21 May 2018), "Using AI to create humanlike computers is a shortsighted goal", MIT Sloan Management Review blog, https://sloanreview.mit.edu/article/rethink-ai-objectives/.

Reich, J. and J.A. Ruipérez-Valiente (2019), "The MOOC Pivot: What happened to disruptive transformation of education?", *Science,* 11 January, Vol. 363/6423, American Association for the Advancement of Science, Washington, DC, pp. 130-131, https://science.sciencemag.org/content/363/6423/130.summary.

Rosenberg, L, et al. (2018), "Artificial swarm intelligence employed to amplify diagnostic accuracy in radiology," presentation at the annual information technology, electronics and mobile communication conference, Vancouver, 2-4 November, www.Artificial_Swarm_Intelligence_employed_to_Amplify_Diagnostic_Accuracy_in_Radiology.

Sankaran, V. (2018), "Meet the people busting scientists who fake images in research papers", The Next Web, 6 November, https://thenextweb.com/science/2018/11/06/meet-the-people-busting-scientists-who-fake-images-in-research-papers/.

Scholz, T. and N. Schneider (2019), *Ours to Hack and to Own: The Rise of Platform Cooperativism, A New Vision for the Future of Work and a Fairer Internet*, OR Books, New York and London.

Shapira, P. and J. Youtie (2017), "The next production revolution and institutions for technology diffusion", in *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-11-en.

Somers, J. (2018), "The scientific paper is obsolete: Here's what's next", *The Atlantic*, 5 April, www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/.

Somers, J. (2017), "The coming software apocalypse", *The Atlantic*, 26 September, www.theatlantic.com/technology/archive/2017/09/saving-the-world-from-code/540393/.

Stankovic, M. (2018), "Using blockchain to facilitate innovation in the creative economy", presentation at the digital technology for science and innovation workshop, Oslo, 5-6 November.

Stoet, G. (2016), "Maths anxiety is creating a shortage of young scientists … here's a solution", The Conversation, 1 June, http://theconversation.com/maths-anxiety-is-creating-a-shortage-of-young-scientists-heres-a-solution-58889.

Susskind, J. (2018), *Future Politics,* Oxford University Press, United Kingdom.

Tapscott, D. (2015), "Blockchain revolution: How the technology behind Bitcoin is changing money, business and the world", Don Tapscott, http://dontapscott.com/2015/06/blockchain-revolution-the-brilliant-technology-changing-money-business-and-the-world/ (accessed 20 June 2019).

Tegmark, M. (2017), Life 3.0: Being Human in the Age of Artificial Intelligence, First Vintage Books, New York.

Valiant, L. (2013), *Probably Approximately Correct*, Basic Books, New York.

van Rossum, J. (2018), "The blockchain and its potential for science and academic publishing", presentation at the digital technology for science and innovation workshop, Oslo, 5-6 November.

Webber, M.E., R.D. Duncan and M.S. Gonzalez (2013), "Four technologies and a conundrum: The glacial pace of energy innovation", *Issues in Science and Technology,* Vol. xxix/2, Winter, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, University of Texas at Dallas, www.issues.org/29.2/Webber.html.

Winick, E. (2018), "A cyber-skills shortage means students are being recruited to fight off hackers", *MIT Technology Review*, 18 October, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/612309/a-cyber-skills-shortage-means-students-are-being-recruited-to-fight-off-hackers/?source=download-metered-content.

Worstall, T. (2016), "Are ideas getting harder to find? Not really, no, unless you measure by ideas already found", *Forbes*, 20 December, www.forbes.com/sites/timworstall/2016/12/20/are-ideas-getting-harder-to-find-not-really-no-not-if-you-measure-by-ideas-already-found/#66eaf6334840.

Youn, H. et al. (2015), "Invention as a combinatorial process: Evidence from US patents", *Journal of the Royal Society Interface*, 6 May, Royal Society, London, https://doi.org/10.1098/rsif.2015.0272.

Zubașcu, F. (2017), "'Digital revolution' will underpin next EU research programme, says Commissioner", Sciencebusiness.net, 11 December, https://sciencebusiness.net/framework-programmes/news/digital-revolution-will-underpin-next-eu-research-programme-says (accessed 20 June 2019).

Notes

[1] Last year, more than 1.2 million new papers were published in the biomedical sciences alone, bringing the total number of peer-reviewed biomedical papers to over 26 million. However, the average scientist reads only about 250 papers a year, www.nature.com/news/scientists-may-be-reaching-a-peak-in-reading-habits-1.14658.

[2] In this connection, one recent study argued that gender stereotyping was instrumental in the United Kingdom losing its globally pre-eminent position in computing after World War II (Hicks, 2017).

[3] Professor Li's full remarks at the 2017 Global StartupGrind Conference can be found here: www.startupgrind.com/blog/cloud-will-democratize-ai.

[4] Tegmark (2017) provides many similar examples.

[5] See, for example, Authenteq (https://authenteq.com/), which uses DLTs to provide digital identity verification.

2 How are science, technology and innovation going digital? The statistical evidence

Fernando Galindo-Rueda

Directorate for Science, Technology and Innovation, OECD

Chapter 2 examines the digitalisation of science and innovation drawing on statistical measurement and analysis by the OECD's Working Party of National Experts on Science and Technology Indicators, including material featured in the OECD report *Measuring the Digital Transformation*. This chapter maps the ICT specialisation of research and the growth of scientific production and government funding of research related to artificial intelligence. It examines the multidimensional nature of the digital transformation of science. This chapter also shows how innovation in firms can be linked to the adoption of digital technologies and business practices. It concludes by summarising possible next steps for OECD's own measurement agenda.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

Hardly a day goes by without the traditional or social media highlighting how digitally driven scientific or technological breakthroughs might transform daily life. Much computer speech and image recognition has attained human-like levels of performance, while self-driving cars are gradually improving their safety record. Media attention to such breakthroughs is provoking a deeper reflection among policy makers concerned with science, technology and innovation (STI). How is the nature of science and innovation itself changing? How, if at all, should this change be managed?

The public's exposure to the accumulation of anecdotal evidence on the digital transformation of science and innovation builds views of change shaped by how close they are to specific developments. But how widespread are those specific developments? Which practices have fully broken into the mainstream? Which practices remain the preserve of relatively small communities at the leading edge? Do different facets of digitalisation complement or offset each other? Is debate excessively focused on practices that are no longer at the forefront, and are incipient signals about the direction of change being missed?

Addressing these questions requires a comprehensive view of how science and innovation are "going digital". The digital revolution is based on the growing possibilities to create and use data, information and knowledge, and ultimately to support decision making, science and innovation policy. As such, it requires data and measurement that help map the ongoing transformations, their causes and their effects.

This chapter reports on some key features and trends in the digitalisation of science and innovation. To that end, it draws principally on statistical measurement and analysis under the aegis of the OECD's Working Party of National Experts on Science and Technology Indicators (NESTI), including contributions featured in the report titled *Measuring the Digital Transformation: A Roadmap for the Future* (OECD, 2019a), a publication that provides a broad statistical and measurement-oriented view of digitalisation and accompanies the OECD report *Going Digital: Shaping Policies, Improving Lives.* Both in and outside the OECD, work to measure digitalisation is also a basis for collective choices about the data that policy makers wish to have and act upon (see Chapter 7). This chapter provides a number of reflections on measurement gaps and what can be done, and is being done, to address them.

Given the breadth of digitalisation's influence, and the available evidence, some perspective is needed. Historically, the development of science and technology have been intertwined. Innovation in measurement tools provided a means to improve scientific understanding of nature, and this knowledge also turned out to be essential for innovation. Each wave of widespread technological development has raised the question of what makes it truly distinctive and unique and how it might affect science and innovation (Furman, 2016). For the current wave of digitalisation, several core questions emerge about the distinctiveness of new digital technology. What does it enable that was previously impossible or prohibitively expensive? In addition, how will the key features of digital technology, e.g. various externalities, contribute to further developments that could lead to its more intensive use?

Chapter 2 examines how the science system contributes to developing capabilities that can support the digital transition and how the former is impacted by changes in the possibilities and costs associated with digital economic activity (Goldfarb and Tucker, 2017). In science, as in several other fields, the greater information availability brought about by the digital revolution does not necessarily result in greater information quality. Not surprisingly, then, considerable effort in science and innovation aims to deploy digital technologies to help make information useful for meaningful and reliable quality assurance, classification and prediction.

As a result, this chapter also dedicates space to discussing trends and features of research activity related to automating human-like cognitive functions through artificial intelligence (AI). AI is considered to be both a general-purpose technology – i.e. it has a wide domain of applications – as well as a new method of research and invention (Agrawal, Gans and Goldfarb, 2018; Cockburn, Henderson and Stern, 2018; Klinger, Mateos-Garcia and Stathoulopoulos, 2018). Other developments, such as those related to developing computer enabled tamper-proof mechanisms for trust and assurance, are not covered here for reasons of space and limited statistical evidence, but can be just as important.

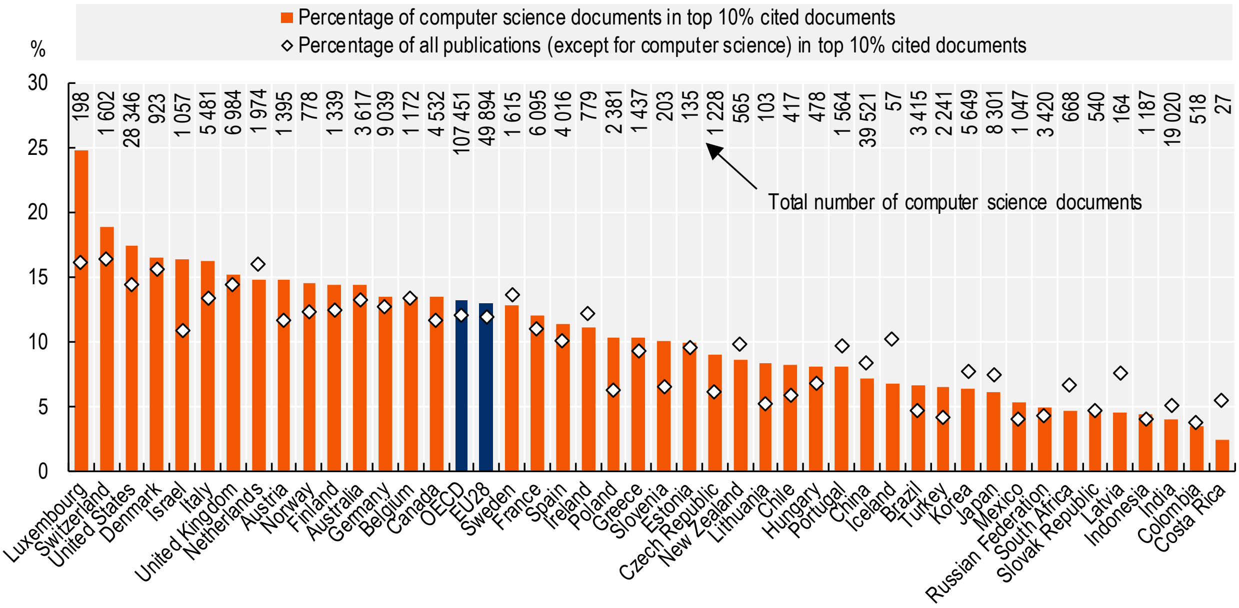
Science going digital

Scientific research on digital technologies

Advances in scientific knowledge are key to developing new digital technologies. Over the last decade, the People's Republic of China (hereafter "China") almost trebled its contribution to computer science journals. In so doing, it overtook the United States in the production of scientific documents in this field. However, China's share of documents that are in the world's top-cited (top 10%, normalised by type of document and field) is still close to 7%, well below the United States at 17% (Figure 2.1).

Figure 2.1. Top 10% most cited documents in computer science, by country, 2016

Percentage of domestic documents (fractional counts) in the top 10% citation-ranked documents



Notes: Computer science publications consist of citeable documents (articles, conference proceedings and reviews) featured in journals specialising in this field. "Top-cited publications" are the 10% most cited papers normalised by scientific field and type of document. Instead of counting a publication repeatedly if two or more countries contribute to it, fractional counting distributes such publication across contributors so that all publications have the same equal weight.
Source: OECD (2019a), *Measuring the Digital Transformation: A Roadmap for the Future*, https://doi.org/10.1787/9789264311992-en.

StatLink https://doi.org/10.1787/888934075697

China's share of highly cited papers has nonetheless more than doubled since 2006. This makes it the second largest producer of highly cited computer science publications worldwide. In some countries, such as Italy, Israel, Luxembourg and Poland, scientific research in the field of computer science has a much higher citation rate than overall scientific production in those countries. Nearly 20% of computer science publications by Switzerland-based authors feature among the world's top 10% cited scientific documents. This figure reaches 25% for Luxembourg, although with a much smaller level of scientific production.

Scientific research and artificial intelligence

Scientific production

AI research has aimed for decades to allow machines to perform human-like cognitive functions. Breakthroughs in computational power, the availability of data and algorithms have raised the capabilities of AI. In some narrow fields, its performance increasingly resembles that of humans. Such advances have allowed computers to distinguish between objects in images and videos, and interpret text through natural language processing, with growing levels of accuracy (OECD, 2017). The 2017 edition of the *OECD Science, Technology and Industry Scoreboard* provided initial evidence on the rapid growth in scientific documents that refer to machine learning – the general method underpinning current advances in data-driven AI – between 2003 and 2016. Interest in AI has triggered several measurement efforts, as documented in Box 2.1.

Box 2.1. Measurement of AI in research, technology and innovation

The OECD supports governments through policy analysis, dialogue and engagement, and identification of best practices. Significant effort is put into mapping the economic and social impacts of AI technologies and applications and their policy implications. This includes improving the measurement of AI and its impacts, as well as shedding light on important policy issues. These issues include labour market developments and skills for the digital age, privacy, accountability of AI-powered decisions, and the security and safety questions that AI generates (OECD, n.d. a).

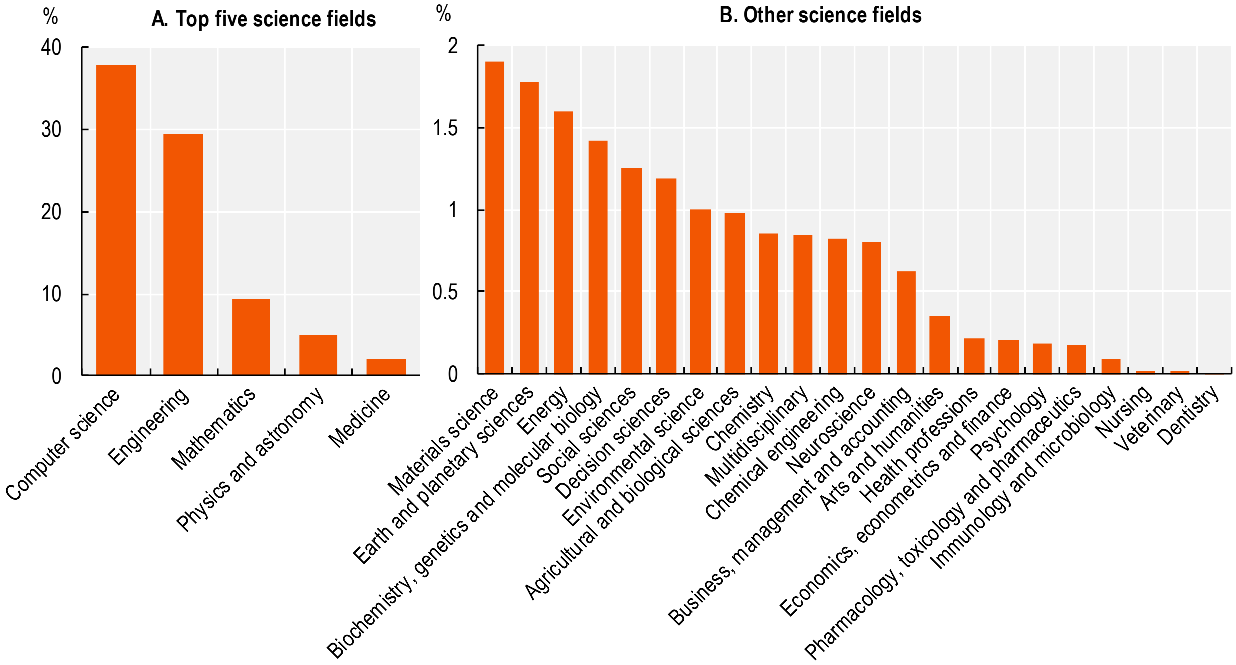
Recent OECD analysis has looked at areas as diverse as scientific publications, conference proceedings, patenting, open-source software and venture capital investment. One such study has used data from Crunchbase, a commercial database on companies around the world, and found that AI start-ups had attracted around 12% of worldwide private equity investments in the first half of 2018, up 3% in 2011. US-based start-ups account for two-thirds of total investment since 2011 (Breschi, Lassébie and Menon, 2018; OECD, 2018c). China has seen a dramatic upsurge in AI start-up investment since 2016. From just 3% in 2015, Chinese companies attracted 36% of global AI private equity investment in 2017. In addition to AI measurement work reported elsewhere in this chapter, recent measurement of AI at the OECD includes analysis in collaboration with Germany's Max Planck Institute for Innovation and Competition using data on patents with and analysis of open-source software publishing. Since 2014, AI open-source software recorded in GitHub grew about three times as much as the rest of open-source software. The number of AI IP5 patent families (namely those registered in the five major intellectual property [IP] offices) went up from close to 1 000 in 2001 to 2 500 in 2014 (Yamashita et al., forthcoming).

Several other public and private, national and international organisations have an active interest in measuring AI. Recent examples include reports by Elsevier (2018) on scientific publications and WIPO (2019), principally on patenting. The Electronic Frontier Foundation, which campaigns to protect civil liberties from digital threats, has started to measure and contextualise progress in AI. This not-for-profit organisation is working to assemble an open-source, online repository of data points on AI progress and performance (Simonite, 2017), benchmarking AI-enabled machine performance compared to humans. The AI Index, backed by the One Hundred Year Study on Artificial Intelligence, was established at Stanford in 2015 to examine the effects of AI on society. This initiative prioritises measurement and uses multiple sources (Shoham et al., 2018), including company reports and executive management surveys (such as Bughin et al., 2017 and McKinsey, 2018).

Text mining of keywords in scientific publications shows that Computer science is the most prevalent domain in AI-related science. It accounts for slightly more than one-third of all AI-related documents published between 1996 and 2016 (Figure 2.2). More than a quarter of all AI-related scientific publications and conference proceedings have appeared in Engineering journals and close to 10% in Mathematics journals. About 25% of the science involving AI (either drawing on AI or contributing to its general advancement) is found in a wide array of other scientific disciplines. These include Physics and astronomy, Medicine and Materials science, among others, demonstrating the growing pervasiveness of AI-related scientific research.

Figure 2.2. Scientific fields contributing to or making use of AI, 2006-16

Journal classification of AI-related scientific documents, as percentage of all AI-related documents



Notes: AI = artificial intelligence. See Box 2.2 for further information.
Source: OECD (2019a), *Measuring the Digital Transformation: A Roadmap for the Future*, https://doi.org/10.1787/9789264311992-en.

StatLink https://doi.org/10.1787/888934075716

Box 2.2. How is AI-relatedness measured in scientific publications and how can it be interpreted?

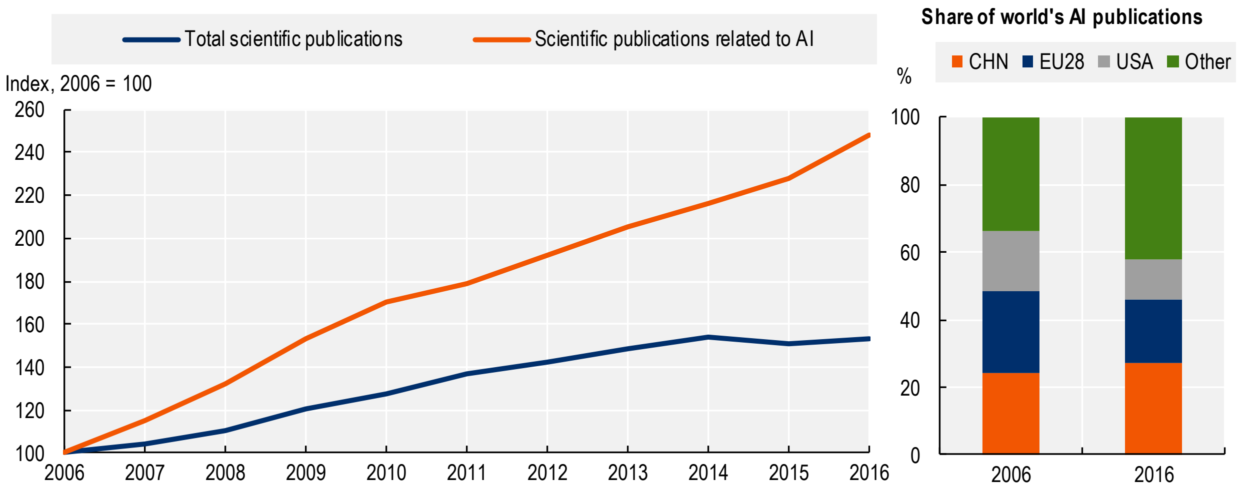
NNNAI-related documents are identified on the basis of Scopus-indexed articles, reviews and conference proceedings using a list of keywords to search on the abstracts, titles and author-provided keywords of scientific documents. The AI keywords are selected on the basis of high co-occurrence with terms frequently used in journals classified as AI-focused (a subcategory of Computer Sciences) by Elsevier, the publisher and provider of bibliographic information and related services.

In the OECD analysis, which focuses on documents published between 1996 and 2016, only those documents with two or more selected keywords were classed as AI documents in order to reduce the risk of including non AI-related documents. Relatedness in this context encompasses instances in which the document presents findings related to existing or new AI procedures. It also includes instances in which the document reports findings based on the application of AI procedures.

The ability to distinguish systematically between enabling and outcome dimensions of AI in the corpus of document titles, abstracts and keywords relies on the consistent recording both of research methods used and findings. As found in the AI literature, automated classification procedures can be substantially enhanced through richer data sources. This implies that analysis could be improved through access to the entire body of documents subject to analysis.

Figure 2.3. Trends in scientific publishing related to AI, 2006-16

Index of publication counts

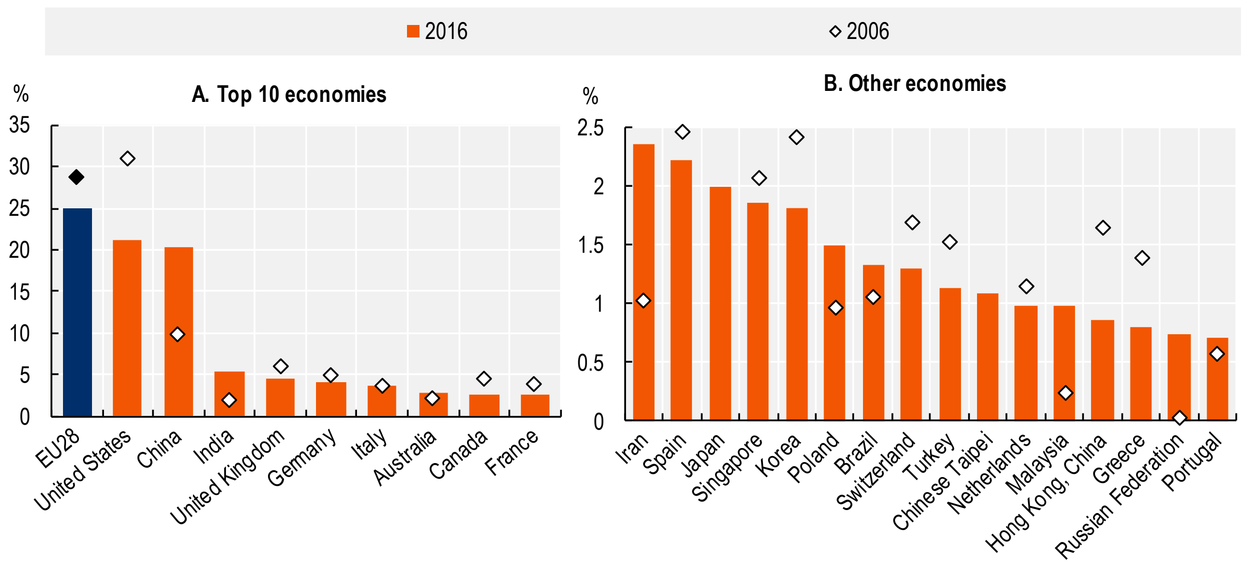


Notes: AI = artificial intelligence. See Box 2.2 for further information.
Source: OECD (2019a), *Measuring the Digital Transformation: A Roadmap for the Future*, https://doi.org/10.1787/9789264311992-en.

StatLink https://doi.org/10.1787/888934075735

Figure 2.4. Top-cited scientific publications related to AI, 2016 and 2006

Economies with the largest number of AI-related documents among the 10% most cited publications



Notes: AI = artificial intelligence. Economies' shares in global AI top-cited publications are based on fractional counts. See Box 2.2 for further information.
Source: OECD (2019a), *Measuring the Digital Transformation: A Roadmap for the Future*, https://doi.org/10.1787/9789264311992-en.

StatLink https://doi.org/10.1787/888934075754

Scientific publishing related to AI (Box 2.2) has experienced a remarkable expansion over the past decade. From 2006 to 2016, the annual volume of AI-related publications grew by 150%, compared to 50% for the overall body of indexed scientific publications (Figure 2.3). China is now the largest producer of AI-related science, in terms of publications, and is fast improving the quality of its scientific production in this area.

Back in 2006, China was already the largest producer of AI-related scientific publications, and grew its global share to 27% by 2016. In turn, the global publication shares accounted for by the EU28 and the United States declined over the same period, to 19% and 12% respectively. Also of note has been the rapid growth of AI-related publishing in India, which in 2016 contributed 11% of the world total. In other areas, however, different AI-related scientific publications have different levels of what is termed “citation impact”. Since it can be misleading to count all publications equally, further analysis has been carried out by focusing on AI-related publications attaining the highest citation rates (the top 10% most cited documents globally) within their respective journal disciplinary domains.

As shown in Figure 2.4, the EU28 and the United States are still responsible for the largest shares of highly cited AI-related publications (i.e. those featuring among the world’s top 10% most cited publications). However, from 2006 to 2016 their shares declined from 29% to 25% for the EU28, and from 31% to 21% for the United States. China, India, Iran and Malaysia all more than doubled their share of the world’s top-cited AI publications over the past decade.

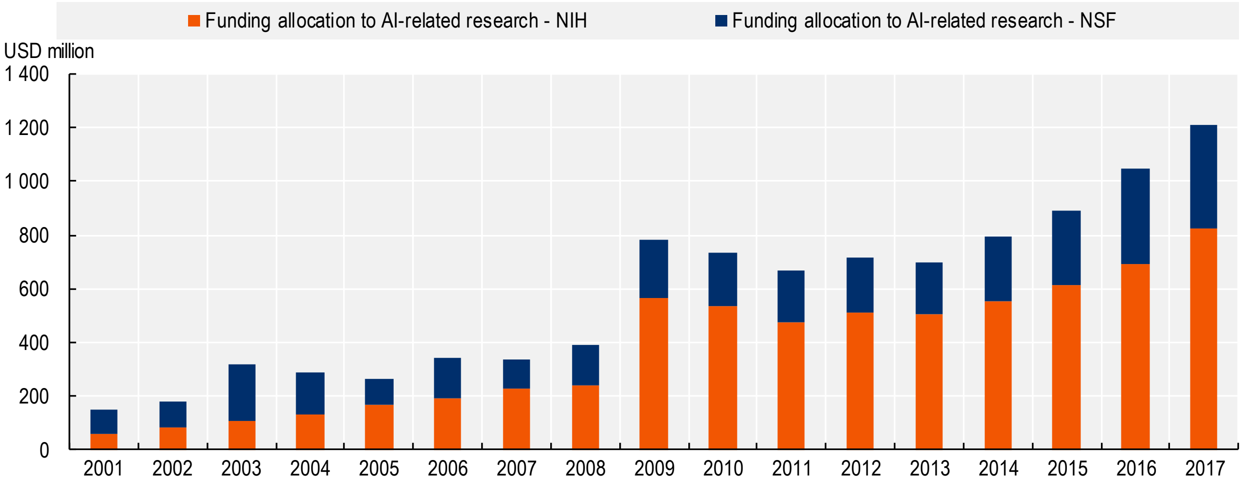
Public funding of scientific research on AI

Given the transformative potential of AI, it is worth examining the scale and nature of government and business investment. There has been a plethora of policy announcements across countries that are difficult to compare. A 2016 White House report indicated that the United States invested USD 1.1 billion in “AI research and development [R&D]” in 2015, rising to USD 1.2 billion in 2016 (NSTC, 2016). The European Commission estimates it has dedicated close to 13% of its R&D budget to information and communication technology (ICT) since 2014 (EC, 2018). The United Kingdom’s Engineering and Physical Sciences Research Council has allocated more than GBP 400 million (USD 527 million) for research related to data science and AI through different mechanisms (BEIS and DCMS, 2018). In December 2017, Korea’s Ministry of Science and ICT announced plans to dedicate in 2018 the equivalent of USD 1.5 billion to AI and related areas in support for the “fourth industrial revolution” (EDaily, 2017). Japan’s Prime Minister Abe established a Strategic Council for AI Technologies in 2016 to ensure a co-ordinated approach across ministries and agencies for AI research, including new AI labs and complementary R&D centres.

Because AI does not fit neatly into pre-established taxonomies of R&D funding, detailed information sources at the micro level are needed to produce reliable and relevant statistical information. Available data systems and statistics lack systematic granular information about what publicly funded researchers work on, as opposed to what they publish. This makes them ill equipped to address subject-specific questions. Data on government-funded projects (often allocated on a competitive basis) provide a useful but partial view of the funding landscape that is most accurate when project-based funding dominates over other resource allocation mechanisms for scientific research funding. No international data infrastructure brings together research funding agencies’ databases on the basis of an explicit agreement that renders them comparable. A number of commercial providers grant related information services based on data collected from publicly available sources or bilateral data-sharing agreements. The OECD is seeking to addresses this information gap by assessing the feasibility of a shared data resource for analysis through the Fundstat pilot project. The OECD has also begun new work to map research funding trends using case studies for demonstration purposes and focusing on AI given its high policy relevance.

To date, the case studies have focused on two major US agencies, the National Institutes of Health (NIH), one of the world’s main funders of biomedical research, and the National Science Foundation (NSF), which covers several areas including civilian computer science research.[1] The analysis uses funding data from 2001 to 2017 from the NIH RePORTER database (over 1.2 million granted applications) and the NSF Award Search System 2018 (over 200 000 granted applications). Over less than two decades, the share and volume of AI-related funding has grown significantly for both agencies. AI-related funding in 2017 (Figure 2.5) represented close to USD 820 million for NIH (i.e. 3.6% of total NIH health R&D funding) and USD 388 million for NSF (7.3% of NSF R&D funding).

Figure 2.5. Estimated NIH and NSF funding for AI-related R&D, 2001-17

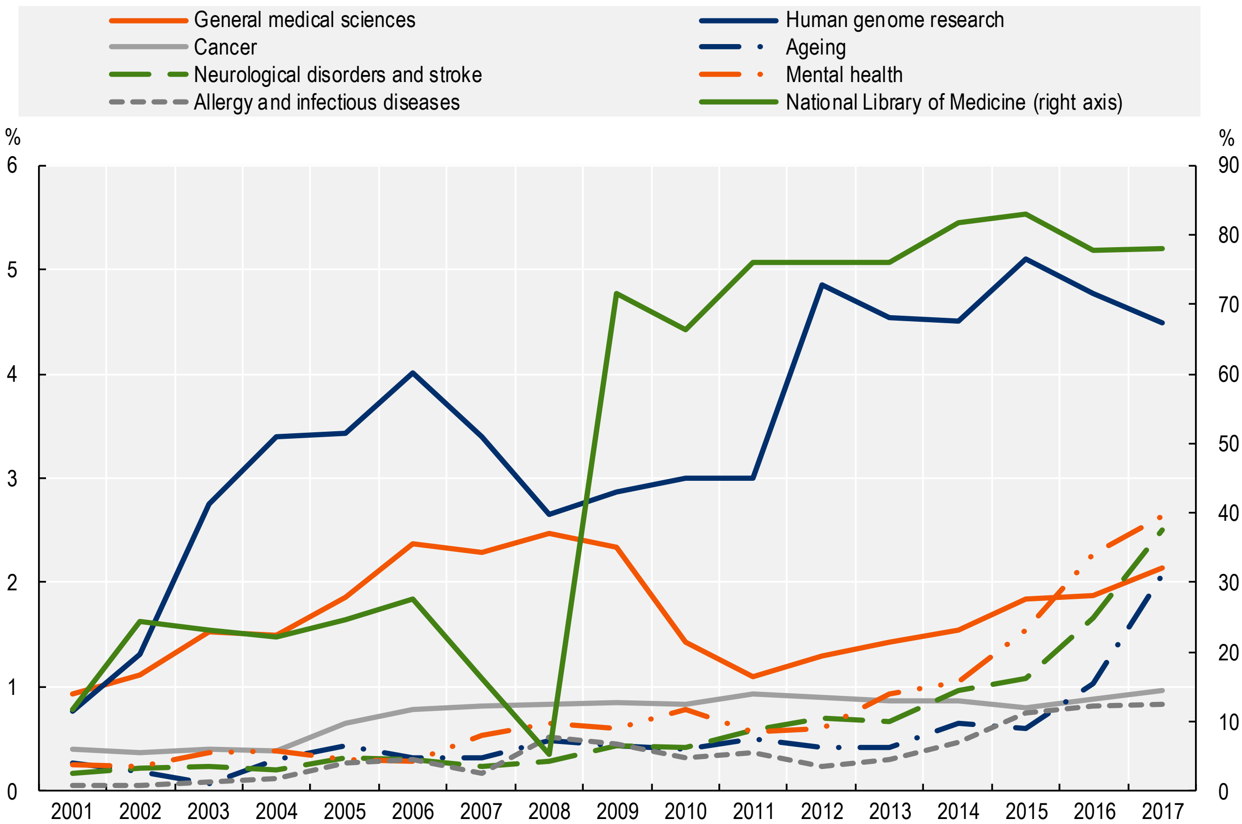


Notes: AI = artificial intelligence; NIH = National Institutes of Health; NSF = National Science Foundation. This is an experimental indicator.
Source: OECD calculations based on *NIH RePORTER* (database) and *NSF Award Search* (database) (accessed 1 December 2018).

StatLink https://doi.org/10.1787/888934075773

Figure 2.6. Estimated share of AI-related R&D funding within NIH institutes

"AI intensity" for selected institutes with the largest estimated amounts of AI funding



Notes: AI = artificial intelligence; R&D = research and development; NIH = National Institutes of Health. This is an experimental indicator. For clarity of presentation, and with the exception of the National Library of Medicine, responsible NIH institutes' names are presented by referring solely to their missions/subjects.
Source: OECD calculations based on *NIH RePORTER* (database) (accessed 1 December 2018).

StatLink https://doi.org/10.1787/888934075792

Analysis of NIH-AI funding data shows which institutes appear to make more intensive use of AI, as implied in the awards granted (Figure 2.6). The National Library of Medicine (NLM) (secondary axis) accounts for the largest share of AI-related research within NIH (about one-third of the total). It also has the highest internal AI intensity at close to 80%, followed by the National Human Genome Research Institute at 5%. In total funding terms, NLM is followed by the National Cancer Institute, which has an AI intensity of less than 1%.

Figure 2.7 shows the incidence of AI-related R&D within the NSF directorates with responsibility for managing the funding for different disciplinary domains. AI intensity in 2018 is more than 35% in the case of Computer and information sciences (displayed on the secondary axis), up from less than 10% in 2001. This is followed by Engineering (general) at 11%, up from nearly 2% in 2012.

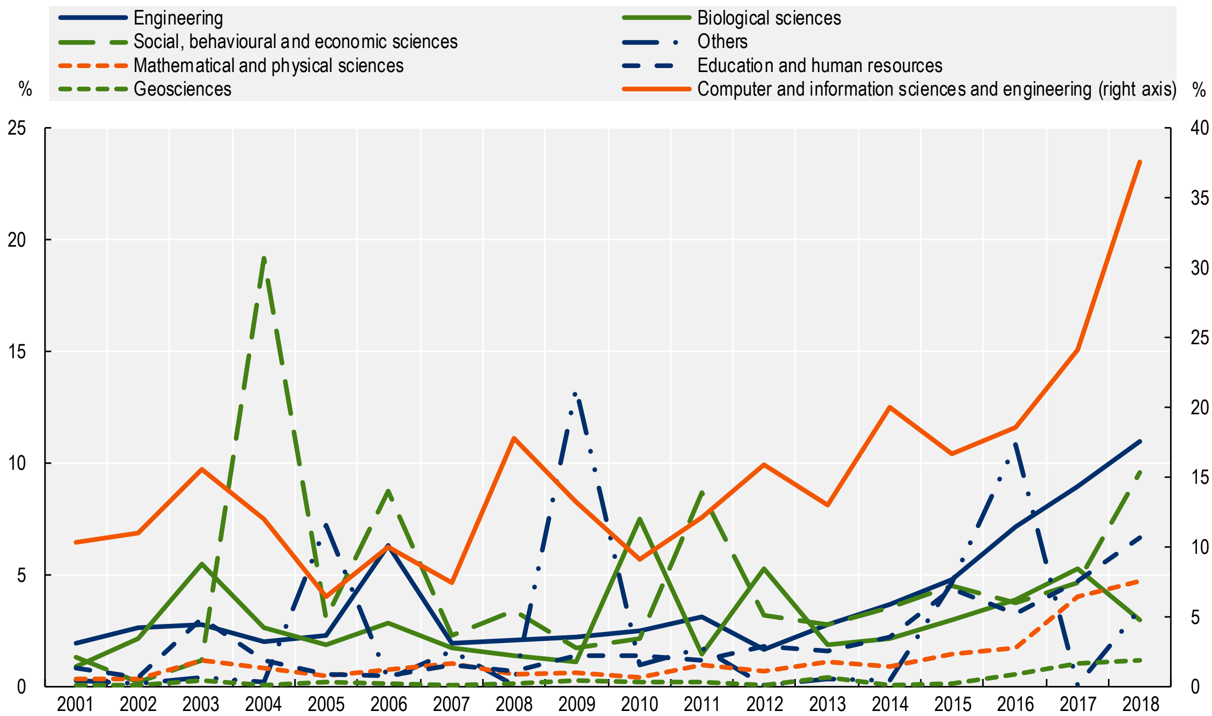
The use of funding data through text analysis is a promising avenue for understanding developments in AI research. Funding data help develop a timelier and more finely grained picture that connects funding agencies, their missions and traditional disciplinary areas. This can be an important complement to measurement on AI in related domains. The challenge is to work towards securing comprehensive data sources with high-quality text descriptions about the nature of R&D projects across several countries. More than a big data challenge, this is a co-ordination challenge that policy makers can help address, particularly in light of the OECD *Recommendation of the Council on Artificial Intelligence* (OECD, 2019b). The OECD council recommendation does explicitly state that governments

> *"should consider long-term public investment, and encourage private investment, in research and development, including interdisciplinary efforts, to spur innovation in trustworthy AI […]".*

Monitoring this recommendation requires concerted action.

Figure 2.7. Estimated share of AI-related R&D funding within NSF disciplines

"AI intensity" for selected disciplinary directorates with the largest amounts of AI funding



Notes: AI = artificial intelligence; R&D = research and development; NSF = National Science Foundation. This is an experimental indicator.
Source: OECD calculations based on *NSF Award Search* (database) (accessed 1 December 2018).

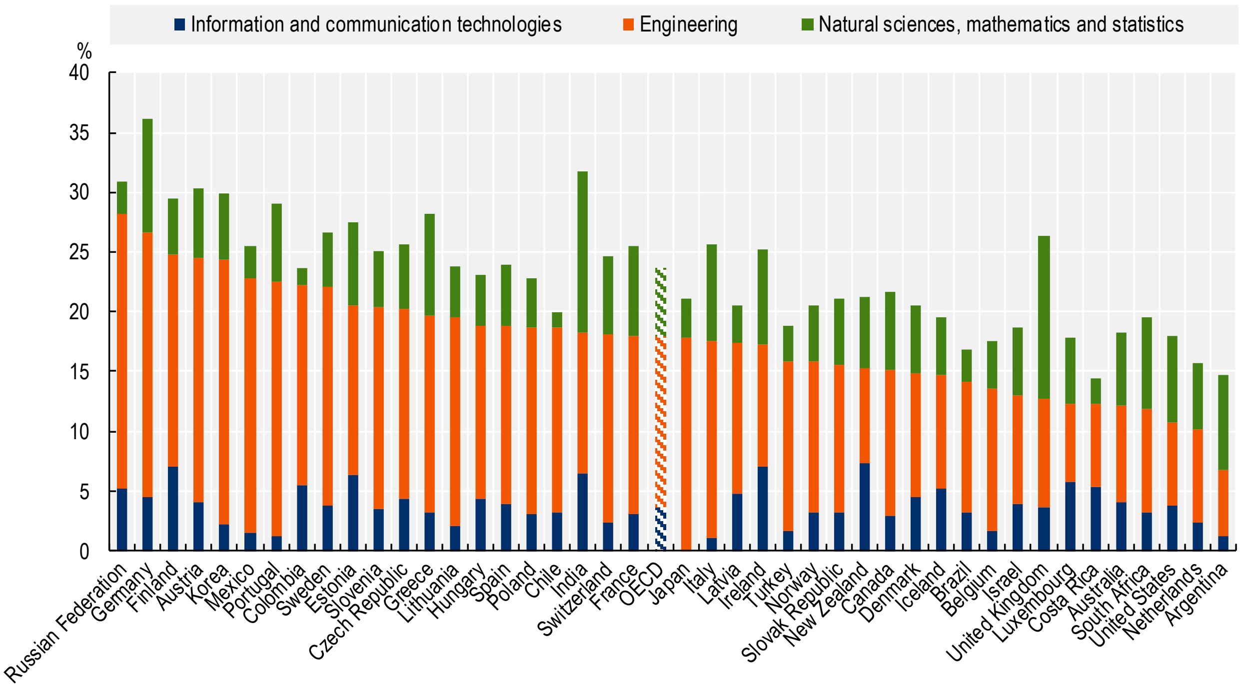
StatLink https://doi.org/10.1787/888934075811

The science system and its contribution to the development of digital skills

Any overview of how science and innovation are digitalising must examine how the system helps develop skills and competences critical to the digitalisation process within science itself and across society and ultimately utilises them. Figure 2.8 presents the distribution of new tertiary graduates in the natural sciences, engineering and ICT fields for 2016. It shows that Estonia, Finland, India and Ireland have the largest shares of graduates in designated ICT fields.

Figure 2.8. Tertiary graduates in natural sciences, engineering and ICT fields, 2016

As a percentage of all tertiary graduates



Notes: ICT = information and communication technology. Data on ICT graduates for Japan are included in other fields. The Netherlands excludes doctoral graduates. Data for China not included because of reporting differences. Natural sciences and engineering account for about 25% of higher education institution graduates (60% for new doctorates). http://en.moe.gov.cn/Resources/Statistics/edu_stat2017/national/index_2.html.
Source: OECD (2018a), *Education at a Glance: OECD Indicators*, https://doi.org/10.1787/eag-2018-en.

StatLink https://doi.org/10.1787/888934075830

Data from the OECD's publication Education at a Glance show differences in numbers of graduates in ICT subjects at different levels of attainment (Table 2.1). For example, European countries graduate many doctoral students relative to those with lower levels of attainment. Conversely, in Korea, the United States and India relatively few individuals graduate at doctorate level given the numbers of graduates at the bachelor's level. This may be due to differences in the opportunity cost of staying on for postgraduate study.

Higher education institutions (HEIs) can also prepare individuals to make use of advanced ICT skills in domains other than the computer sciences. Initiatives like the Open Syllabus Project can provide a basis for analysing the content of instruction in HEIs across different subjects. They can also provide insight into trends in the teaching of digitally based methods (OSP, n.d.). Researchers have used data from this project to compare offered tuition with skills demand and developments in scientific research. For example, Börner et al. (2018) compare features of academic syllabi, scientific publications and job advertisements. They show that the distribution of skills taught in the classroom is three to four times closer (in terms of content similarity) to skills described in research articles than the skills specified in job advertisements. Skills related to specific software and computational tools (often referred to as data science related [Box 2.3]) are found in the three types of documents. However, they tend to be highly specialised (not present in many courses, for example). Conversely, general research, management, problem-solving and management skills are both central to courses and job ads. Skills related to computational tools appear to show mutual predictability across scientific publications and job requirements, as if course offerings both anticipate and react to employers' needs. This highlights a form of close interdependency between science and industry in this particular area, one not captured by standard indicators of science-industry knowledge flows.

Table 2.1. Graduates in ICT at different levels of attainment, selected countries, 2016

	Bachelor's level	Master's level	Doctorate level
France	9 370	9 827	630
Germany	15 931	8 380	1 021
Korea	7 837	1 018	154
United Kingdom	15 275	6 733	1 136
United States	69 436	41 002	1 951
India	338 062	211 693	507
Russian Federation	31 087	29 251	1 860

Notes: ICT = information and communication technology. Data on China are not available because of lack of comparable data under new ISCED-Fields classification.
Source: OECD (2018a), *Education at a Glance: OECD Indicators*, https://doi.org/10.1787/eag-2018-en.

Box 2.3. Data science and data scientists

The US NIH define "data science" as "the interdisciplinary field of inquiry in which quantitative and analytical approaches, processes, and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data" (NIH, 2018). Google's chief economist, Hal Varian, foresaw this trend when he argued in 2009 that the "sexy job in the next 10 years" would be "statistician" (Varian, 2019). This prediction has in a sense come true for those who are known as data scientists (OECD, 2018b).

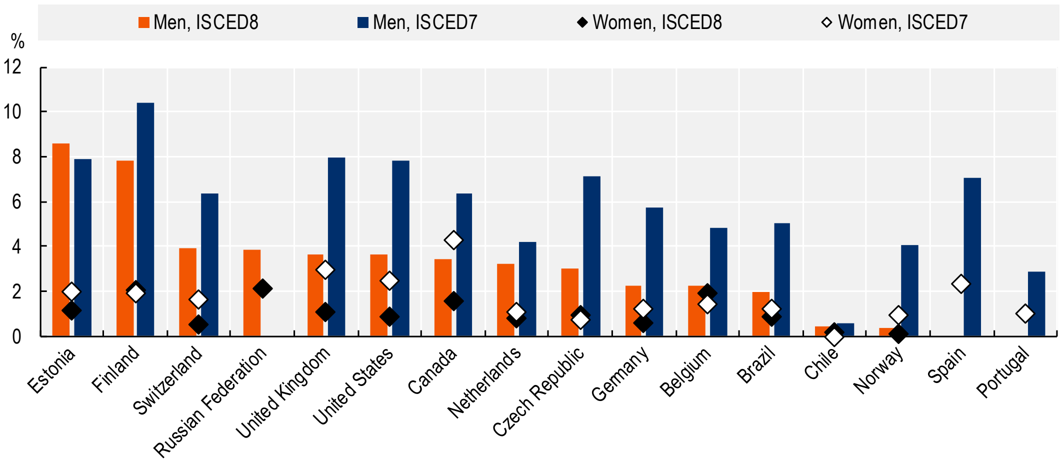
The term "data scientist" is now widely used in business and management contexts not conventionally associated with scientific research. It refers to individuals with formal training at the junction of computer and decision sciences, modelling, statistics and applied mathematics. The particular combination of knowledge and skills, however, goes beyond those used in traditional business analytics posts. It allows data scientists to harness and interpret vast and growing amounts of data and information. Ultimately, this connects them to organisational decision making.

Are universities training a sufficient number of individuals who can do advanced research on digital tools and systems? Evidence from the 2017 OECD collection of data on the Careers of Doctorate Holders (CDH-light) shows that ICT doctorates account for a relatively small share of the doctorate population, typically with lower shares than at master's or lower levels of tertiary attainment (Figure 2.9). Available figures indicate that at both doctorate and master's levels, the share of ICT graduates is much higher among men than women.

While the history of computer science has seen periods when, like in the 1960s, women made up the majority of computer programmers, doctoral education among women was rare. It was only in 1965 that the first doctorate in computer science was awarded to a woman – Mary Keller – in the United States. In 2005, the proportion of women among entrants to doctoral programmes in the United States was just below 20%, a value close to the OECD average (OECD, 2018a). In most countries, the share of female entrants to doctoral programmes is below 30%, which is less than for engineering programmes. These figures are similar to entry shares at bachelor's or equivalent levels.

Figure 2.9. Individuals holding master's (ISCED7) and doctorate (ISCED8) level degrees in ICT, 2016

As a percentage of graduates in all fields, by sex and attainment level

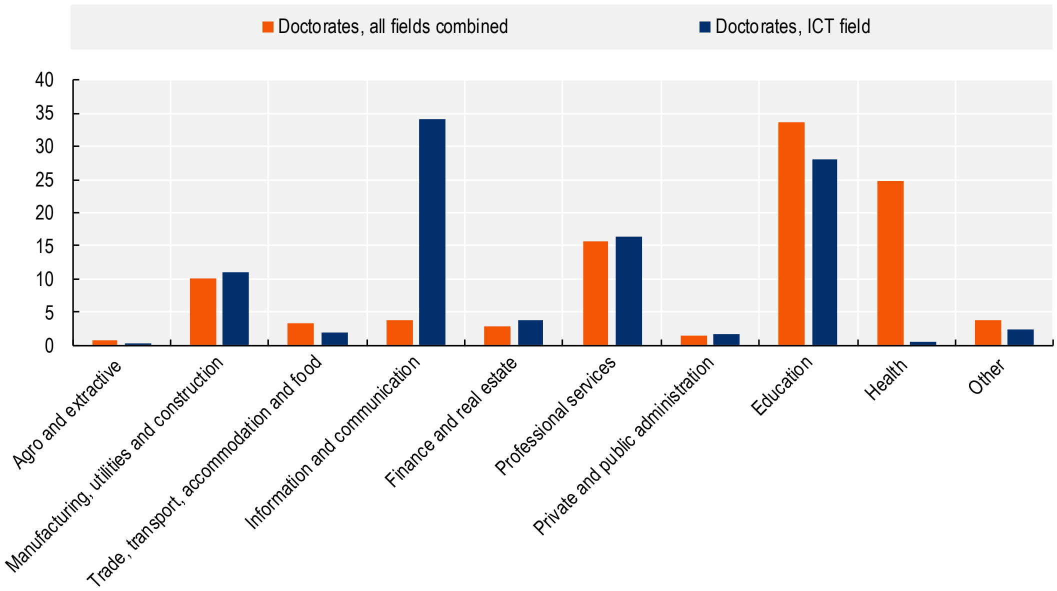


Note: ICT = information and communication technology.
Source: OECD calculations based on OECD (n.d. b), *Careers of Doctorate Holders database*, http://oe.cd/cdh.

StatLink https://doi.org/10.1787/888934075849

Figure 2.10. The distribution of ICT doctorates across industries

As a percentage of all doctorates with a degree in ICT or any field



Notes: ICT = information and communication technology. Estimates based on data for Belgium, Brazil, Canada, Finland, Germany, the Netherlands, Switzerland and the United Kingdom.
Source: OECD calculations based on OECD (n.d. b), *Careers of Doctorate Holders database*, http://oe.cd/cdh.

StatLink https://doi.org/10.1787/888934075868

In contrast with the low gender diversity of doctorate-educated individuals in the ICT area, these graduates are among the most likely doctorates to have been born abroad among the economies covered in the CDH2017 data collection. This indicates that the supply of skills in this area is potentially more exposed to policies that tighten up residential visa or nationality requirements. There are also significant reallocations within digitally oriented fields. In the United States, the number of doctoral recipients from domestic universities in Computer science increased by 20%, while in Electrical, electronics, and communications engineering decreased by 3% over the decade from 2007 to 2017. This compares to an overall growth for all engineering fields of 27% and 13% for all fields of science (National Science Foundation, 2018).

These individuals with high research competences in ICT-related subjects are found principally in the ICT industry, followed by professional services (which includes R&D specialist firms) and higher education (Figure 2.10). Holders of ICT doctorates are also more oriented to work in the business sector than the average doctoral graduate. CDH data also show that doctorate holders in the field of ICT are significantly more mobile across jobs than their counterparts in other fields. For example, in the United States, 30% of ICT doctorates have changed jobs in the last year, compared to a 15% average across fields.

Scientific research enabled by digital technology

As discussed in Chapters 1 and 3, digitalisation is changing the way research is conducted and disseminated. To examine the emerging patterns of digitalisation in science, the International Survey of Scientific Authors (ISSA) (Box 2.4) asked a global sample of scientists a number of questions. These included such questions as whether digital tools make scientists more productive; to what extent they rely on big data analytics, or share data and source codes developed through their research; and to what degree they rely on a digital identity and presence to communicate their research. Preliminary survey results reveal contrasting patterns of digitalisation across fields.

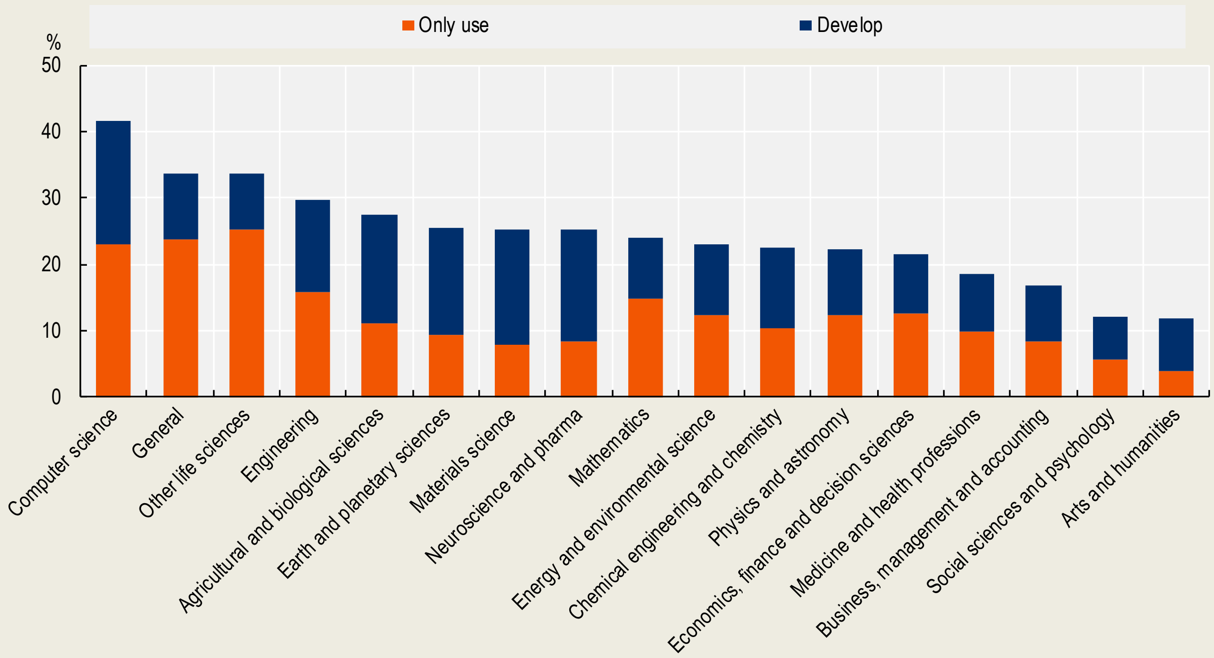
The use of advanced digital tools, including big data, is a defining feature of the computer sciences, followed by multidisciplinary research, mathematics, earth and materials sciences and engineering (Figure 2.12). The life sciences (with the exception of pharmaceuticals) and the physical sciences (other than engineering) report the largest relative efforts to make data and/or code usable by others. There are smaller systematic differences in the reported use of productivity tools, which happen to have much higher general adoption rates. Scholars in the engineering domains report using productivity tools relatively less frequently. Interestingly, the fields making less use of advanced digital and data/code dissemination tools – namely those in the social sciences, arts and humanities – are more likely to engage in activities that enhance their digital presence and external communication (e.g. use of social media).

Box 2.4. The OECD International Survey of Scientific Authors

During the last quarter of 2018, the OECD contacted a large, randomly selected group of authors of scholarly documents. The group was asked to respond to an online survey aimed at identifying patterns, drivers and effects of digitalisation in scientific research. This OECD ISSA obtained rich information from nearly 12 000 scholars worldwide about their use of a broad range of digital tools and related practices, in addition to other key demographic and career information.

Answers to 36 questions were analysed to identify four major "latent" factors. These represent how likely scientists are to i) make use of productivity tools to carry out regular tasks such as retrieving information and collaborating with colleagues; ii) make data and code outputs available to others; iii) use or develop unconventional data and computational methods; and iv) maintain a digital identity expanding their communication with peers and the public in general. Analysis of a variable closely correlated with the third factor shows the digitalisation of science is not limited to scientific fields that specialise in computer science or information technology engineering. More detailed results and analysis from this study will be available on the project website http://oe.cd/issa.

Figure 2.11. Use and development of big data across scientific domains, 2018



Notes: This is an experimental indicator. “Other life sciences” include: Biochemistry, Genetics, Molecular biology, Immunology and microbiology. “Big data” refers to authors who answer that their teams use or develop “data with size, complexity and heterogeneity features that can only be handled with unconventional tools and approaches, e.g. Hadoop”. Estimates are weighted and take into account the sample design as well as non-response.
Source: OECD calculations based on OECD (n.d. c), *OECD International Survey of Scientific Authors 2018*, http://oe.cd/issa.

StatLink https://doi.org/10.1787/888934075887

Differences in digitalisation patterns are also marked by personal and sectoral employment characteristics. Younger scientists are more likely to engage in all four dimensions of digital behaviour. This confirms digitalisation patterns found in ICT use surveys addressed to individuals in the general population. Female scientists are less likely than their male counterparts to use and develop advanced digital tools. However, they are more likely to engage in enhancing their digital presence, identity and communication, even after accounting for differences in field and country.

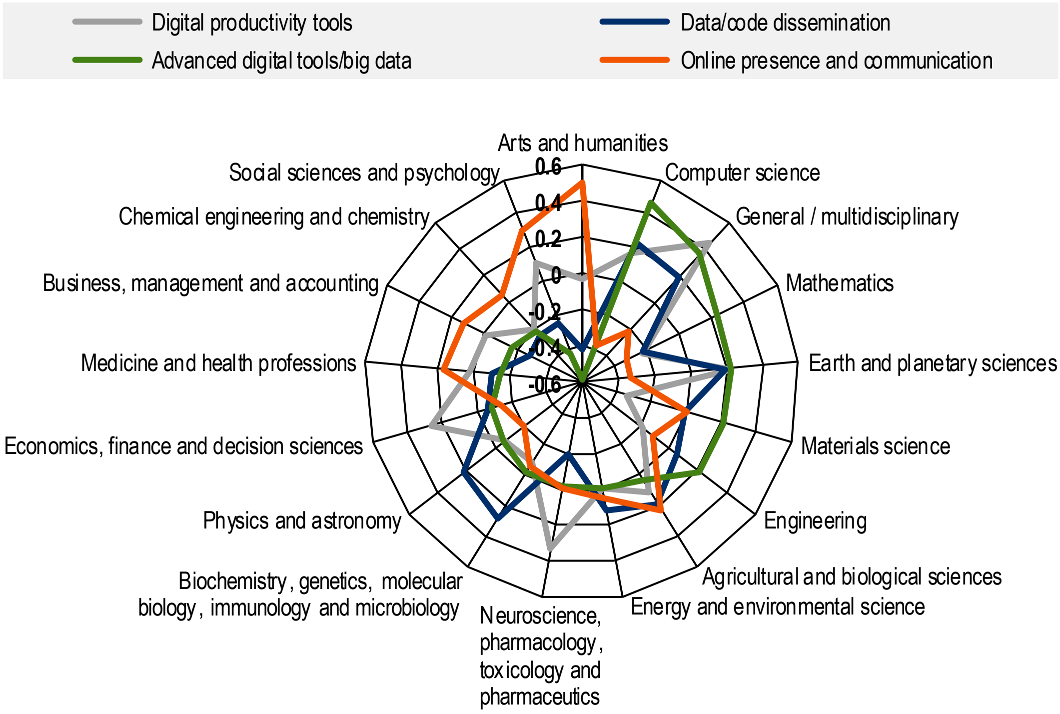
Scientific authors that work in the business sector are also more likely than those in other sectors to make use of advanced digital tools linked to big data and less likely to engage in data/code dissemination activities and online presence and communication. By contrast, authors in the higher education sector are more prone to use digital productivity tools (indeed most of those presented in the survey are related to academic tasks), as well as online presence and communication.

Research paradigms and digitalisation

Since digital tools can transform how scientific research is conducted, ISSA survey respondents were allowed to describe their scientific research work with respect to the use of theory, simulations, empirical non-experimental and experimental activity, and combinations among these. Scientific research practices correlate with digital practices in complex ways. Researchers engaged in computational and modelling work (37% of the sample) are the most likely to use advanced digital tools. However, they are also less likely to engage in online presence and communication activities. Together with researchers involved in experimental work (49%), they are also the most likely to engage in data and code dissemination practices, for example through platforms such as GitHub.

Figure 2.12. Patterns of digitalisation in science across fields, 2018

Average standardised factor scores for four different facets of digitalisation, by field



Notes: This is an experimental indicator. This figure presents average scores for four latent factors representing different facets of digitalisation for each scientific field. The factor analysis is based on responses by scientists to 36 questions relating to digital or digitally enabled practices. These are combined in four synthetic indicators that have been normalised to have overall zero average and identical variance.
How to read this figure: computer science's highest score for the factor representing use of advanced digital tools (grey line) represents high relative intensity on this facet. Conversely, a low relative intensity is seen on the digital facet representing online presence and communication (dotted line) for scientists in this area.
Source: OECD calculations based on OECD (n.d. c), *OECD International Survey of Scientific Authors 2018*, http://oe.cd/issa.

StatLink https://doi.org/10.1787/888934075906

Those reporting work on gathering information (37%) are surprisingly not among those most likely to disseminate data and code. This suggests considerable scope for digitalisation of their data diffusion activity. Among this group, the use of digital productivity tools is nonetheless high. Those involved in theoretical work (46%) tend to make limited use of most digital practices. The incidence of digital practices among those undertaking empirical, non-experimental work (45%) is most common in the social sciences. It is relatively constrained in terms of data/code dissemination (creating a challenge for replicability) and advanced digital tools.

Open science and digitalisation

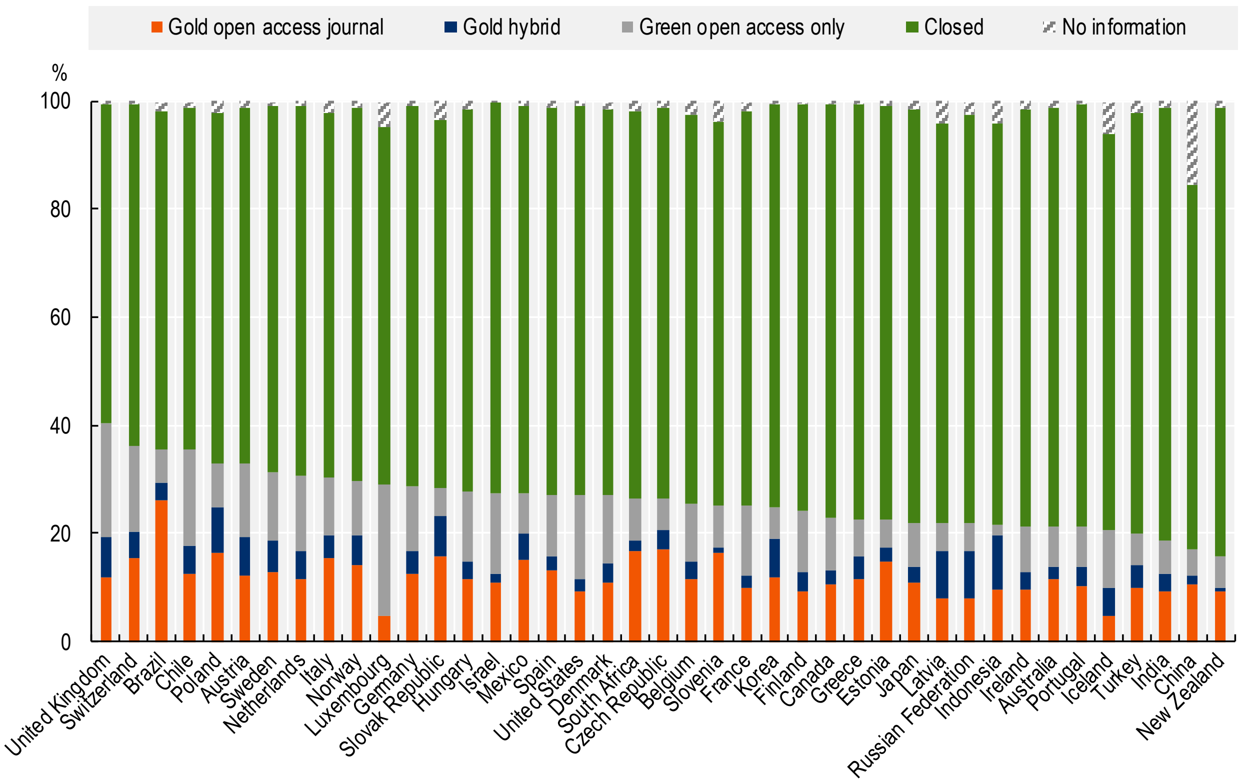
One important avenue of enquiry relates to the scope for digitalisation to address some perceived structural problems in how research is collectively organised. As Chapter 3 discusses, digitalisation offers a variety of opportunities for open science practices. For example, digitalisation can help reduce transaction costs; promote data reuse; increase rigour and reproducibility; and decrease redundant research. It can also better involve patients, consumers and others; facilitate researcher transparency in sharing information on processes and results; and improve connections between a larger variety of actors to produce more innovative approaches and solutions (Gold, 2016). Open Science encompasses multiple dimensions, including unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and complementary incentives. Broadening access to scientific publications, data and code is at the heart of open science so that potential benefits are spread as widely as possible (OECD, 2015b). Interest is growing in monitoring the use of such practices (Gold et al., 2018).

Open access to documents

Access to scientific research articles plays an important role in the diffusion of scientific knowledge. Digital technology facilitates the sharing of scientific knowledge to promote its use for further research and innovation. Open access (OA) indicators reported in OECD (2017) reveal that 60% to 80% of content published in 2016 was, one year later, only available to readers via subscription or payment of a fee (Figure 2.13). Journal-based OA (usually termed "gold" OA) is particularly noticeable in Brazil, as well as in many other Latin American economies. Repository-based OA (also known as "green" OA) is especially important for authors based in the United Kingdom. About 5% of authors appear to be paying a fee to make their papers publicly available in traditional subscription journals (also known as "gold hybrid" OA).[2]

Figure 2.13. Open access of scientific documents, 2017

As a percentage of a random sample of 100 000 documents published in 2016, by country of affiliation



Source: OECD (2017), "Open access of scientific documents, 2017: As a percentage of a random sample of 100 000 documents published in 2016", https://doi.org/10.1787/sti_scoreboard-2017-graph66-en.

StatLink https://doi.org/10.1787/888934075925

Assessing the extent to which OA publications receive more citations than non-OA publications helps policy makers evaluate the social costs and benefits of alternative mechanisms for funding scientific publication. This has led to efforts to measure the "open access citation advantage". Bibliometric analysis confirms previous findings of a mixed picture (OECD, 2015b; Boselli and Galindo-Rueda, 2016), as not all forms of OA appear to confer a citation advantage. OA is in general associated with higher citation rates among documents covered by major indices. However, this does not apply to documents published in OA journals, which on average tend to be more recent and present lower historical citation rates. Repository-based (green) OA systematically confers a citation advantage. In most cases, higher citation rates are generally found for “gold hybrid” documents. These are articles published in subscription journals whose authors pay publishers a fee to enable free online access on the part of potential readers. The ISSA1 study showed that researchers had a positive willingness to pay to disseminate their result conditional on their paper being accepted. The results from the ISSA1 and ISSA2 studies confirm that authors of documents in gold OA journals tend to report significantly lower earnings, point to strong and self-reinforcing prestige effects that are dissociated from dissemination objectives in the digital era (Fyfe et al., 2017). Evidence points to OA increasingly becoming the norm. Moreover, incumbent high prestige journals look likely to take advantage of their current citation advantage. This leads to the fundamental question: what type of OA model will prevail in the longer run for signalling quality?

Open access to data and code

Measuring and understanding access to data and code are also important for mapping open science practices. The ISSA2 study has gone beyond probing the access status of publications. It also considers the status of other research outputs, in particular the code and data reported by authors to have been developed as part of the published research. The study shows that less than half of respondents in all science fields deliver data or code to a journal or publisher as support to their publication. The use of repositories for data archiving and dissemination seems to be most common among respondents in the life sciences. Informal data or code sharing among peers seems to be the main way researchers in all fields make data available to others.

The publication of research data or code does not imply that other researchers can easily use and reuse them. Use might be impaired if the access costs are prohibitive or access implies other challenges. For this reason, the ISSA2 survey asks about charging policy. It also asks about attributes that are part of the open science principles of findability, accessibility, interoperability and reusability.

The practice of adopting standard mechanisms for requesting and securing data access seems to be uncommon in all disciplines. Less than 30% of respondents indicated using such mechanisms when sharing their data or codes. Likewise, a low percentage of respondents (about 10%) applied a data usage licence to their data. Reusability of data seems to be ensured mainly through the development and provision of detailed and comprehensive metadata, especially in the physical sciences and engineering. Compliance with standards that facilitate data combination with other sources is more common in health and life sciences, whereas it seems to be less diffused in the physical sciences and engineering.

In all fields, authors tend to report several barriers to access of scientific outputs. These include formal sharing requirements set by publishers, funders or the respondent’s organisation; IP protection systems; and resources necessary for dissemination. Career objects and peer expectations were pre-eminently reported as drivers of enhanced access. Privacy and ethical considerations tend to limit access to scientific outputs in health sciences. Dissemination costs in terms of time and money are deemed strong barriers. However, respondents do not consider capabilities for managing disclosure and sharing as important either way.

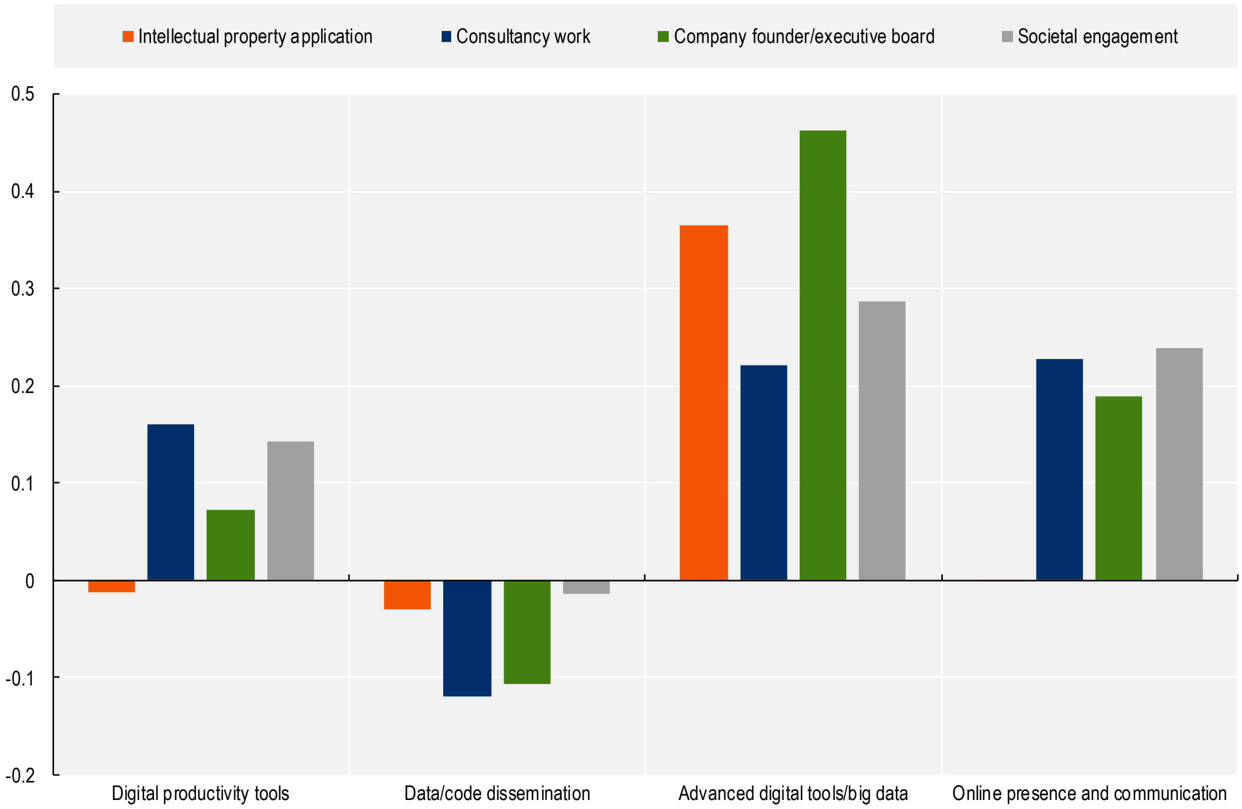
Digitalisation and the broader impacts of science

Another key policy question is the extent to which scientists that engage in non-academic activities exhibit different patterns of digital competence. Data from the 2018 ISSA suggest that scientists who have applied or registered for IP protection; done consultancy work; started new companies or served as executives; and engaged in various societal outreach activities, such as supporting the work of museums and charities tend to exhibit also higher levels of competence in advanced digital tools (Figure 2.14).

Those scientists who have started companies or served as executives – about 20% of the sample – had the biggest advantage in advanced data competences; the gap is close to one-half of the standard deviation for this latent factor. The gap is also particularly large for persons engaging in IP application or registration (reported in about one-fifth of cases), and significant too for those undertaking consultancy work and societal engagement.

Figure 2.14. Digital activity of scientific authors by engagement in external activities, 2018

Difference in digital intensity scores between authors active and non-active in external activities



Notes: How to read this chart. Scientists who have founded companies or served as executives have an expected latent competence in advanced digital tools that is 0.45 standard deviations larger than those who have not. In contrast, their expected competence in digital productivity tools is much closer to that of others, with a difference of less than 0.1 standard deviations. See notes for Figure 2.12.
Source: OECD calculations based on OECD (n.d. c), *OECD International Survey of Scientific Authors 2018*, http://oe.cd/issa.

StatLink https://doi.org/10.1787/888934075944

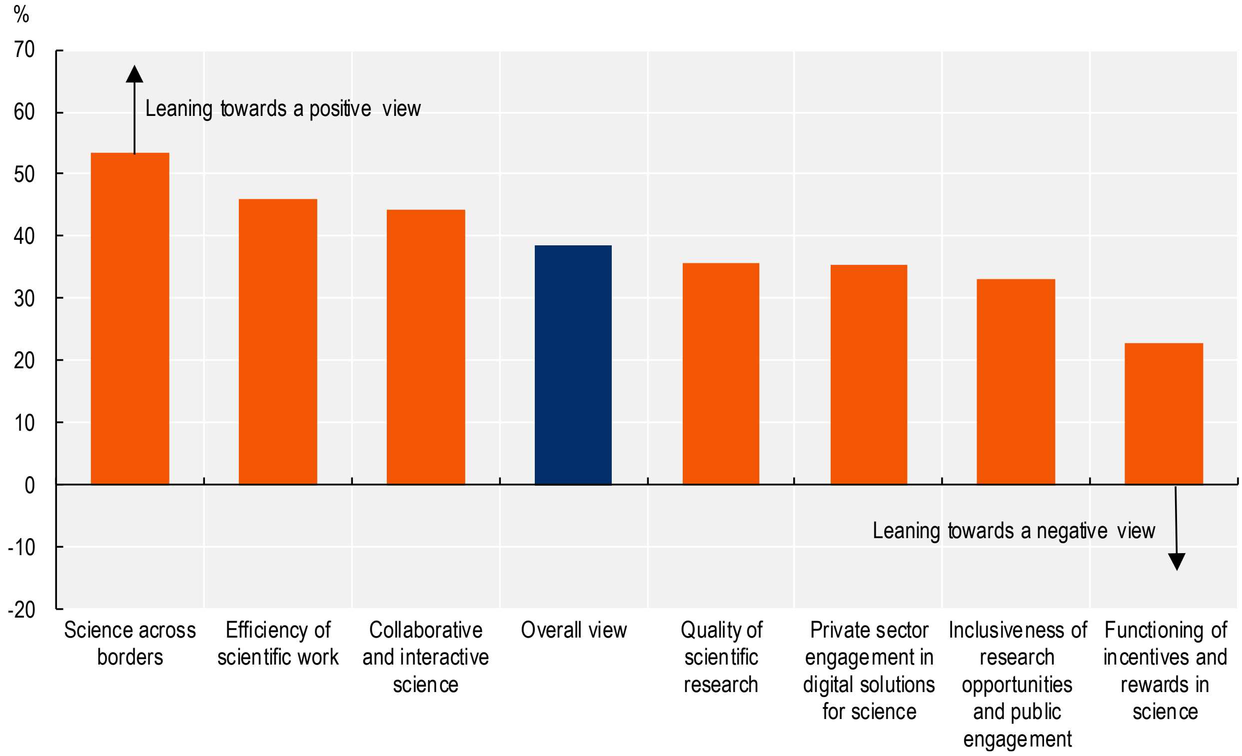
All this points to the high demand for these skills in the economy and society. The digital advantage in terms of individuals' online presence and communication is particularly marked for those engaged in societal outreach activities (also political work, not reported in the chart) and consultancy work. There is no significant difference in this digital factor for those active in IP and those who are not.

Looking ahead: Scientists' perspectives on digitalisation and its impacts

How do scientists themselves view the digital transformation of scientific research and its impacts? Evidence from the 2018 ISSA study suggests that scientists are on average positive across several dimensions (Figure 2.15). Many respondents feel that digitalisation has positive potential to promote collaboration, particularly across borders, and improve the efficiency of science. While remaining positive, scientists appear less optimistic regarding the potential impact of digitalisation on the system of incentives and rewards. Specifically, they are concerns about being rated on the basis of their digital "footprint", such as their publications and citations, as well as downloads of their work. They also have reservations about whether digitalisation can bring scientific communities and scientists together with the public (inclusiveness). Finally, they sometimes question the role of the private sector in providing digital solutions to assist their work. Younger authors are generally more positive than their older peers, except with respect to the impacts of digitalisation on the incentive system, which may reflect concerns about their future careers.

Figure 2.15. Scientists' views on the digitalisation of science and its potential impacts, 2018

Average sentiment towards "positive" digitalisation scenarios, as percentage deviation from mid-viewpoint



Notes: This is an experimental indicator. Survey respondents were asked to rate opposing scenarios on different dimensions from (1 = fully agree with a negative view) to (10 = fully agree with a positive view). For interpretability, weighted average scores on each dimension and the general summary view (weighted average across dimensions) are presented as percentage deviations from the midpoint. This means, for example, that with respect to the subject of "Science across borders", respondents are on average 50% oriented towards the positive outcome, relative to the neutral perspective. Weighted average scores take into account the sample design and non-response.
Source: OECD calculations based on OECD (n.d. c), *OECD International Survey of Scientific Authors 2018*, http://oe.cd/issa.

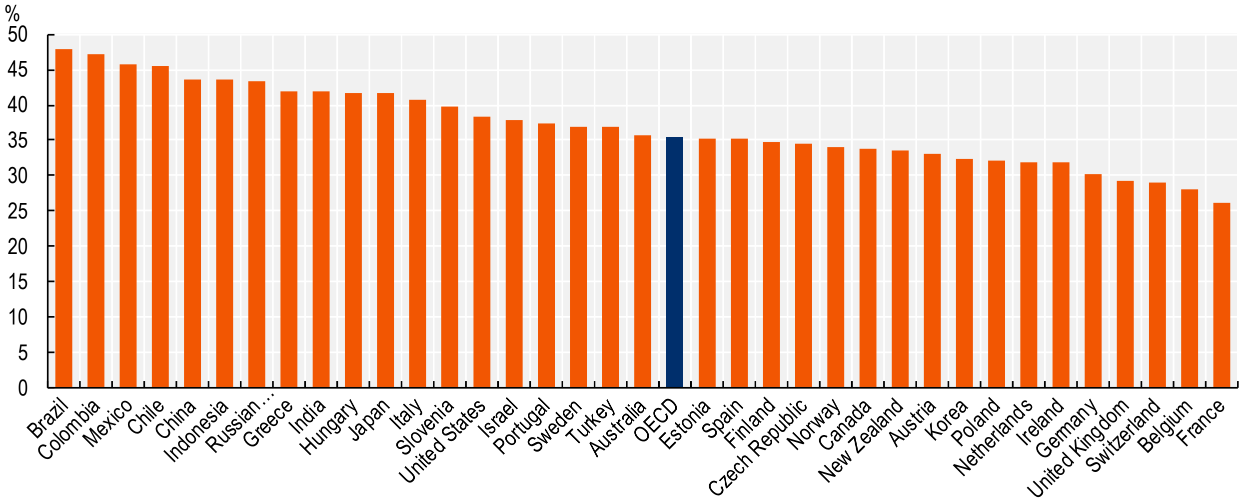
StatLink https://doi.org/10.1787/888934075963

Across countries, the average sentiment towards the impacts of digitalisation (Figure 2.16) seems consistent overall with results from broader population surveys on attitudes towards science and technology (OECD, 2015e). Scientists in emerging and transition economies appear to be more positive on average towards the impacts of digitalisation on science. The position of scientists in the most R&D-intensive European economies is more reserved, while still positive in the main. These results do not imply that scientists are by and large dismissive of the potential pitfalls of digitalisation. A minority, but still a significant number, of respondents tended to agree with "negative" statements about the impacts of digitalisation on science. They were concerned, for example, about the promotion hypothesis-free research in computationally intensive data-driven science. For these respondents, digitalisation could also accentuate divides in research between those with advanced digital competences and those without. It could also encourage a celebrity culture in science, premature diffusion of findings and individual exposure to pressure groups. Digitalisation could also lead to use of readily available but inappropriate indicators for monitoring and incentivising research.

Finally, they agreed with the statement that digitalisation could concentrate workflows and data in the hands of a few companies providing digital tools.

Figure 2.16. Scientists' views on the digitalisation of science, by country, 2018

Average sentiment towards a "positive" digitalisation scenario, as percentage deviation from the mid-range of possible views



Notes: This is an experimental indicator. Cross-country comparisons should be interpreted with caution as the population of corresponding scientific authors is not uniformly representative of their scientific community. Economies with less than 75 survey responses have been removed. Average scores are weighted and take into account the sample design and non-response. See notes for Figure 2.15.
Source: OECD calculations based on OECD (n.d. c), *OECD International Survey of Scientific Authors 2018*, http://oe.cd/issa.

StatLink https://doi.org/10.1787/888934075982

Technology and innovation going digital

Development of digital technologies

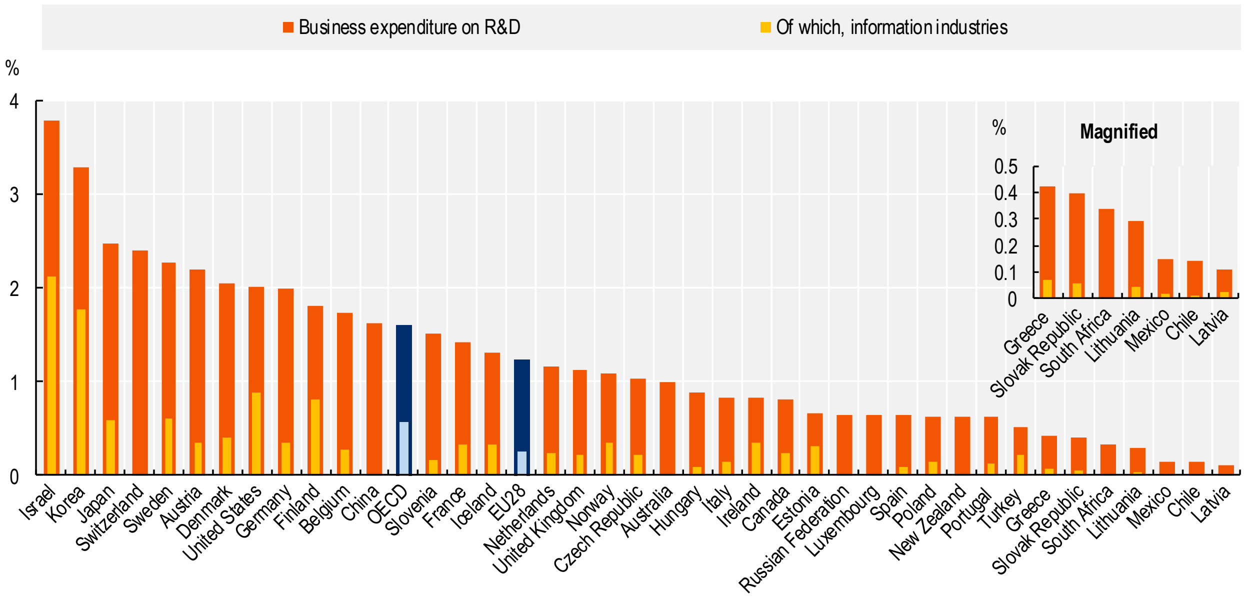
R&D in ICT industries and ICT-driven R&D

As an activity defined by the pursuit of new knowledge, R&D is important in driving advances in digital technologies. Businesses are the main source of R&D. Information industries are particularly strong contributors in countries with high business R&D intensity, accounting for just over half of all business R&D in some cases (Figure 2.17). Information industries also represent over 40% of business R&D in Estonia, Finland, Ireland, Turkey and the United States, confirming the knowledge-intensive nature of these industries.

Estimates of business R&D by industry fail to gauge perfectly the extent to which R&D contributes to digitalisation. This is important in the case of software because many firms invest in it for internal use and as a basis for providing other goods and services. By missing out software R&D (Box 2.5) in other sectors, the value of R&D in the software and information industries underestimates the total R&D aiming to generate new software. For instance, while software publishers in the United States account for 10% of all R&D performed and funded by companies, three times as much money was actually dedicated by US firms R&D aimed for software products or software embedded in other projects or products.

Figure 2.17. Business R&D expenditure, total and in information industries, 2016

As a percentage of GDP

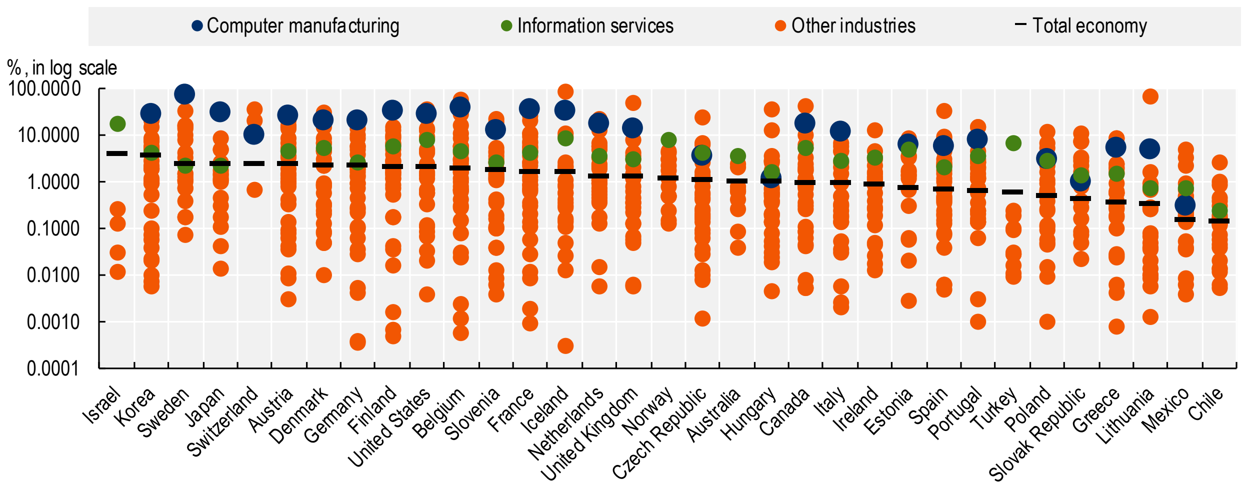


Notes: R&D = research and development; GDP = gross domestic product. "Information industries" are defined according to ISIC Rev.4 and cover ICT manufacturing under "Computer, electronic and optical products" (Division 26), and information services under "Publishing, audiovisual and broadcasting activities" (Divisions 58 to 60"), "Telecommunications" (Division 61) and "IT and other information services" (Divisions 62 to 63).
Source: OECD (2019a), *Measuring the Digital Transformation: A Roadmap for the Future*, https://doi.org/10.1787/9789264311992-en.

StatLink https://doi.org/10.1787/888934076001

Figure 2.18. R&D intensity of ICT and other industries, 2016

As a percentage of gross value added in each industry, log scale



Note: R&D = research and development; ICT = information and communication technology.
Source: OECD calculations based on OECD ANBERD, http://oe.cd/anberd, STAN, http://oe.cd/stan and National Accounts, https://stats.oecd.org/Index.aspx?DataSetCode=NAAG databases (accessed December 2018).

StatLink https://doi.org/10.1787/888934076020

Box 2.5. Software and R&D: A measurement challenge

Software development and R&D are closely intertwined (OECD, 2015a; OECD, 2015c; OECD/Eurostat, 2018). For example, the software industry is among the most R&D-intensive across most countries (Figure 2.18). Following revision of international guidelines in 1993, national accounts (NAs) economic statistics were comprehensively updated, as purchases of software and the own-account production of software were recognised as capital formation (i.e. “real” or “fixed” investment). Subsequent updates in NA systems and practices in many countries expanded this treatment to include firms’ own development of software originals used for reproduction.

The latest (2008) update of international guidelines introduced the classification of R&D as fixed investment. In so doing, it adopted the OECD definition of R&D and its measurement guidelines as the basis for primary data collection. Consequently, national accountants had to deal with the natural overlap between the development of own-account software originals and R&D activity. Own-account software originals were already included as investment in the NA measures of own-account software. Therefore, they were excluded in most cases from the new R&D measures to avoid double-counting in the NA aggregates. This treatment introduced a misalignment between the NA measures and the primary source data underlying the estimates of investment in R&D, produced by organisations that participate in the OECD’s NESTI. This misalignment could increase over time and potentially confuse users if software-generating R&D accounts for an increasing share of R&D. Some countries, such as the United States, are resolving this apparent inconsistency by reclassifying the own-account production of software originals that meets the R&D definition as R&D. It is unclear how other countries will resolve this challenge.

The growing importance of software development as an economic activity also presents a test-case for the measurement of R&D. The criteria provided in the 2015 edition of the *Frascati Manual* (OECD, 2015c) are a case in point. They allow organisations reporting and collecting data for statistical and other administrative purposes (such as the provision of R&D tax incentives) to discriminate between genuine R&D and non-R&D activities. R&D in software includes software development or improvement that expands scientific or technological knowledge, as well as the development of new theories and algorithms in computer science. In contrast, R&D activity in software excludes software development that fails to meet such requirements, e.g. work to support or adapt existing systems, add minor functionality to existing application programs, etc.

Use of digital technology in business and the link between digitalisation and innovation

Although the way in which innovation responds to and influences digitalisation can be mediated by R&D and invention, but it would be wrong to identify them as the same concept. The *Oslo Manual* definition of an innovation (OECD/Eurostat, 2018) refers to a new or improved product or process (or combination of both). It must differ significantly from a unit’s previous products and processes and be available to potential users or brought into use by the unit. Innovation requires that implementation take place: it must transcend the space of ideas and inventions. At a minimum, the innovation has to be new to the organisation. Thus, this is a broad concept that encompasses diffusion processes that involve a significant change from the viewpoint of who is adopting them. In this regard, various digitalisation processes across the economy are effectively innovations for those who implement them. Data from business innovation surveys show the information services industry generally exhibits the largest rates of reported innovation (e.g. 75% in the case of France). This may partly reflect higher rates of obsolescence, which call for more frequent innovation.

Digitally based innovations can be found in any sector. They comprise product or process innovations that incorporate ICT (the product itself can be a digital good or service). They also comprise innovations that rely to a significant degree on the use of ICT for their development or implementation. A wide range of business process innovations entail fundamental changes in the organisation's ICT function and its interaction with other business functions and the products delivered.

The latest edition of the Oslo Manual aims at ensuring that guidance fully reflects changes induced by digitalisation. For example, it recognises data development activities, along with software, as a potential innovation activity. Data accumulation by companies can entail significant direct or indirect costs. For example, a firm may give away for free, or at a discounted price, goods or services that generate information valuable for advertising products. The manual proposes to focus on developing measures of "digital competence". This multifaceted construct seeks to reflect a firm's ability to deal with digitalisation in a broad sense. Potential indicators, still to be harmonised ways in surveys, relate to:

- levels of digital integration within and across business functions
- access to an ability to use data analytics to design, develop, commercialise and improve products, including the ability to secure data about the (potential) users of the firm's products and how they interact with the products (Rindfleisch et al., 2017)
- access to networks and use of appropriate solutions and architectures
- capacity to manage privacy and cybersecurity risks
- adoption of appropriate business models for digital environments and platforms.

In addition to these internal capabilities, the manual recommends capturing, among the various external factors influencing innovation, information on the extent to which a firm uses digital platforms or is exposed to competition from them. Consumer and societal perspectives such as trust are also relevant to digitalisation. This measurement agenda requires close co-ordination with surveys on ICT use in firms. The latter are the responsibility of the OECD's Committee on Digital Economy Policy and its Working Party on the Measurement and Analysis of the Digital Economy.

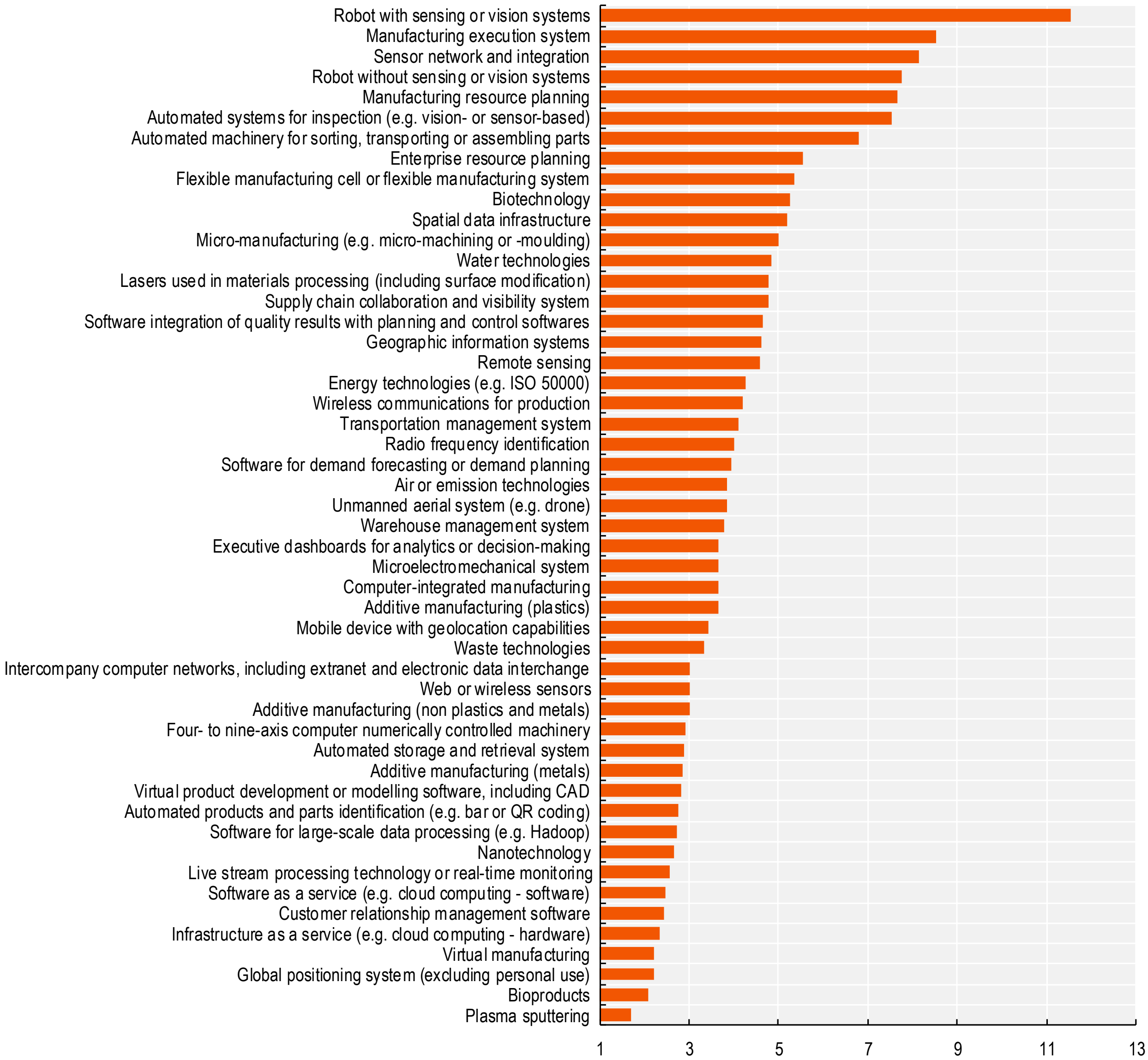
The OECD is highlighting country experiences in collecting data to motivate the collection and analysis of information at the junction between ICT adoption and innovation. It is also showcasing the data's potential relevance for international comparative analysis. One example is a recent study of patterns of advanced technology and business practices (ATBPs) among Canadian firms. This was conducted within the scope of Statistics Canada (STC)'s 2014 Survey of Advanced Technology. Joint OECD-STC analysis (Verger et al., forthcoming) has helped map ATBP portfolios via factor analysis. This has revealed seven main categories of ATBP specialisation: logistics software technologies; management practices and tools; automated production process technologies; geomatics and geospatial technologies; bio-and-environmental technologies; software and infrastructure as a service; and additive and micro manufacturing technologies. The data indicate a strong complementarity between management practices and production and adoption of logistics technologies.

As shown in Figure 2.19, Verger et al. (forthcoming) has found the rate of use of ATBPs to be generally positively correlated to the size of firms. This is especially so in the area of automated production process technologies, where scale appears to be important. However, software and infrastructure as a service (i.e. including cloud computing) is a noticeable exception; unlike technologies such as robotics, it is similarly diffused in small and medium-sized enterprises (SMEs) and large firms. This latter finding underlines one of the distinctive features of the digital economy: the attractiveness of such technologies for SMEs and their potential role in enabling scaling up.

Characterising industries by ATBP use patterns complements the standard classification systems for industries. Such systems are mainly informed by the type of goods and services delivered, rather than the processes used to produce them. The correlation between R&D intensity and technology is high in manufacturing industries and low in services. Most non-manufacturing sectors have low R&D intensity, even though many are technology-intensive. These findings confirm the limitations of using R&D measures for building technology taxonomies of industries that include services.

Figure 2.19. Advanced technology usage in Canada: Large firms vs. SMEs

Relative odds of using advanced technology for large firms vs. SMEs



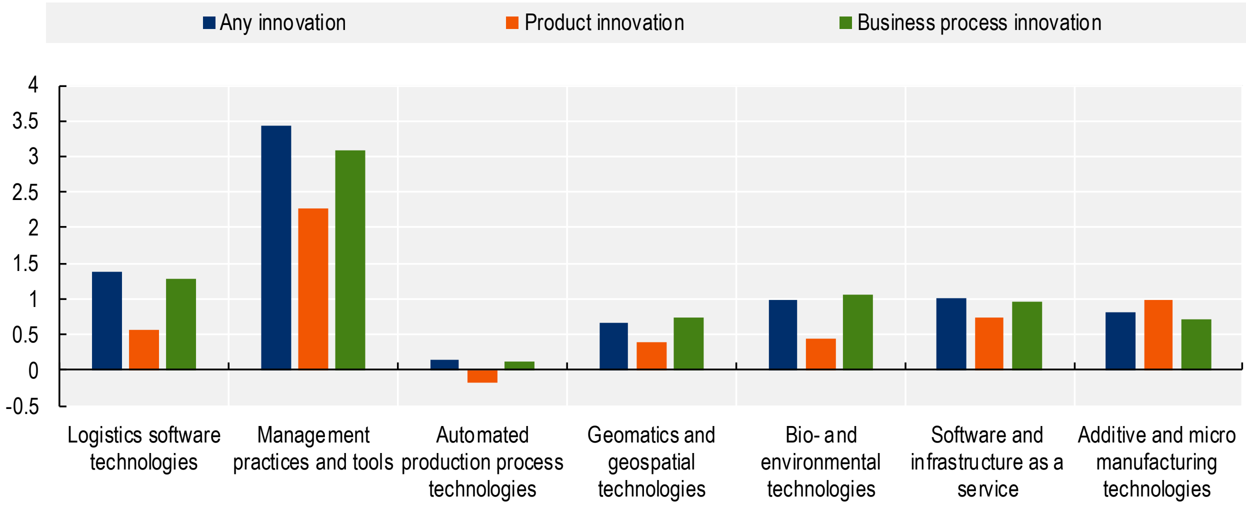
Notes: SMEs = small and medium-sized companies. How to read this chart: large companies are nearly 12 times more likely to use robots with sensing or vision systems than SMEs.
Source: Verger et al. (forthcoming), “Exploring patterns of advanced technology and business practice use and its link with innovation: An empirical case study based on Statistics Canada’s Survey of Advanced Technologies”.

StatLink https://doi.org/10.1787/888934076039

Lastly, the OECD-STC quantitative case study found that innovation is highly correlated with the use of certain business practices and advanced technologies (Figure 2.20). Regression results suggest that using advanced technologies doubles the odds of reporting innovations. The odds of innovating are trebled for users of selected business practices. The results also indicate complementarity between technology and management in explaining innovation. A positive relationship is also found between the *development* of technologies and innovation, especially for products, pointing at the advantages of being lead adopters.

Figure 2.20. The link between innovation and the adoption of technology and business practices, Canada, 2014

Estimated log odds ratios of reporting an innovation between technology and/or practice users and non-users



Note: Estimates control for technology development activity, country of ultimate ownership control, and business size and industry.
Source: Verger et al. (forthcoming), "Exploring patterns of advanced technology and business practice use and its link with innovation: An empirical case study based on Statistics Canada's Survey of Advanced Technologies".

StatLink https://doi.org/10.1787/888934076058

This analysis suggests that aspects of the SAT survey can be adopted more widely. With relevant adaptations, they could help assess the combined role of innovation, technology and management in business performance. A key challenge is to build consensus on which technologies and practices should be the focus of innovation surveys. Another challenge is how to implement approaches that compare data across countries, industries and longitudinally (given rapid technological change and obsolescence). At present, there is strong demand for specific analysis of the role of AI in business innovation strategies and activities.

Conclusion

Digitalisation is everywhere in STI, but with varying depth and perspective

Ministers from OECD countries and partners at the OECD Ministerial Meeting in Daejeon (Korea) in 2015 recognised that the rapid evolution of digital technologies is revolutionising STI (OECD, 2015d). These technologies, it was noted, are changing the way in which scientists work, collaborate and publish. They are also increasing reliance on access to scientific data and publications, and opening new avenues for public engagement and participation in science and innovation. At the same time, they are facilitating the development of research co-operation between businesses and the public sector, and contributing to the transformation of innovation.

At the time, the OECD was asked to monitor this transformation. It was also invited to convene the international community working on STI data and indicators to develop new thinking and solutions for generating empirical evidence to guide policy. The 2016 OECD Blue Sky Forum (http://oe.cd/blue-sky) identified the digitalisation of STI both as a priority object of measurement and as a fundamental enabler of future statistical and analytical work (OECD, 2018b). This principle guided the OECD's work on the digitalisation of science and innovation in 2017-18. This chapter summarises the main results of this work. It presents selected and new evidence arising from recent work on measuring digitalisation in science, its potential drivers and impacts. The indicators presented also raise further questions.

The evidence presented has put a focus on the potential synergies and trade-offs faced by those in decision-making roles in the science and innovation system:

- The geography of scientific activity in computer science and AI, measured by publications, has rapidly shifted. Formerly emerging economies like China have increased the quantity and quality of their publications, as implied by their citation impact.
- Research on AI is increasingly embedded in government agencies' funding of R&D across different missions and disciplinary areas. The example based on two major US funding agencies should soon be extended to other agencies and countries. However, this requires a concerted effort to maintain high-quality project information and to make it available for research policy purposes.
- Research careers in the area of computer science in the OECD area open a broad range of opportunities within and outside academia, but fail to attract a significant share of women. Research careers in this area are more inclusive of individuals born or raised abroad, pointing to the importance of policies that influence the mobility of talent and consider changes in demand for skills.
- Digital activity in science is highly pervasive, but there is considerable room for different disciplines to more fully exploit the potential of digitalisation. This is particularly true in the use of advanced tools that can transform the established research paradigms. Furthermore, high digital intensity is associated with many of the third mission activities that policy makers wish to encourage, such as creation of start-ups and societal engagement. There is some evidence of a generational and gender gap in the adoption of the most demanding digital practices.
- By and large, scientists appear to be optimistic about the possibilities brought to the practice of science by digitalisation, especially the youngest. However, many among the latter harbour more reservations about implications for their own careers.
- The adoption of advanced digital technologies appears to be highly correlated with the adoption of complementary business practices; this is closely associated with higher reported innovation. There is also evidence that firm size is a strong determinant of advanced technology adoption. However, among a representative sample of Canadian firms and after accounting for other characteristics, smaller firms are almost as likely to use cloud computing technologies as larger firms. It remains to be seen if this finding can be replicated in other contexts.
- This chapter calls for enhanced measures of organisational competences linked to digital technologies that influence firms' ability to innovate in the current landscape in which platforms play a major role. The ongoing revision of innovation surveys to adapt to the guidance of the newly published latest edition of the *Oslo Manual* is a significant opportunity for countries to reconsider how best to generate insights in this area.

More targeted measurement is required to address specific policy questions. These include how digitalisation can fundamentally expand the range of hypotheses generated and the speed at which competing research hypothesis can be tested. This could help address concerns about declining research productivity, public trust in science, lack of diversity and community engagement. It could also inform policy making so as to avoid a potential misalignment between career incentives and socially beneficial research.

Questions about the role of digitalisation also provide a much-needed stimulus to measure key dimensions of science and innovation once considered too complicated, or even unnecessary, to measure. Understanding how science adopts technologies and organisational practices can ultimately help explain how it can influence the direction of technical change and innovation more broadly.

Digitalisation is a "game-changer" for STI measurement and analysis

Digitalisation represents a major force for change in the generation and use of STI data and statistics. STI systems have become remarkably data-rich: information on innovation inputs and outputs that was only recorded in highly scattered, paper-based sources is now much easier to retrieve, process and analyse

(OECD, 2018b). When researchers and administrators use digital tools, they leave traces that can be used to develop new databases and apply to indicators and analysis. The digitalisation of the patent application and scientific publication processes has already provided rich and widely used data resources for statistical analysis. Digitalisation is rapidly extending to other types of administrative and corporate data, e.g. transactions (billing and payroll data); website content and use metadata; and generic and specialised social media, in which STI actors interact with their peers and society. Data practitioners have viewed these new "big data" as "uncomfortable data", i.e. datasets that are too large to be handled by conventional tools and techniques. But even these uncomfortable data are now more tractable.

The increasingly fuzzy boundary between qualitative and quantitative data is a striking example of how big data is becoming easier to manage. Many information gathering methodologies (e.g. user testing or interviews), for example, were traditionally considered as purely qualitative. However, they can now be conducted on a large scale and results quantified. For example, text, images, sound and video can all be "read" by machines. Natural language-processing tools automate the processing of text data from thousands of survey responses or social media posts into quantifiable data. These techniques can help alleviate some of the common challenges facing STI statistics, such as survey fatigue and unfit-for-purpose classification systems applied differently by human coders. Subsequently, they generate adaptable indicators.

Effective application of these new methods relies ultimately on fit-for-purpose, high-quality systems. These systems need to collect qualitative information consistently and avoid potential manipulation by parties with an interest in the use of the data. Administrative database managers become important gatekeepers of data quality, but information providers still need adequate incentives. Big data implies risks in exploiting datasets with possible defects and biases not recognised by the researchers. It also implies difficulties in evaluating big data techniques and analysis, especially using conventional criteria (such as falsifiability). And it implies complexities in explaining these techniques – and their value as evidence for policy evaluation – to decision makers and the public. In this new environment, work is moving progressively from fixed scales of analysis (such as the nation) towards variable categories, and dealing with vast new databases. This requires a different way of searching for patterns, trends, correlations and narratives.

The changing landscape for surveys has provoked much debate about their future. Some question whether the shift to big data is the precursor to the demise of surveys. Others, paraphrasing Mark Twain, argue that reports of the death of surveys are greatly exaggerated. The manner in which surveys are carried out has indeed changed, as online surveys have largely displaced more expensive non-digital methods. Surveys can therefore be targeted towards areas where other data sources are less effective (Callegaro and Yang, 2018; Jarmin, 2019).

Electronic tools (including do-it-yourself platforms) have "democratised" the process of surveying, making it easier than ever before. This has resulted in an explosion of surveys both in general and in the area of STI studies. However, these surveys often fail to meet basic statistical quality requirements, including for safeguarding privacy and confidentiality. The rapid growth of surveys also represents a growing source of fatigue for respondents; it results in lower expected response rates to non-compulsory (and compulsory, but difficult-to-enforce) surveys and may undermine trust. STI policy makers should co-ordinate, and apply standards to, their sponsored surveys.

New data sources for STI, such as administrative records, commercial databases and the Internet, have considerable transformational power stemming from their multidimensionality, and the possibilities for interconnecting the different types of subjects and objects covered. The strengths of these data sources are hard to reproduce in surveys. Traditionally, surveys were conceived to identify key actors and the presence of pre-defined types of interactions rather than to trace those linkages.

Digital solutions applied to survey tools that build trust and credibility can help bring out the potential of new data sources. Digital technologies are viewed as key components of the move towards "rich data". They are crucial to validating and augmenting the quality of big data sources (Callegaro and Yang, 2018). Rather than competing with alternative sources, surveys look set to focus increasingly on the crucial information that cannot be obtained otherwise. Recent experience shows that trust and credibility will be the most crucial factors determining the success of survey efforts in the digital era.

The experience of the OECD ISSA study confirms the importance, when conducting surveys in the digital age, of ensuring mutual trust between data collector and respondent. The ISSA survey ultimately explores how to develop working knowledge of emerging topics of high policy relevance. This can help provide a potential basis for distributed data collection within countries. It can also create a mechanism for ongoing dialogue between the OECD and the global science community.

STI policy makers need to support the creation and adoption of standards to protect the integrity of data they wish to use to inform their policies, regardless of the source. Furthermore, risk management will become an integral part of science and innovation policy in the digital era. Policy makers will need to consider how to make digitally driven systems, including those based on AI, more trustworthy. As a result, measurement will need to increasingly map risks and uncertainty, and analyse how these impact digitalisation practices and policies. This will be an important component of decisions assessing the merits of different science and innovation policy options in the digital era.

Digital innovation and AI in particular are indeed at the top of the national and international, as reflected in the adoption by the OECD Council of a Recommendation on AI (OECD, 2019b). The OECD council recommendation does explicitly state that governments "should consider long-term public investment, and encourage private investment, in research and development, including interdisciplinary efforts, to spur innovation in trustworthy AI […]". These policy priorities, with their often explicit demand for measurement and new evidence, will guide future OECD statistical and analysis work in this area.

References

Agrawal, A., J. Gans and A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, Cambridge, Massachusetts.

BEIS and DCMS (2018), "Artificial intelligence sector seal: A sector deal between government and the artificial intelligence (AI) sector", *Policy Paper,* Government of the United Kingdom, www.gov.uk/government/publications/artificial-intelligence-sector-deal.

Börner, K. et al. (2018), "Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy", *Proceedings of the National Academy of Sciences*, December, Vol. 115/50, United States National Academy of Sciences, Washington, DC, pp. 12630-12637, https://doi.org/10.1073/pnas.1804247115.

Breschi, S., J. Lassébie and C. Menon (2018), "A portrait of innovative start-ups across countries", *OECD Science, Technology and Industry Working Papers*, 2018/02, OECD Publishing, Paris, http://dx.doi.org/10.1787/f9ff02f4-en.

Boselli, B. and F. Galindo-Rueda (2016), "Drivers and implications of scientific open access publishing: findings from a pilot OECD International Survey of Scientific Authors", *OECD Science, Technology and Industry Policy Papers*, No. 33, OECD Publishing, Paris, https://doi.org/10.1787/5jlr2z70k0bx-en.

Bughin, J. et al. (2017), "Artificial intelligence: The next digital frontier?" *Discussion Paper*, June, McKinsey Global Institute, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

Callegaro, M. and Y. Yang (2018), "The role of surveys in the era of 'big data'", in D. Vannette and J. Krosnick (eds.), *The Palgrave Handbook of Survey Research*, Palgrave Macmillan, Cham, Switzerland, https://doi.org/10.1007/978-3-319-54395-6_23.

Cockburn, I., R. Henderson and S. Stern (2018), “The impact of artificial intelligence on innovation”, in A. Agrawal, J. Gans and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.

EC (2018), *Artificial Intelligence for Europe; Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions*; COM (2018) 237 final, European Commission, Brussels.

EDaily (2017), “Big Data, 5G, autonomous car, etc.”, www.edaily.co.kr/news/read?newsId=01685926616162128&mediaCodeNo=257.

Elsevier (2018), *Artificial Intelligence: How Knowledge is Created, Transferred, and Used*, Elsevier, Amsterdam, www.elsevier.com/research-intelligence/resource-library/ai-report.

Furman, J. (2016), “Is this time different? The opportunities and challenges of artificial intelligence,” presentation at AI now: The social and economic implications of artificial intelligence technologies in the near term, New York University, 7 July, https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160707_cea_ai_furman.pdf.

Fyfe, A. et al. (2017), “Untangling academic publishing: A history of the relationship between commercial interests, academic prestige and the circulation of research”, *Briefing Paper*, May, University of St. Andrews, United Kingdom, https://doi.org/10.5281/zenodo.546100.

Gold, E.R. et al. (2018), “An open toolkit for tracking open science partnership implementation and impact” [version 1; not peer-reviewed], Gates Open Research 2018, 2/54, https://doi.org/10.21955/gatesopenres.1114891.1.

Gold, E.R. (2016), “Accelerating translational research through open science: The neuro experiment”, *PLOS Biology,* Vol. 14/12, e2001259, https://doi.org/10.1371/journal.pbio.2001259.

Goldfarb, A. and C. Tucker (2017), “Digital Economics”, *NBER Working Paper* No. 23684, National Bureau of Economic Research, Cambridge, Massachusetts.

Jarmin, R.S. (2019), “Evolving measurement for an evolving economy: Thoughts on 21st century US economic statistics”, *Journal of Economic Perspectives,* Vol. 33/1, American Economic Association, Nashville, pp. 165-184, https://doi.org/10.1257/jep.33.1.165.

Klinger, J., J.C. Mateos-Garcia and K. Stathoulopoulos (2018), “Deep learning, deep change? Mapping the development of the artificial intelligence general purpose technology”, 17 August, SSRN, http://dx.doi.org/10.2139/ssrn.3233463.

McKinsey (2018), “AI adoption advances, but foundational barriers remain”, Survey, McKinsey Institute, www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain.

National Science Foundation (2018), *Doctorate Recipients from U.S. Universities: 2017*. Special Report NSF 19-301, National Center for Science and Engineering Statistics. Alexandria, VA, https://ncses.nsf.gov/pubs/nsf19301/.

NIH (2018), *NIH Strategic Plan for Data Science*, National Institutes of Health, Bethesda, United States, https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf.

NSTC (2016), *Preparing for the Future of Artificial Intelligence*, Executive Office of the President National Science and Technology Council Committee on Technology, Washington, DC, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

OECD (n.d. a), “Going Digital”, webpage, www.oecd.org/going-digital/ai (accessed 1 June 2019).

OECD (n.d. b), *Careers of Doctorate Holders database*, 2017 CDH-light data collection, http://oe.cd/cdh.

OECD (n.d. c), *OECD International Survey of Scientific Authors 2018*, http://oe.cd/issa.

OECD (2019a), *Measuring the Digital Transformation: A Roadmap for the Future,* OECD Publishing, Paris, https://doi.org/10.1787/9789264311992-en.

OECD (2019b), *Recommendation of the Council on Artificial Intelligence*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

OECD (2018a), *Education at a Glance: OECD Indicators,* OECD Publishing, Paris, https://doi.org/10.1787/eag-2018-en.

OECD (2018b), "Blue Sky perspectives towards the next generation of data and indicators on science and innovation", in *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-19-en.

OECD (2018c), "Private equity investment in artificial intelligence", *OECD Going Digital Policy Note*, OECD, Paris, www.oecd.org/going-digital/ai/private-equity-investment-in-artificial-intelligence.pdf.

OECD (2017), "Open access of scientific documents, 2017: As a percentage of a random sample of 100 000 documents published in 2016", in *Science, Technology and Industry Scoreboard 2017: Knowledge economies and the digital transformation*, OECD Publishing, Paris, https://doi.org/10.1787/sti_scoreboard-2017-graph66-en.

OECD (2015a), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264229358-en.

OECD (2015b), "Making open science a reality", *OECD Science, Technology and Industry Policy Papers,* No. 25, OECD Publishing, Paris, https://doi.org/10.1787/5jrs2f963zs1-en.

OECD (2015c), *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, https://doi.org/10.1787/9789264239012-en.

OECD (2015d), "Daejeon Declaration on Science, Technology, and Innovation Policies for the Global and Digital Age", webpage, www.oecd.org/sti/daejeon-declaration-2015.htm (accessed 1 June 2019).

OECD (2015e), "Public perceptions of science and technology", in *OECD Science, Technology and Industry Scoreboard 2015: Innovation for growth and society*, OECD Publishing, Paris, https://doi.org/10.1787/sti_scoreboard-2015-56-en.

OECD/Eurostat (2018), *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition,* The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg, https://doi.org/10.1787/9789264304604-en.

OSP (n.d.), Open Syllabus Project website, http://opensyllabusproject.org/ (accessed 1 June 2019).

Piwowar et al. (2018), "The state of OA: A large-scale analysis of the prevalence and impact of open access articles", *PeerJ*, 6:e4375, https://doi.org/10.7717/peerj.4375.

Rindfleisch, A et al. (2017), "The digital revolution, 3D printing and innovation as data", *Journal of Product Innovation Management*, Vol. 34/5, Wiley Online Library, pp. 681-690, https://doi.org/10.1111/jpim.12402.

Shoham, Y. et al. (2018), "The AI Index 2018 Annual Report", AI Index Steering Committee, Human-Centered AI Initiative, December, Stanford University, Stanford, United States.

Simonite, T. (2017), "Do we need a speedometer for artificial intelligence?", *WIRED*, 30 August, www.wired.com/story/do-we-need-a-speedometer-for-artificial-intelligence/.

Varian, H. (2019), "Artificial intelligence, economics, and industrial organization", in A.K. Agrawal et al. (eds.), *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Cambridge, United States.

Verger, F. et al. (forthcoming), "Exploring patterns of advanced technology and business practice use and its link with innovation: An empirical case study based on Statistics Canada's Survey of Advanced Technologies", *OECD Science, Technology and Industry Working Papers*, OECD Publishing, Paris.

WIPO (2019), *WIPO Technology Trends 2019: Artificial Intelligence*, World Intellectual Property Organization, Geneva, www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

Yamashita et al. (forthcoming), "Identifying and measuring developments in artificial intelligence", *OECD Science, Technology and Industry Working Papers*, OECD Publishing, Paris.

Notes

[1] In the application area of defence, the Defense Advanced Research Projects Agency is responsible for much of the research funding related to computer science in the United States. Project-level information is not readily available in this case. While AI is not separately identifiable, this agency's unclassified budget estimates for 2019 contain 21 references to AI research. The 2017 funding for R&D, testing and evaluation for the "Defense Research Sciences Program" alone includes USD 145 million for Mathematics and Computer Sciences and USD 46 million for Cyber Sciences.

[2] The OECD analysis of OA was based on a random sample of 100 000 documents drawn from articles, reviews and conference proceedings published in 2016, listed in the Scopus database and having digital object identifiers (DOIs). Assessment of the OA status of the documents was conducted in June 2017 using the R-language based "wrapper" routine for the oaDOI application program interface. It was produced by ImpactStory, an open-source website that works to help researchers explore and share the online impact of their research. The API returns information on the ability to secure legal copies of the relevant document for free and the different mechanisms available: *Gold OA journal*; *Gold hybrid*; *Green OA*. When the DOI cannot be resolved to any source of access information, the result is marked as "No information – status not available". This category is particularly high for China at more than 15%. When the DOI resolves and the return indicates there are no legal open versions available, the document is marked as "Closed". This includes documents under embargo. The oaDOI application and related "unpaywall" browser extension have since been further refined and developed. They now identify an additional category of publications, namely those that are free to read on the publisher's page, but without a clearly identifiable licence (labelled as "bronze"). Most of these documents went unnoticed in previous versions of the oaDOI application and were treated as closed. Piwowar et al. (2018) suggest the percentage of publications in this category is around 15% for the most recent publications with valid DOIs. This brings the percentage of functionally open documents closer in line with evidence in the ISSA2 study, which suggests about 65% of research documents published in 2017 can be freely accessed online one year later. This experience points to the usefulness for policy research purposes of APIs with metadata about scientific research, but it is also a stern reminder of the sensitiveness of results to the methods used.

3 Digital technology, the changing practice of science and implications for policy

Carthage Smith

Directorate for Science, Technology and Innovation

This chapter considers how digital technologies that have arisen out of publicly funded scientific research are now rapidly transforming the practice of research and enabling open science. This transformation is apparent across all of the three main pillars of open science: dissemination of scientific information, access to research data and engagement with stakeholders from outside of research. Recent developments and analysis are presented for each of these areas. This is followed by a discussion of what these developments mean for the governance of science as a whole, including for international co-ordination and co-operation. The chapter builds on earlier work by the OECD's Working Party on Innovation and Technology Policy and the report "Making open science a reality" and synthesises findings from recent work by the OECD Global Science Forum.

Introduction

Digital technologies are transforming science. Much discussion about the digital economy focuses on the dominant role of a small number of multinational companies. In this context it is easy to overlook the fact that public sector science is at the origin of the digital revolution and continues to play a critical role in shaping it. The World Wide Web was first developed at the European Laboratory for Particle Physics in Switzerland to meet the needs of particle physicists. Foundational work on the Internet was supported by the Defence Advanced Research Projects Agency and the National Science Foundation in public laboratories in the United States. Academic researchers are playing a key role in developing the next generation of digital technologies – from quantum computing to biological storage of data. At the same time, science itself is being radically transformed by the digital technologies it has helped create.

Digitalisation is affecting all stages of the scientific process – from agenda setting and experimentation to knowledge sharing and public engagement. In so doing, it is facilitating the transition towards a new paradigm of open science. The transformative and sometimes disruptive effects of digital technologies are apparent across all fields of science, but manifest differently in different communities. Scientific domains that have historically been data intensive and co-operative, such as particle physics or astronomy, face different challenges to much of medical research or social sciences that have been less data-centric. In contrast, these latter fields have a stronger history of societal engagement, which is also being transformed by digitalisation.

Open Science, in its broadest sense, is about making the scientific process more open and inclusive for all relevant actors. There are three main pillars: open access (OA) to scientific publications and information; enhanced access to research data; and broader engagement with stakeholders from within and beyond the scientific community. Strengthening these three pillars could increase the efficiency and effectiveness of science, and accelerate the translation of scientific findings into innovations and socio-economic benefits. Achieving this and realising the full benefits of open science, while minimising the associated risks, will require new policies and careful balancing of mandates and incentives. It will also require long-term strategic investment in digital infrastructure and skills.

This chapter considers the three pillars of open science: how digitalisation is changing established practices, the opportunities and challenges that this entails and what this means for policy. It then discusses the meaning of these developments for the governance of science as a whole, including for international co-ordination and co-operation. The chapter builds on earlier work by the OECD's Working Party on Innovation and Technology Policy and the report "Making open science a reality" (OECD, 2015). It synthesises some key findings from recent work done by the OECD Global Science Forum. This work includes an overall framework for open science (Dai, Shin and Smith, 2018) and specific policy reports relating to new forms of data and ethics (OECD, 2016), data repositories (OECD, 2017a, 2017b); agenda setting (OECD, 2017c); and access to research infrastructures (OECD, 2017d). These reports are complemented by insights from other recent OECD activities, including workshops, surveys and references to other relevant information.

Accessing scientific information

The results of scientific research have traditionally been published in specialist scientific journals, following review of submitted manuscripts by peers. The costs of managing the review process and journal production and distribution, have been recovered by charging readers (or academic libraries). Over time, a large and profitable industry has grown up around scientific publishing and many professional scientific societies have come to depend on publishing income to offset the costs of other services that they provide for their communities. As the scientific community has grown, the number of scientific journals has massively increased, together with the overall subscription charges to access these journals. Even within academia, only the better-endowed institutions, mainly in developed countries, have been able to keep up with this expansion. With the advent of the World Wide Web and online publishing, the marginal costs of disseminating scientific information have been reduced almost to zero, opening up new possibilities for more inclusive and broader access to scientific information. New OA publishing models have emerged (gold, green, hybrid etc.) and pre-print servers (such as arXiv.org in physics, or bioRxiv.org for biology), mega-journals (such as PLOS One), institutional repositories and online scientific information aggregators (such as PubMedCentral or LENS.org) are making access to scientific information easier and more inclusive. This transition to new science publishing models has raised concerns about the quality and sustainability of the scientific record. Ensuring these has been an important aspect of the added-value that commercial publishers – in partnership with scientific societies – have provided. This role has been integrated into their traditional business models. In the new OA publication era, it is less clear how editorial and peer-review processes will work and how the academic record will be maintained and updated. Estimates for the costs of publication vary considerably. Better information will be required to move away from a reader-pays market model to a high-quality and sustainable upstream, or author pays, model. It is notable in this regard that cOAlition S – a consortium of research funders – has identified lack of transparency on OA publication costs and fees as an obstacle to promoting OA publishing (Science Europe, n.d.).

As new actors enter the science publishing arena, there is considerable concern about the growth in predatory online journals. These journals charge authors for publication, but carry out little or no review and quality control. Their publications are contaminating the scientific record and can undermine public trust in science. An online catalogue of predatory journals created by the librarian Jeffrey Beall in 2008 became an important reference site for the scientific community. Since the catalogue ceased to be updated in 2017, its absence has been lamented and there have been a number of subsequent efforts to revive it (Weebly.com, n.d.). Predatory journals must be publicly identified, and researchers discouraged from seeking publication in them.

The sheer volume of scientific papers is overloading both legitimate journals and researchers who try to keep up with them. The growth in scientific papers is in keeping with the expansion of the global scientific community. It also partly reflects academic incentive systems and the "publish or perish" dynamic. Scientists have reached "peak reading" and too many research papers are of inadequate quality.[1] Even the most prestigious journals are having problems with quality assurance and the number of retractions is increasing. In some research fields, including life sciences and psychology, there is a reproducibility crisis, with many peer-reviewed and published findings being impossible to replicate. High-profile cases of scientific misconduct have become apparent in publications across all areas of science. Online forums such as Retraction Watch are helping the research community to identify questionable publications and the Committee on Publication Ethics is providing valuable guidelines to assist editors in dealing with these, but the numbers continue to increase (Brainard and You, 25 October 2018).

While digital tools cannot address the underlying causes of information overload and lack of scientific rigour, they can help manage these issues. Information and communication technology (ICT) can assist in organising, sharing and analysing large volumes of scientific information. Emerging tools and platforms enable researchers to rapidly identify and access papers that correspond to their interests (e.g. IRIS.AI., n.d.). Articles can be automatically "recommended" to scientists based on previous online search histories. Anti-plagiarism software, combined with data linkage systems such as Crossref, is helping editors and publishers with quality control. These tools, however, depend on the broad adoption of standards and unique digital identifiers, which can be supported at the policy level. For example, several research funders have joined publishers in mandating the use of Open Researcher and Contributor IDs (ORCIDs) for individual researchers.

Digitalisation is creating new possibilities for peer review, which remains at the core of the scientific publishing process. In some fields, including physics and astronomy, there is a tradition of making results available on line for open review and comment prior to formal publication. Pre-print archives and open peer review are now being tested in other fields, including life sciences (Cold Spring Harbor Laboratory, 2018). Looking to the future, this could be imagined as one part of a tiered process for publication, with more scientific information being shared earlier and commented on by the community and only a fraction of this eventually being formally published in journals. Some fields are also testing post-publication peer review, which can potentially help ensure the quality and rigour of the scientific record. Sites such as Pubpeer (Pubpeer Foundation, n.d.) are playing an important role in enabling the community to report and discuss concerns about published results. Technologies such as blockchain can potentially help ensure the fidelity of peer review, while accelerating the process and rewarding reviewers (Blockchain for Peer Review, 2019).

As indicated previously, the predominant model for communication of scientific information to date has been via the release of peer-reviewed publications at the end of the research process. However, much useful information that is generated during research, including negative results that may be important with regards to reproducibility, are never shared. While at one level scientists are over-loaded with information, at another level the information that can be readily accessed is often inadequate to critically evaluate, replicate and build on what is published. Again, digital technologies can help address this challenge. Online open lab notebooks, can provide access to the primary experimental data and information linked to publications and also help to ensure appropriate accreditation. The landmark publication of the detection of gravitational waves that led to a Nobel prize in Physics in 2017, for example, was accompanied by OA to the experimental records in a Jupyter notebook. The scientific article of the future may be more than just a narrative with summary results. It may also include direct links to all supporting data and a record of the process by which that data was generated and analysed (Schapira, 2018).

The publication of scientific articles in journals is intimately coupled to the evaluation and rewards systems for science. This means that changes to publication practices can directly affect scientific careers. This is critically important in the current transition period, when many science funders are mandating OA publication (Science Europe, n.d.) but promotion and tenure, and, in some cases, institutional funding, continue to be largely determined by publication in high-impact, pay-for-access, journals. Mandates need to be matched by incentives and changes to current evaluation systems if the transition to OA publication is to be accelerated. A stronger focus on article-based metrics rather than journal impact factors is one way to assist this transition.

Digitalisation also provides opportunities to communicate scientific results and information in different ways that can complement or even replace traditional scientific articles in journals. Not all scientific disciplines are equally dependent on scientific articles as their main means of communicating results. In some areas of social sciences books are the main output of academic work and in computing sciences, conference proceedings are the most important mechanism for sharing results. Again, digitalisation and online tools can increase access to these outputs.

The use of social media, such as Facebook and Twitter, has transformative potential across all fields of science. Already, science blogs (e.g. LSE, n.d.) are becoming essential information sources, and increasingly cited, in scientific articles. The publication of scientific papers is now frequently accompanied by tweets. Alternative metrics or “altmetrics” are being developed to measure the impact of traditional scientific publications via their uptake in social media networks. Such metrics can clearly provide interesting information. However, further experimentation is required to test what kind of impact they are really measuring and how their deployment in evaluations might affect scientific behaviour and trust in science.

Enhanced access to research data

Data that is used in research and/or generated by research is the lifeblood of the science enterprise. Some fields of science are facing a reproducibility crisis and OA to the data [and code] that provides the basis for published scientific results is important as it allows for verification of those same results. Secondary analysis of data and application of the same data in different research fields can provide new scientific insights. Greater access to data can help to make science more inclusive and productive by allowing new actors to engage in the scientific process. The integration of data from diverse sources is important for science to be able to address complex societal challenges. Research data can also be an important substrate for innovation and economic growth. This is particularly the case when data are combined with mathematical algorithms, models and high-performance computing.

The OECD first advocated for greater access to data from publicly funded research in 2006. Since then, both the rationale and the tools for enabling greater access have been strengthened considerably. The

OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007) laid out 13 overarching principles that have stood the test of time. More recently, with an added emphasis on open science, the essence of this earlier normative work has been distilled into four concise findability, accessibility, interoperability and reusability (FAIR) principles: research data should be Findable, Accessible, Interoperable and Re-usable. The FAIR principles have been widely adopted across countries and the focus is now on how they can be implemented at the operational level, where issues such as standards, security and protection of privacy need to be addressed. Funding, infrastructure and skills are also limiting factors. It is increasingly recognised that, as data volumes increase, the costs of stewarding this data become prohibitive, while, at the same time, much of the data probably has little secondary value. The mantra is moving towards making research data "as open as possible and as closed as necessary", as opposed to making all data open to everyone (OECD, 2018a).

Online data and associated services have dramatically changed many fields of science, from genomics to astronomy. The Global Earth Observation System that combines huge amounts of data from space, ocean and terrestrial observation devices, is essential for understanding the planet we live on and how it is changing. Social networking data is providing new insights into human behaviour and even the spread of disease (HealthMap, n.d.). Nevertheless, and despite broad agreement on the FAIR principles, a number of significant obstacles inhibit access to data. These include: i) costs and business models for data repositories; ii) trust and transnational barriers; iii) privacy and ethical considerations; iv) access to cyber-infrastructure and skills for data management and analysis; and v) incentives and rewards. The first three obstacles are considered in the paragraphs that follow, while the fourth and fifth are discussed at the end of this chapter.

Business models for data repositories

Research data repositories are the main focus for implementation of FAIR data principles. However, as data volumes and user demands expand, the costs of data management are straining research budgets. Recent analysis of almost 50 data repositories, across diverse areas of research, has identified key actions to improve long-term sustainability of these critical infrastructures (OECD, 2017e).

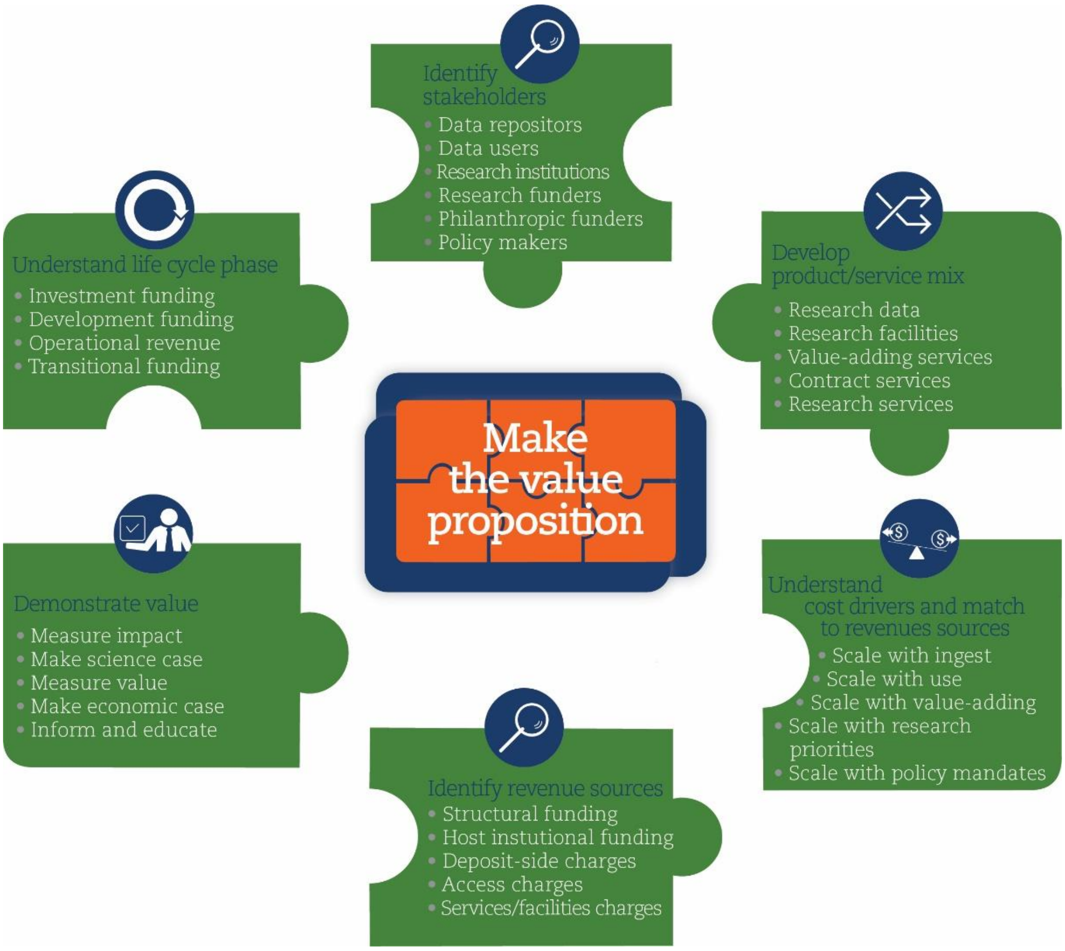
Hence, repositories need to be considered as an integral part of the infrastructure for research and they need to have clearly articulated business models (Figure 3.1). This, in turn, affects how they are funded and, in particular, how public funding is allocated. Many valuable data resources start out with short-term project funding but then struggle to be sustainable. Mandates for OA need to be matched with incentives, including appropriate funding. Opportunities for cost-optimisation, including scale effects and technological advances, need to be actively pursued. Where the commercial sector provides repositories and associated data services for research, it should be consistent with the aim of enhanced long-term access. Monopoly arrangements, which can have longer-term negative consequences, should be avoided.

Trust and transnational barriers

International data networks play an important role to assure data quality across borders. The sharing of research data across national borders is critical for many areas of science and, in most cases, this depends not just on single global data repositories but on federated international data networks. Examples of such networks include the multidisciplinary World Data System, the International Virtual Observatory Alliance (IVOA) in astronomy and the Inter-University Upper Atmosphere Global Observation Network. These networks can play an important role in data quality assurance, with membership being conditional on compliance with agreed standards and recognised accreditation systems (e.g. Data Seal of Approval, n.d.).

As is the case for individual repositories, these networks also need to have well-defined business models and value propositions. However, several additional challenges are associated with the establishment and maintenance of such networks (OECD, 2017b). The main barriers to sharing data across borders are the lack of policy coherence and trust between different communities.

Figure 3.1. Key elements of a business model for a data repository



Source: OECD (2017a), "Business models for sustainable research data repositories", https://doi.org/10.1787/302b12bb-en.

Despite the growing acceptance of the FAIR principles as an aspirational aim, at the operational level there is considerable discordance around what data should be available to who and how – there is an absence of commonly agreed legal and ethical frameworks for sharing different types of public research data. Although no one model fits all, a number of organisational issues need to be addressed for networks to operate effectively. These range from aligning different objectives and user needs to governance arrangements. Ensuring inclusiveness and respecting cultural differences and capacity limitations can be problematic. Cutting across all of this are issues related to the adequacy of funding, and there is a need for funders to participate in relevant international discussions and fora, such as the Research Data Alliance, to improve co-ordination of their strategies and support for data infrastructure.

While the technical issues should not be underestimated, establishing trust is perhaps the main obstacle to enhancing data access and implementing the FAIR principles. This applies both from the perspective of the data provider and the user (Box 3.1). In recent OECD work on sharing scientific data and information during crises, it was striking that lack of trust was identified as the major obstacle to cross-sectoral and transnational co-operation (OECD, 2018b). There are a number of policy actions that can be taken to address issues of trust. Some of these relate to technology, such as blockchain, or the adoption of standards and processes, e.g. the use of safe havens for working on sensitive data. However, trust is fundamentally a sociological issue and building trust requires dialogue and shared understanding.

Box 3.1. Barriers to data sharing: The researcher perspective

A number of governments and research funding agencies are beginning to mandate increased sharing and/or OA to research data. However, four key issues impede data sharing by researchers. Each of these is amenable to policy interventions:[1]

- **Trust**. This obstacle refers to the mutual mistrust between scientists (Do I trust the data? Will I receive credit for my data if someone else uses them? Will my data be used appropriately?). In the case of personal data, the need to ensure trust between human subjects/patients and users is also important. Where commercial sector users are involved, then issues of trust can be further amplified.
  - Policy options: Put in place processes for data tracking and citation; adopt trusted repository accreditation systems and support international data networks; strengthen ethics committees by including data experts; and organise public dialogues on personal data and privacy and develop consensus around key issues, such as consent, anonymisation and commercial use.
  - Good practices: Data Seal of Approval for repository certification; establishment of the Ada Lovelace Institute in the United Kingdom to ensure data and artificial intelligence (AI) work for people and society (Nuffield Foundation, n.d..
- **Burden**. This obstacle relates to the time, expertise, and resources required by providers to make their data available and the time invested by users to discover available data.
  - Policy options: Develop national strategic plans, including long-term funding plans, for sustainable research data infrastructure (data repositories and services); require data management plans and provide funding to implement them in association with grant awards; provide dedicated funding to develop new data services; and identify and address data skills gaps in the research workforce.
  - Good practices: European Open Science Cloud Strategic Implementation Roadmap; Australian National Data Service Skills support services (ANDS, 2018).
- **Motivation, credit and reward**. There is little incentive for scientists to make their data openly available. While publication of research results is critical for career advancement, there is little reward for developing and sharing useful data resources.
  - Policy options: Develop new indicators/measures for data sharing and incorporate these into institutional assessments and individual researcher evaluation processes; promote the use of unique digital identifiers for individual researchers and for data sets to enable citation and accreditation; and develop attractive career paths for data professionals, who are necessary for the long-term stewardship of research data and provision of services.
  - Good practices: Open Research Funders Group work on incentivising the sharing of research outputs through research assessment (ORFG, n.d.).
- **Governance and legal frameworks**. A lack of understanding and clear guidance on data privacy regulations can inhibit data sharing by scientists. Likewise, in the absence of clear guidelines and relevant expertise, institutional review boards (IRBs) may act as a barrier to data sharing.
  - Policy options: Identify and support trusted brokers to mediate access to data; support the development of standardised data management plans and data use agreements; where appropriate, involve lay persons/patients in governance and oversight structures; encourage citizen science projects; and ensure that IRBs include the necessary expertise in data science, including legal aspects.
  - Good practices: National Services Scotland Save Haven for secure access to health service data (ISD Scotland, 2018); Science Europe initiative for the development of domain-specific data management protocols (Science Europe, 2018).

1. This was identified at an INCF-OECD workshop on data sharing in dementia research, Stockholm, September 2015.

Data privacy and ethics

Trust is a particularly important issue with regards to access to personal data. New forms of personal data are becoming available in digital format from many sources, ranging from supermarket transactions to social media. These have enormous potential value for research, particularly when combined with administrative data, public health records or more traditional population survey data. Such data combinations can provide important new understanding into human behaviour, economic systems and the social determinants of health and well-being (OECD, 2013).

The rapid advance of technology is raising ethical questions about the use of personal data that go beyond the scope of existing legal agreements. Several legal frameworks, most notably the General Data Protection Regulation (GDPR) in Europe, provide some guidance and establish agreed limits on the use of personal data. However, the technology is developing so fast that new possibilities for data use in research raise ethical dilemmas that transcend these frameworks. Something can be legal without being ethically acceptable. Indeed, this is implicit in the GDPR's provision for ex ante Data Protection Impact Assessments (DPIAs) when a process involving data is likely to pose a high risk to people's rights and freedoms.

Increasing concern about data privacy and security has created an urgent need for science to adapt its governance and review mechanisms (OECD, 2016). The previously accepted requirements for the use of human subject data in research were informed consent and anonymisation. However, both of these are now being questioned as a consequence of advances in ICTs. For example, is it possible to get informed consent for specific purposes from all the individuals in very large sets of social media data? Can personal data from one source be truly anonymised when its linkage to other personal data is required for research?

There is a critical role for institutional review boards and/or research ethics committees to ensure oversight of what research is being conducted with new forms of personal data – or in the language of the GDPR – to carry out DPIAs. These bodies need to be empowered, supported and have the expertise necessary to assess the balance between protecting personal privacy and ensuring the public good. Social consensus will need to be established both within and beyond the scientific community as to what the appropriate limits are on the use of new forms of data.

It is difficult to foresee everything and, along the way, mistakes will certainly be made. Transparency and accountability will be critical to building a consensus on the use of new forms of data. Policy makers play an important role in ensuring the right governance frameworks are in place and supporting the necessary consultation and consensus building processes. In the United Kingdom, for example, public consultation has been important in establishing the core policies and operations of the Administrative Data Research Network (Verwulgen, 2017).

Broader engagement in science

Broader engagement in science is the third pillar of open science. Digitalisation is opening the scientific process to a variety of societal actors, including patient groups, citizen scientists, non-governmental organisations, industry and policy makers. This shift has considerable potential to improve the quality, relevance and translation into practice of scientific research. Societal engagement can take place across the research process – from agenda setting to co-production of research and dissemination of scientific information. Depending on the emphasis, societal engagement encompasses concepts such as responsible research and innovation and transdisciplinary research. Engagement, which depends on access to scientific information and data, is being transformed by the use of digital tools.

Many countries promote citizen engagement to help ensure research relevance and promote transparency and trust in science. If science is to provide solutions for pressing societal challenges, then arguably it needs to be more closely engaged with society. In this context, digitalisation is providing powerful new tools to assist with societal engagement.

The first, and perhaps most critical, step in citizen engagement is to frame research agendas and set priorities for research investment. Recent OECD work has focused on this, including an in-depth analysis of key features and lessons learned from a variety of open agenda-setting exercises for research (OECD, 2017c). These exercises ranged from broadly focused citizen consultations and dialogues to inform international and national agenda setting through to more local and community-specific co-design processes.

OECD (2017c) identified ten key issues for consideration in designing effective open agenda-setting processes. These begin with clear articulation of the rationale for a consultation; selection of an appropriate methodological approach; and consideration of resource implications and impact assessment. If these three areas are addressed, then open agenda setting can make research more relevant and may also generate new research questions. A case in point is the Great New Zealand Science Project, a national campaign to define research priorities. Citizens expressed the need for more research on care issues as opposed to drug development for the elderly. There is a substantial body of work and tested methods for citizen engagement (Engage2020, n.d.; PARTICIPEDIA, n.d.). As the interest in open agenda setting expands, these previous experiences can provide valuable lessons.

Research infrastructures (RIs) provide a variety of shared services to the research community in all fields of science. Digitalisation is changing the operations of these infrastructures in many ways. Several of these RIs are at the forefront of the big data revolution, including the development of related hardware, software and standards. RIs are at the centre of many issues relating to open science – from information and data management to data security, privacy protection, analysis and training, and citizen science. Indeed, the growth of data and the policy emphasis on FAIR data are putting the financial sustainability of many RIs at risk (OECD, 2017e).

At a more mundane level, RI managers, funders and potential users are faced with a simple and persistent challenge: identifying what RIs exist, what they can do and how they can be accessed. Scientists are likely familiar with the main RIs used routinely in their own field. However, access to facilities and resources in other research areas is increasingly required. Other potential RI users – from companies, the public sector or civil society – can find it difficult to explore the possibilities of RIs related to their interests. Optimising the use of RIs depends first on accessing systematically collected, up-to-date information. This is where digitalisation potentially provides a solution.

Recently published work (OECD, 2017d) included an in-depth analysis of eight initiatives that are using dedicated digital platforms to promote broader access to, and more effective use of, RIs. These platforms ranged from digital catalogues providing standardised metadata on the resources available in a specific scientific domain, to national and regional service platforms enabling virtual access or online reservations of facilities.

Greater co-operation across borders on definitions, standards and interoperability of digital platforms is needed to provide sustainable high-quality service for users. The OECD work identified seven areas requiring attention. Of these, the most important is the need for international co-operation around definitions, standards and interoperability. Different countries and institutions are developing ad hoc solutions to meet their own specific data aggregation needs. However, there is limited long-term planning and co-operation with other actors. The Mapping of European Research Infrastructure Landscape initiative, for example, consulted broadly with the community to develop a set of definitions, glossaries and RI classifications together with a metadata model. These have all been made openly available for other users, but take up has been limited.

RIs, such as telescopes, can provide a focus for citizen science, i.e. the engagement of people who are not professional scientists in research processes. In the field of astronomy, for example, lay persons are helping to classify images of the night sky that are shared on line (Zooniverse, n.d.. More broadly, many fields are promoting citizen science as a way of both addressing unique issues and of promoting public trust in science. Digitalisation is rapidly changing what is feasible, enabling new approaches to crowdsourcing and access to untapped intellectual resources to solve problems (OECD, 2015; Dai, Shin and Smith, 2018).

Beyond data collection and analysis, ICT can also help engage the networked public in novel forms of discovery. For instance, in 2011, players of an online protein-folding game – Foldit – outperformed scientists by discovering the structure of a protein involved in the Mason-Pfizer monkey virus. This discovery was facilitated by complex software that permitted visualisation of protein shapes, allowing the employment of shape recognition and modification skills by persons not necessarily trained in biochemistry (University of Washington, 2012).

At the more applied level, many companies are using online crowd sourcing platforms, such as InnoCentive, to help solve technological challenges, with significant prizes being awarded to problem solvers. Hackathons, that bring together interested actors, on line, are a common way of addressing software development challenges and are increasingly being organised in association with traditional scientific congresses. Through Kaggle, which is owned by Google, data scientists and users get together on line to find solutions to problems presented by research teams and private companies.

Opening up science to engage new actors from civil society raises new issues in terms of preserving quality, ensuring proper attribution and ethics (Bonney et al., 2014). Engaging the right audience and promoting effective participation can be a particular challenge, especially when dealing with issues that are value laden. From the policy perspective, defining where citizen science approaches might be most valuable in specific contexts and how best to achieve this will require careful consideration (Box 3.2).

Box 3.2. Policy challenges for citizen science

Citizen science is a relatively recent, diverse and evolving approach to research. It presents both opportunities and challenges. Among other issues, more needs to be known about the following:

- The quality of scientific output. Concerns exist that valid scientific methods are not followed in some projects managed by non-scientists and that quality control through peer review is often absent.
- The types of science project for which citizen science might best be used. Not all research lends itself to citizen participation, which can significantly increase (or decrease) the overall cost of projects.
- The trade-off between participant anonymity and the opportunity to earn peer recognition through publication.
- The financial implications of crowdsourcing science. In particular, might financial incentives be used to attract firms and participants with specialised talent? Who owns the outputs if they have potential commercial value?
- How the efficiency of citizen science might be improved. For instance, software might be used to track participant performance in some tasks, possibly avoiding the need for other participants to replicate these tasks.

Promoting and steering open science systems in the digital world

There are many actors – institutions and individuals – with different roles and responsibilities in the scientific enterprise. These actors also often have different, and even contradictory expectations, for science. For instance, a dean in a top-ranked research university may be primarily interested in high-impact publications accredited to his/her institution. Conversely, the ministry that provides research funding may be more interested in open data for innovation. Moreover, in the digital world, distance and location matter less than access to data and information. This increases the emphasis on international collaboration and/or competition and presents new challenges for the governance of science as whole. While digitalisation could make science both more inclusive and more productive, in the transition from the old to the new a number of important policy issues need to be addressed. These cut across the whole of science and are manifest at different scales, from local to global as discussed in the following paragraphs.

Policy makers can play an important role in promoting the development and implementation of frameworks, common definitions and standards. As ICTs develop and open new possibilities, it is becoming clear that formal legal frameworks, IP regulations and standard-setting processes are lagging. Commercial actors, and, in some instances, specific research communities are establishing de facto standards for operating in the digital world. These are determining how information, data and technologies are used. In the best cases, community standards are adopted that ensure interoperability and openness. IVOA, for example, established community standards that have enabled researchers and interested citizens across the world to use astronomical data. In other cases, standards may reflect specific interests and severely limit access to and usage of scientific data and information. Likewise, ownership and licensing arrangements for digital information and data can either promote openness or limit access and reuse. With regards to text and data mining, for example, several countries have recently revised their copyright regulations to limit restrictions for research. Maintaining an optimal balance between protecting IP (which can promote innovation) and openness (to improve the efficiency and effectiveness of research) is an ongoing policy challenge. It is interesting that in some areas of medical research, public and private sector actors are building new open science partnerships in which OA and sharing of data, information and downstream technologies are the norm (e.g. SGC, 2019).

Ensuring the provenance and traceability of scientific data and information are important with regards to quality assurance, accountability and accreditation. Digital traces of individuals, research groups, institutions and their scientific outputs are becoming an essential part of the evaluation and impact assessment processes for science (see Chapter 7). These depend on the use of open digital object identifiers (DOIs), including ORCIDs for researchers. Policy makers can play an important role in promoting the routine use of such DOIs.

Mandates and incentives are valuable policy tools to promote open science, provided that they are used carefully. Mandates and incentives are often most effective when used in tandem (OECD, 2015). This is illustrated by the recent launch of Plan S (Science Europe, n.d. b) that aims for full and immediate OA to publications from publicly funded research. All recipients of research funding from any of the coalition partners promoting Plan S will be mandated to publish in compliant journals. At the same time, funders are working together on new incentives for open research. Plan S, for example, refers to the San Francisco Statement on Research Assessment (DORA, n.d.), which states that research needs to be assessed on its own merits rather than on the basis of the venue in which it is published. The proposals for Plan S include a transition period during which journals are expected to become compliant. However, it is clear that reward and recognition systems for research will need to value OA publishing if this is to be broadly adopted by the academic community.

Similarly, for data sharing, new indicators and measures will be required not only to monitor how mandates for enhancing access are being implemented but also to incentivise activities to implement FAIR data (Ali-Khan, 2018). And societal engagement activities will require similar incentivisation. It is notable that the latest development of the Research Excellence Framework for the evaluation of UK higher education institutions puts increasing emphasis on scientific outputs other than journal publications (REF2021, n.d.).

Digitalisation is transforming science very rapidly and this raises issues with regards to the skills that are required by the current and future scientific workforce. Digital skills are high on the education agenda in all OECD countries and, from an economic perspective, having an appropriately trained population is considered to be one of the key determinants of future productivity and growth (OECD, 2017f). From a research policy perspective the key questions are: what are the additional or specific digital skills that are, or will, be required for data intensive science? How do these skills need map onto the scientific workforce? How will the necessary skills be provided and what does this mean for science education and training?

"Data scientist" is a generic term that encompasses many different skills and roles (see Table 3.1 and also Chapter 2) and although the needs vary from one field to another there is a general consensus that more data scientists are required in public research. It has been suggested that for the European Open Science Cloud to be effective 500 000 data experts will need to be trained over the next five to ten years. The difficulty of meeting such a shortfall is compounded by a number of factors. It is not clear exactly what the needs are, while, at the same time, a plethora of new education and training courses for digital skills for science are appearing. Looking from the opposite perspective, it is not known whether these new educational and training courses are adequately addressing the real needs and gaps. Data science, in its different manifestations (Table 3.1), often does not align well with existing academic credit and reward systems that depend on publication outputs as opposed to code and data products. New career structures and professions will need to be developed, e.g. for data stewards. Moreover, there is intense competition from the commercial sector for digitally skilled individuals who, in "hot" areas such as AI, can earn salaries well above what is offered in academia. A strategic approach, that takes all these various factors into account, is required and this should consider how public and private actors can work together to develop human capital in ways that are mutually beneficial.

Table 3.1. Digital science personnel and roles

Data scientist	A data scientist is a practitioner of data science. It is a generic term that encompasses many fields of specialised expertise.
Data analyst	This is someone who knows statistics. Analysts may know programming or may be expert in Excel. Either way, they can build models based on low-level data. Most importantly, they know which questions to ask of the data.
Data engineer	Operating at a level close to the data, data engineers write the code that handles data and moves them around. They may have some machine-learning background.
Data steward	A data steward is a person responsible for the management of data objects including metadata. These people think about managing and preserving data. They are information specialists, archivists, librarians and compliance officers.
Research software engineer	A growing number of people in academia combine expertise in programming with an intricate understanding of research. Research software engineers may start as researchers who spend time developing software to progress their research. They may also come from a more conventional software development background and are drawn to research by the challenge of using software to further research.

Note: Depending on the field of research some of these roles may be combined in a given individual. They may be supporting or service provision roles or fully embedded in research projects.
Source: This categorisation and the definitions are derived from the ongoing work of an OECD-GSF Expert Group on Digital Skills for Data Intensive Science and more detailed glossaries for digital science (CASRAI, n.d., Research Data Domain website, https://dictionary.casrai.org/Category:Research_Data_Domain; Science Europe, n.d. a, "Science Europe Data Glossary Main Page", http://sedataglossary.shoutwiki.com/wiki/Main_Page).

There is a need for long-term strategic planning and effective co-operation across countries and continents. Many countries are making significant investments in the digital infrastructure necessary to support science. This includes data repositories, as well as cyber-infrastructure such as high-performance or cloud computing. They are also investing heavily in "next-generation" technologies such as quantum computing. Within Europe, there is a major initiative to integrate these national initiatives within the European Open Science Cloud. Similar developments are taking place in the United States, and on a smaller scale, in Africa and other regions. These initiatives include both public and commercial sector service providers, e.g. for data storage and computing, and it will be important going forward to ensure their long-term sustainability and adaptability and avoid the "lock in" that can arise when effective monopolies develop. Global bodies, such as the Research Data Alliance (RDA, n.d.), that bring together data scientists and policy makers to develop community standards, technical fixes and social networks have an important role to play.

Building trust both within the scientific community and between science and society remains the most pressing and difficult challenge for science in the digital world. With regards to the use of personal data in science, the challenges are fairly well known. Solutions are being developed and tested, including new mechanisms of governance and engagement with the public. However, there is a more ubiquitous challenge for science as a whole that will require more complex multifactorial solutions. While open science holds great promise, it arrives at a time when trust in experts is being questioned and “alternative facts” are becoming common currency in social networks and political forums. Open science means more transparency and accountability, but it also means more scrutiny and more questioning from actors for whom access to science was previously restricted. As witnessed in relation to debates on climate change or the safety of vaccination, some groups will readily appropriate, distort or re-interpret scientific information and data to their own ends. It is critically important in today’s more open environment that the integrity of science itself is maintained, and that science is rigorous and published research results are reproducible. Digital technologies such as blockchain and AI can potentially assist in this quest (see Chapter 1). However, appropriately skilled research personnel and the right incentives and reward structures will be even more critical.

Conclusion

By enabling a new paradigm of open science, digitalisation is disrupting long-established scientific practices, norms and institutions. Recent work from the OECD and many other organisations has demonstrated that digitalisation, which has its origins in public research, is also having a huge impact on how this research is being conducted. This is opening up exciting new opportunities and, at the same time, throwing up new challenges.

As with any disruptive change, different actors resist some of the emerging directions. Commercial publishers have served the scientific community well over many decades. They are understandably reluctant to change their business models. Scientists have built careers around “ownership” of data collections and are reluctant to share. Universities are used to being assessed on the number and/or quality of publications rather than data outputs or citizen engagement. Academic peer review, evaluation and promotion processes have been similarly focused on research excellence. Career paths have been designed for researchers in traditional scientific disciplines; they are poorly adapted to the new inter- and transdisciplinary opportunities of the digital world or the need to attract highly skilled data scientists in research support roles. Research funders are used to funding large-scale RIs over the long term. However, their mechanisms are less well adapted to the multitude of distributed data resources and services upon which research increasingly depends.

Science systems as a whole are having to adapt rapidly and this inevitably entails a mix of some things that are completely new, adjustments in much of what already exists and renewal of what cannot or does not adapt. Of course, science continually evolves and so this scenario is not unique. However, the extent and depth of the impacts of digitalisation on science and the speed of change are likely beyond what science systems have experienced since World War II. Strategic planning, flexibility and careful development and implementation of policies will be necessary to ensure that we build on the best of the past in taking forward the future. There is also an opportunity that should be grasped to address and correct some of the emerging problems in science, including lack of reproducibility, lack of diversity in academia and precarity in research careers. This transition is occurring at a critical moment when trust in science needs to be assured. Achieving this will require vision, policy action and joint commitment from multiple stakeholders with an interest in the scientific enterprise. It needs to be both top-down (policy driven) and bottom-up (community led).

Different fields of science and different organisations and countries are at different stages of adaptation to open science in the digital world. This provides important opportunities for mutual learning, exchange of good practices and co-operation. Scientific policy makers need to crowdsource within their own community, engaging other relevant actors as necessary, to identify existing or new solutions to support the positive evolution of science in the digital age.

References

Ali-Khan, S.E. et al. (2018), "Defining success in open science [version 2; referees: 2 approved]", *MNI Open Research*, Vol. 2/2, https://doi.org/10.12688/mniopenres.12780.2.

ANDS (2018), *ANDS/Working with Data – SKILLS*, Australian National Data Service, www.ands.org.au/working-with-data/skills (accessed 27 May 2019).

Begley, C.G and L.M. Ellis (2012), "Raise standards for preclinical cancer research", *Nature*, Vol. 483, Nature Research, Springer, pp. 531-533, www.nature.com/nature/journal/v483/n7391/full/483531a.html.

Blockchain for Peer Review (n.d.), "Towards a fairer and more transparent peer review process", webpage, www.blockchainpeerreview.org/ (accessed 27 May 2019).

Bonney, R. et al. (2014), "Next steps for citizen science", *Science*, Vol. 343/6178, American Association for the Advancement of Science, Washington, DC, pp. 1436-1437, http://dx.doi.org/10.1126/science.1251554.

Brainard, G. and J. You (25 October 2018), "What a massive database of retracted papers reveals about science publishing's 'death penalty'", Science News blog, www.sciencemag.org/news/2018/10/what-massive-database-retracted-papers-reveals-about-science-publishing-s-death-penalty.

CASRAI (n.d.), Research Data Domain website, https://dictionary.casrai.org/Category:Research_Data_Domain (accessed 27 May 2019).

Cold Spring Harbor Laboratory (n.d.), BioRxiv: The Preprint Server for Biology website, www.biorxiv.org/ (accessed 27 May 2019).

Dai, Q., E. Shin and C. Smith (2018), "Open and inclusive collaboration in science: A framework", *OECD Science, Technology and Industry Working Papers*, No. 2018/07, OECD Publishing, Paris, https://doi.org/10.1787/2dbff737-en.

Data Seal of Approval (n.d.), Data Seal of Approval website, www.datasealofapproval.org/en/ (accessed 27 May 2019).

DORA (n.d.), Improving How Research is Assessed website, https://sfdora.org/ (accessed 27 May 2019).

Engage2020 (n.d.), Action Catalogue website, http://actioncatalogue.eu/ (accessed 27 May 2019).

HealthMap (n.d.), HealthMap website, www.healthmap.org/en/ (accessed 27 May 2019).

IRIS.AI (n.d.), "Research discovery with artificial intelligence", https://iris.ai/ (accessed 27 May 2019).

ISD Scotland (2018), *Use of the NSS National Safe Haven*, www.isdscotland.org/Products-and-Services/EDRIS/Use-of-the-National-Safe-Haven/ (accessed 27 May 2019).

LSE (n.d.), "LSE blogs expert analysis and debate from LSE website", webpage, http://blogs.lse.ac.uk/ (accessed 27 May 2019).

Noorden, R.V. (5 February 2014), "Scientists may be reaching a peak in reading habits", Nature News blog, www.nature.com/news/scientists-may-be-reaching-a-peak-in-reading-habits-1.14658.

Nuffield Foundation (n.d.), "The Ada Lovelace Institute", webpage, www.nuffieldfoundation.org/ada-lovelace-institute (accessed 27 May 2019).

OECD (2018a), "Enhanced access to publicly funded data for STI", in *Science, Technology and Innovation Outlook 2018*, OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-en.

OECD (2018b), *Scientific Advice During Crises: Facilitating Transnational Co-operation and Exchange of Information*, OECD Publishing, Paris, https://doi.org/10.1787/9789264304413-en.

OECD (2017a), "Business models for sustainable research data repositories", *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, https://doi.org/10.1787/302b12bb-en.

OECD (2017b), "Co-ordination and support of international research data networks", *OECD Science, Technology and Industry Policy Papers*, No. 51, OECD Publishing, Paris, https://doi.org/10.1787/e92fa89e-en.

OECD (2017c), "Open research agenda setting", *OECD Science, Technology and Industry Policy Papers*, No. 50, OECD Publishing, Paris, https://doi.org/10.1787/74edb6a8-en.

OECD (2017d), "Digital platforms for facilitating access to research infrastructures", *OECD Science, Technology and Industry Policy Papers*, No. 49, OECD Publishing, Paris, https://doi.org/10.1787/8288d208-en.

OECD (2017e), "Strengthening the effectiveness and sustainability of international research infrastructures", *OECD Science, Technology and Industry Policy Papers*, No. 48, OECD Publishing, Paris, https://doi.org/10.1787/fa11a0e0-en.

OECD (2017f), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, https://doi.org/10.1787/9789264276284-en.

OECD (2016), "Research ethics and new forms of data for social and economic research", *OECD Science, Technology and Industry Policy Papers*, No. 34, OECD Publishing, Paris, https://doi.org/10.1787/5jln7vnpxs32-en.

OECD (2015), "Making open science a reality", *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris, https://doi.org/10.1787/5jrs2f963zs1-en.

OECD (2013), "New data for understanding the human condition: International perspectives", OECD, Paris, www.oecd.org/sti/inno/new-data-for-understanding-the-human-condition.pdf.

OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, https://doi.org/10.1787/9789264034020-en-fr.

ORFG (n.d.), Open Research Funders Group website, www.orfg.org/ (accessed 27 May 2019).

PARTICIPEDIA (n.d.), PARTICIPEDIA website, https://participedia.net/ (accessed 27 May 2019).

Pubpeer Foundation (n.d.), Pubpeer – The Online Journal Club website, https://pubpeer.com/static/about (accessed 27 May 2019).

RDA (n.d.), Research Data Alliance website, www.rd-alliance.org/ (accessed 27 May 2019).

REF2021 (n.d.), Research Excellence Framework website, www.ref.ac.uk/ (accessed 27 May 2019).

Schapira, M. (2018), "Open lab notebooks to increase impact and accelerate discovery", 26 January, Springer Nature, Data Dialogues, https://researchdata.springernature.com/users/81403-matthieu-schapira/posts/29655-open-lab-notebooks-to-increase-impact-and-accelerate-discovery.

Science Europe (n.d. a), "Science Europe Data Glossary Main Page", webpage, http://sedataglossary.shoutwiki.com/wiki/Main_Page (accessed 27 May 2019).

Science Europe (n.d. b), "Why plan S: Open access is fundamental to the scientific enterprise," webpage, www.coalition-s.org/why-plan-s/ (accessed 27 May 2019).

Science Europe (2018), "Practical guide to the international alignment of research data management", https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/.

SGC (n.d.), Structural Genomics Consortium website, www.thesgc.org/ (accessed 27 May 2019).

University of Washington (2012), "Gamers succeed where scientists fail: Molecular structure of retrovirus enzyme solved, doors open to new AIDS drug design", *ScienceDaily*, 19 September, www.sciencedaily.com/releases/2011/09/110918144955.htm.

Verwulgen, I. (2017), "The ADRN and the public's voice: Making administrative data available for research while gaining public trust", *International Journal of Population Data Science*, Vol. 1/1, p. 155, Swansea University, United Kingdom, https://doi.org/10.23889/ijpds.v1i1.174.

Weebly.com (n.d.), Beall's List of Predatory Journals and Publishers website, https://beallslist.weebly.com/ (accessed 27 May 2019).

Zooniverse (n.d.), Zooniverse website, www.zooniverse.org/ (accessed 27 May 2019).

Note

[1] In 2016, more than 1.2 million new papers were published in the biomedical sciences alone, bringing the total number of peer-reviewed biomedical papers to over 26 million. However, the average scientist reads only about 250 papers a year (Noorden, 5 February 2014). By some measures the quality of scientific literature has been in decline. Some recent studies have found that most biomedical papers were not reproducible (Begley and Ellis, 2012).

4 Digital innovation: Cross-sectoral dynamics and policy implications

Dominique Guellec, Caroline Paunov and Sandra Planes-Satorra

Directorate for Science, Technology and Innovation, OECD

With a focus on the agri-food, automotive and transportation, and retail sectors, Chapter 4 explores the impacts of digital transformation on innovation and identifies sector-specific dynamics. In view of such impacts, the chapter evaluates how innovation policies should adapt to promote vibrant and inclusive innovation ecosystems effectively. Examples of novel innovation policy approaches implemented in various countries are provided. The chapter also synthesises key findings from the OECD's Working Party on Innovation and Technology Policy and, specifically its Digital and Open Innovation project. It explores in detail the changes needed in innovation policy in the digital age, considering the impacts of the digital transformation on innovation across sectors.

Introduction

Digital transformation is a multifaceted phenomenon that is impacting innovation in all sectors of the economy. New digital technologies, including artificial intelligence (AI), have enabled the creation of completely new digital products and services and the enhancement of traditional ones with digital features. Production processes are also subject to substantial change, with new modes of human-to-machine interaction (see Chapter 5). New opportunities are emerging across innovation processes – from research (e.g. the use of big data analytics, large-scale computerised experiments), to development (e.g. new techniques of simulation and prototyping) and commercialisation (e.g. use of marketplace platforms).

Despite the dramatic changes, the impacts of digital transformation on innovation in specific sectors are largely unknown or anecdotal. Since industries significantly differ in their products and processes, their structures and in how they innovate, the impacts of digitalisation on innovation are also likely to differ. For instance, end products in primary sectors such as food or mining remain largely unchanged. Conversely, the media, music and gaming industries, to name a few, have completely digitised their product and service offering. Another example is the wide deployment of robots in the automotive industry, while automation remains at early stages in sectors such as agriculture and retail. There is, however, little systematic evidence about the sector-specific impacts of digital transformation on innovation. Understanding these differences matters for policy aimed at supporting innovation systems, because countries' industry composition differs markedly.

How is the digital transformation changing the innovation practices of firms?

This section explores how digital transformation is changing the nature of business innovation. Digital technologies have lowered information-related production costs and increased the "fluidity" of innovative products. Digitised knowledge (i.e. knowledge that takes the form of data) and information can circulate and be reproduced, shared or manipulated instantaneously by any number of actors regardless of their location. As a consequence of changes in costs and fluidity, four trends affect innovation practices across all sectors of the economy in the digital age (Figure 4.1).

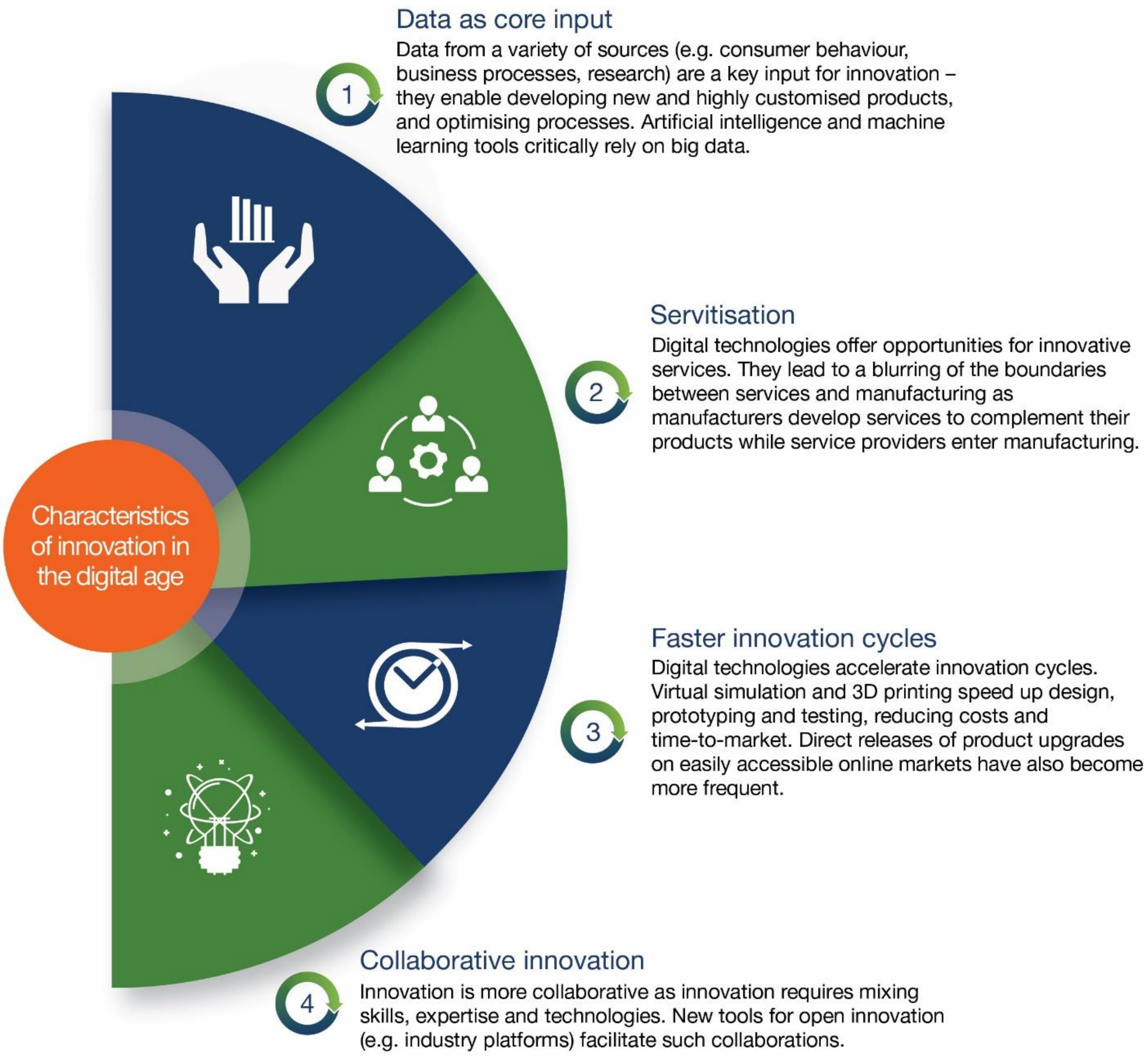
Data are a key input for innovation

Data are increasingly used in innovation processes. They help explore new areas of product and service development. They help gain critical insights about market trends, consumer demand and the behaviour of competitors. And they optimise development, production and distribution processes; tailor the product and service offering to specific demands; and rapidly adjust to changes in demand. The emergence of smart and connected products, as a result of increased sensing, connectivity and data embedded in products, significantly contributes to the generation of new data.

Data have allowed the development of completely new services and business models. These have been enabled by the availability, and capacity to exploit, large amounts of real-time data. Examples include smart farming services, peer-to-peer accommodation services (e.g. Airbnb), on-demand mobility services (e.g. Uber), and platforms to search, compare and book accommodation and transportation options (e.g. Booking).

Business data are increasingly used to optimise processes within firms, but also within supply chains. Manufacturing sectors exploit abundant real-time shop-floor data to identify patterns and relationships among discrete processes. This allows manufacturers to optimise data by reducing waste, saving energy, increasing flexibility and better using assets, among other areas (OECD, 2017). For example, UPS, a multinational logistics company, uses a fleet management system enhanced by data analytics. It allows for route optimisation, increasing the efficiency and flexibility of delivery processes and reducing fuel consumption. Data are also used to predict the needs of production systems, significantly lowering maintenance costs compared to unplanned maintenance and repair. In agriculture, data from a multiplicity of sensors can help farmers optimise use of water and other inputs to boost yields.

Figure 4.1. Characteristics of innovation in the digital age



Source: OECD (2019), *Digital Innovation: Seizing Policy Opportunities*, https://doi.org/10.1787/a298dc87-en.

Services innovation enabled by digital technologies

Digital technologies offer opportunities for the creation of entirely new digitally enabled services. Predictive maintenance services, for example, use the Internet of Things (IoT), which involves the deployment of sensors and actuators connected to software systems. Other emerging services include on-demand transportation (e.g. Uber); and web-based businesses. New digital technologies have also propelled expansion of the sharing economy and greater customisation. Renting-as-a-service models can replace selling of equipment, for example, while businesses can harness software and data to adapt products to customers' specific needs.

Such changes also contribute to a blurring of boundaries between manufacturing and services innovation. On the one hand, manufacturing firms increasingly offer innovative services to complement goods – a process known as "servitisation" of manufacturing. For instance, John Deere, an agriculture machinery producer, has developed a software platform that provides farm-management support services based on sensor data. On the other hand, service providers increasingly invest in digital technology to improve their activities. For instance, big retailers invest intensively in data collection and analytics capabilities (e.g. to personalise promotions and predict consumer trends), augmented and virtual reality (VR) (e.g. to develop digital fitting rooms) and the IoT (e.g. to improve inventory management).

Innovation cycles are accelerating

Digital innovations (such as three-dimensional [3D] printing and increasingly sophisticated simulations) introduce new and rapid innovation cycles by, among other routes, accelerating the processes of product design, prototyping and testing. Engineers and designers across manufacturing industries increasingly use "digital twins" (i.e. a 3D VR version of a production process or a product) to experiment with designs.

New technologies also stimulate market launch of testing (beta) versions that are regularly updated to incorporate consumer feedback. This is common practice for software launches. Many firms are also adopting a "lean start-up" method, which consists of creating minimum viable products that can be brought to market. Once launched, producers collect feedback from users and integrate it into the next development round. For example, GE Appliances' FastWorks system, based on lean innovation principles, involves consumers early in the development of new products such as refrigerators (General Electric, 2017).

Innovation is becoming more collaborative

Innovation ecosystems are becoming more open and diverse. Firms interact with research institutions and firms due to three reasons. First, they gain access and exposure to a richer pool of expertise and skills complementary to their own competences (e.g. data analytics). Second, such collaborations allow sharing of the costs and risks of uncertain investments in digital innovation. Third, reduced costs of communication allow greater interaction among actors engaged in innovation (e.g. firms, public research institutions [PRIs]), regardless of their location.

Collaborations take different forms, including the following:

- **Data sharing**. The non-rivalrous nature of data allows various actors from different organisations to use the same database simultaneously, even if they are located around the world. This has stimulated firms to share their data for research and innovation purposes, often with universities and research organisations, or trusted business partners. Challenges and hackathons are other popular tools for sourcing external ideas to foster data-driven innovation.
- **Partnerships**. Partnerships with large technology firms, digital start-ups and PRIs are becoming more common in the digital age. Their goal is to join efforts to foster joint value creation, expand market potential and combine strengths. In so doing, they allow the closing of skills or competence gaps. Collaborations with digital start-ups, in particular, have also boomed in recent years. Such collaborations are seen as "digital accelerators", with the flexibility needed to develop new disruptive technologies (Lund, Manyika and Robinson, 2016).
- **Platforms**. Industry platforms are products, services or technologies created by one or several firms. They provide the foundation upon which different actors can innovate by developing complementary products, services or technologies using digital tools (Gawer and Cusumano, 2014). These platforms can thus serve as the effective industry standard. They make development processes more efficient and less costly, and reduce time-to-market for new products. An example is the SmartDeviceLink Consortium, an open-source platform for smartphone app development for vehicles, created by Ford and Toyota. Firms also use crowdsourcing platforms to source ideas from outside the organisation (either the general public or a pool of accredited experts). In so doing, they aim to solve a specific problem or challenge, including finding new product or design ideas. Such initiatives are often conducted through intermediary platforms, such as InnoCentive.
- **Acquisitions**. The acquisition of innovative firms (particularly start-ups) by established firms is also a channel for collective innovation. Start-ups play a role in discovering and testing new markets and business models. When successful, they can be acquired by larger firms with access to capital and marketing channels that can help to scale a successful product.

The impacts of the digital transformation on innovation across sectors

How are digital technologies integrating different sectors?

Digital technologies are integrating and transforming sectors in different ways. This section explores transformations in the agriculture, automotive and retail sectors, which are considered representative of primary, secondary and tertiary sectors more generally.

Agri-food sector

In agriculture, intelligent and digitally connected machinery (IoT) enables the development of "precision farming". This allows systems that help farmers improve the accuracy of operations. Such systems can also optimise the use of inputs (e.g. water, fertilisers, pesticides) to give each plant (or animal) exactly what it needs to grow optimally. Tractors and other agricultural machinery are equipped with a large number of sensors that capture information related to crops (e.g. soil conditions, irrigation, air quality, presence of pests). Drones equipped with sensors are also increasingly used for crop scouting and spraying. Data captured by *in situ* sensors, drones and satellites allows better monitoring of crop health, assessment of soil quality and optimisation of input use, thus having positive effects on productivity.

The introduction of robots is another trend in farming. Fruit-picking, harvesting and milking are examples of the repetitive and standardised tasks performed by agricultural robots. Robots also generate data that can be exploited for different purposes. For instance, Lely Industry, a manufacturer of milking robots, collects data from robots to exploit information regarding the feed, animal health and milk quality of individual cows (Lely, 2016). Although agri-robots are generally in early stages of development, they are expected to increase efficiency and allow for more automated and precise agricultural practices.

Large agriculture machinery producers and input suppliers, such as John Deere, are using large amounts of data collected through the IoT from farm applications and robots. They combine them with other data such as on the weather or markets to develop "smart farming" services. These use big data analytics and AI to inform farm-management decision making (Wolfert et al., 2017). Such systems can help farmers decide when to plant or harvest, to choose the type of crop to plant depending on soil conditions and market prices, and to automatically instruct agricultural robots to perform certain tasks. Precision and smart farming are still mainly restricted to large producers. Smaller producers are less likely to adopt precision farming technologies due to the costs of investment, and of learning how to use them and adapt production processes.

The agri-food supply chain is starting to use the IoT to trace the origins and track the whereabouts of products, as well as their transportation and storage conditions. In this way, it improves transparency in the value chain. Blockchain and other distributed ledger technologies are also expected to offer opportunities for increasing the traceability of food products from harvest to point of sale. Major food companies are collaborating with IBM to apply blockchain to make food supply chains more transparent and traceable and to streamline payments (Tripoli and Schmidhuber, 2018).

Automotive industry

In the automotive sector, rapid developments in digital technology are completely reshaping the industry. These include vehicle innovations (e.g. car connectivity, autonomous driving), innovations in production (with smart factories or Industry 4.0 applications) and new business models (with the provision of after-sales services and expansion into on-demand mobility services).

Digital technology has given rise to connected cars that generate data from the physical world, receive and process data, and connect to other cars and devices. Connected cars allow for enhanced driver safety and convenience. New services include automatic emergency calls after an accident and real-time road hazard warnings for drivers, car repair diagnostics and systems of time-saving networked parking. In addition, navigation systems optimise route planning by considering real-time traffic conditions.

Developments in autonomous driving are being propelled by advances in the fields of robotics, AI, machine learning (ML) and connectivity. There are five different levels of automation – from driver assistance to complete automation. All new car models offer driving assistance systems. These take over parts of the vehicle motion control and support the driver with certain tasks such as parking and speed-keeping – but the driver is still in charge of driving. From a technical viewpoint, technology for highly automated driving in controlled environments is quite mature (VDA, 2015). At full driving automation, cars drive independently and react to their environment without intervention of the driver. Such systems are being tested in pilot projects (PSC/CAR, 2017), but opinions differ greatly on when full autonomy might be achieved.

The automotive industry is also a leader in developing "smart factories". It is adopting a variety of Industry 4.0 applications, including Internet-connected robotics, data analytics, and cloud and high-performance computing (HPC), among others. For instance, Hirotec, a Japanese auto parts manufacturer, uses ML and data analytics to predict and prevent failures. This drastically reduces the cost of unplanned downtime (Hewlett-Packard, 2017). BMW has set the goal of knowing the real-time status of all important machines producing components from all their suppliers using IoT applications (Ezell, 2018). Kern and Wolff (2019) provide other examples of investments by carmakers and automotive suppliers to foster efficiency, and automate production and supply-chain processes.

Firms in the automotive industry are also providing new services related to their products. Three areas of focus are the provision of new after-sales services (e.g. predictive maintenance); the development of alternatives to car ownership (e.g. vehicle subscription services); and expansion into on-demand mobility services (e.g. creation of own car-sharing brands).

The retail sector

In the field of retail, digital innovations aim at enhancing the consumer experience (both in physical and online shopping) and optimising processes (e.g. logistics, warehouse management). The largest investments focus on data collection (e.g. purchasing and browsing data) and data analytics capabilities. Such data provide insights on consumer needs and preferences that are used to customise the shopping experience, for instance by sending personalised advertisements and promotions. For example, Sephora uses data from customers' online shopping histories by employing beacons in their stores. These beacons send smartphone notifications when customers near an item they had previously added in a digital shopping cart (Pandolph, 2017).

Innovations in physical stores are expanding. Smart dressing rooms, for instance, might recommend specific items of clothing. Digital mirrors can allow customers to try on and compare several outfits, among other things. And automatic payment systems allow customers to skip check-out lines. AmazonGo, for example, recently established a cashier-free store in Seattle. By deploying sensors, cameras and other digital technologies, the store allows for automatic payment of products that customers take off the shelf, without the need to scan bar codes (Amazon, n.d.). Innovations in online retail include applications for designing or personalising products (e.g. shoes) through 3D visualisations. The automatic reordering of products may also become more common. The Amazon Dash Replenishment Service, for example, allows connected devices (e.g. washing machines, coffee machines) to reorder products automatically (e.g. laundry detergent, coffee beans) when supplies are running low. However, all of these innovations remain marginal and are mainly deployed by large retailers.

The retail sector is using the IoT and robotics to better manage inventories (e.g. in warehouses) and optimise other processes. AI is also opening avenues for predictive analytics to strengthen forecasting and improve stock management. For example, Otto, a German online retailer, uses consumer data and a deep learning algorithm to predict what customers will buy a week before they order. The algorithm, which has 90% accuracy, has led Otto to introduce an innovative stock management system that automatically purchases products from third-party brands (The Economist/Capgemini, n.d.).

Why are the implications of digital transformation likely to differ across sectors?

The implications of the digital transformation likely differ across (and within) sectors due to a range of factors that can be grouped along three dimensions described below.

Opportunities for innovation using digital technology

Future technology developments are inherently uncertain. Yet given the important variety in the nature of a sector's products and processes, some sectors will likely be disrupted to different extents by specific digital technologies (e.g. AI, IoT, drones, VR, 3D printing). Similarly, the transformation will probably take different forms and develop at different speeds. Depending on sectoral characteristics, digital technologies offer different opportunities for the following:

- **Digitalising final products and services**. While some industries have completely digitised their products over past decades (e.g. the media, music and gaming industries), others remain mainly physical, such as food and consumer products. Many industries present a mix of digital and physical components in their final products, with the digital ones often becoming progressively more important. In the automotive industry, vehicles increasingly integrate digital features. Advanced infotainment systems and other functionalities enabled by connectivity and data analytics, for example, are becoming key considerations in consumer purchasing decisions.
- **Digitalising business processes**. The extent to which digitalisation affects sectors' business processes may differ. It depends on the nature of the activities and the characteristics of production (e.g. whether it involves the assembly of physical products, if the sector is characterised by long supply chains, etc.). In particular, digital technologies offer opportunities for digitalisation (and automation) of production processes; for interconnecting supply chains; and, for improving interactions with the final consumer.
- **Creating new digitally enabled markets and business models**. New markets or market segments enabled by digital technologies, often adjacent to traditional sectors, have been created over recent years. E-commerce, car-sharing services and Fintech services are well-known examples. While new business models are emerging across the economy, the scale and disruption potential of these trends vary across sectors. In some cases, those business models may displace traditional ones (e.g. travel agencies). In other cases, the two models may co-exist and expand the product or service offering (e.g. brick-and-mortar existing simultaneously with online retail stores).

Data needs and challenges for innovation

Data have become a key input for innovation (Table 4.1). However, sectoral differences also arise because access to data differs across sectors. For instance, in some sectors, data needed for innovation are more sensitive than in others (such as patient data for healthcare innovations). They may also be less widely available (such as farming data for the development of smart farming services, given the low digital technology uptake in agriculture). The nature of data privacy and safety challenges also differ. The protection of data generated by connected cars and transportation systems is critical to avoid cyber-attacks that could put road safety at risk. At the same time, misuse and leakage of personal data are more problematic in the retail sector. Some sectors may also be more attractive than others to digital talent, creating differences in the capacity to exploit data. Data ownership conditions may likewise be a barrier to innovation in some sectors. This is particularly the case in sectors like agriculture where data are captured by some actors but exploited by others.

Unequal access to data across firms can create an uneven playing field within the same sector. This, in turn, can contribute to higher market concentration. Amazon and Google, for example, have higher capacity to access large amounts of consumer data compared to other retailers.

Table 4.1 presents some of the differences across the agri-food, automotive and retail sectors regarding the type of data needed for innovation purposes, and the opportunities and challenges related to those data types.

Table 4.1. Differences in data requirements for sectoral applications

	Data needs	Main challenges
Agriculture (precision agriculture)	• Aggregated sensor data from many farms (captured by sensors in the fields or mounted on machinery or drones) • Satellite data (GIS, meteorological data, satellite imagery on crops)	• Low levels of digital technology adoption and high cost of uptake, particularly for small farms • Data sharing (resistance by farmers) • Data quality and integration • Building trusted data analytics
Automotive industry (connected cars)	• Data on driver behaviour, car status and location • Historical data on car performance (for predictive maintenance services) • GIS, real-time traffic information	• Skills to exploit data • Data integration • Data privacy (risks, e.g. usage-based insurance contracts) • Road safety (risk of cyber-attacks)
Retail (personalisation of consumer experience)	• Customers and transactions data • Personal data on social media and search websites	• Skills to exploit data • Data integration • Personal data privacy (risk e.g. price discrimination)

Note: GIS = geographic information system.
Source: Paunov and Planes-Satorra (2019), "How are digital technologies changing innovation? Evidence from the agriculture, automotive and retail sectors", https://doi.org/10.1787/67bbcafe-en.

Digital technology adoption and diffusion trends

Digital technology adoption is heterogeneous across sectors (Calvino et al., 2018). Industry estimates, for instance, show that sectors such as automotive and financial services are leading AI adoption, relative to the tourism and construction sectors, among others (Bughin et al., 2017). Key factors influencing adoption include:

- **Capabilities to take up new digital technologies**. Skills for adoption differ across sectors. For instance, sectors such as agriculture and construction, characterised by relatively high shares of low-skilled workers, register low uptake for digital technology. Capacities needed for digital technology adoption include skills at the individual level (e.g. information and communication technology skills, data expertise or previous related knowledge) and at the organisational level. The latter include the capacity to fine-tune organisational structures, adjust processes, redefine strategies and tasks, and manage emerging risks, among others. In this sense, the capacities of managers and their understanding of digital transformation dynamics are critical.
- **Pressures from market competition**. The emergence of new players in the market (i.e. digital start-ups or tech firms that enter existing markets or create new activities adjacent to traditional sectors) is pushing incumbents to innovate. However, such pressures seem to be more critical in some sectors than others. For instance, in the automotive industry, the market entry of firms such as Alphabet (investing in the development of self-driving cars) and Zipcar (offering car-sharing services) is pressuring incumbents to embrace new digital innovations.

  Some sectors are particularly affected by the emergence of new platforms. For example, the emergence of digital platforms is significantly reshaping the tourism industry. Booking.com, for example, enables consumers to search, compare and book accommodation and transportation options. As another example, sharing economy platforms such as Airbnb provide for peer-to-peer accommodation services.
- **Specific sectoral characteristics and structures**. Sectoral characteristics also influence the pace of digital technology adoption. Digital technologies, in particular, permeate the activities of different types of actors within the sector. These range from small and medium-sized enterprises (SMEs) to large firms, start-ups and research institutions. Large firms are usually early adopters of new technologies. This is mainly due to their access to the resources needed to invest in new technologies and the greater presence of workers with relevant technical expertise. However, large firms may also suffer from inertias, hierarchical and rigid structures, and legacy systems that can hamper their transformation (Rogers, 2003; Zhu, Kraemer and Xu, 2006).

Firms integrated into global value chains may be more exposed to digital technologies, and have higher incentives to adopt them. Their suppliers may adjust more rapidly to requests from upstream producers to adopt new practices, and receive support to implement them. For instance, Toyota supports its suppliers in implementing their new production systems (Kern and Wolff, 2019). The diffusion of digital technology also relies on access to critical infrastructure, such as broadband Internet connection. This may be a challenge for sectors and firms located in more remote or rural areas. Firms in agriculture are a case in point.

- **Changes in consumer demands**. Changes in consumer needs and demand are also driving the digital transformation of sectors. For example, in the field of transportation, younger generations (especially in urban areas) show a higher preference for on-demand schemes, rather than car ownership. In retail, consumers show increasing preference for a combination of physical and online shopping, along with the quick delivery of products purchased on line.
- **Level of resistance to change**. Resistance to change may also differ across sectors, depending on several factors. First, resistance may correlate to awareness of the opportunities offered by digital technologies. It could also depend on the perceived and actual challenges for specific stakeholders from adopting digital technologies, including job displacement or retraining requirements. Finally, it could depend on absorption capacities and the state of development of sector-specific digital technology applications. Low levels of technology adoption may also reflect consumers' resistance to change, which differs across products. For instance, the adoption of e-commerce was initially slow, while user rates of mobile transportation services differed strongly across countries. Similarly, there may be more resistance to accepting robots for personal care services than for new transportation services.

How should innovation policy be adapted to the digital age?

Effective policy for innovation in the digital age requires governments to adopt policy mixes that respond to the changing context created by the digital transformation. The new mix should comprise five key policy objectives (Table 4.2): ensuring access to data for innovation; providing support and incentives to innovation and entrepreneurship; building a strong public research system and having a skilled labour force; fostering collaborative, competitive and inclusive innovation ecosystems; and setting national policies that account for the global context and citizens' concerns.

Digital transformation calls for changes that affect all innovation policies, but to varying degrees. Some domains of policy need to adapt their target or content to digital innovation, while essentially preserving their core objectives. This includes, for instance, policies supporting entrepreneurship, digital technology adoption by SMEs and the development of general-purpose technologies. Other domains need more change, including rethinking of the policy rationale: that includes public research policy (moving towards open science).

Access to data has become a major new theme in all policy domains relating to innovation, such as support for business innovation, public research and competition policy. Data has also become a policy domain itself, and subject to issues such as confidentiality and privacy that directly impact innovation.

Policy makers need to address several new challenges. These include ensuring greater responsiveness and agility of policies and setting national policies in view of developments in global markets. They must also provide information and foster dialogue so that citizens are well-informed of the realities of new technologies and can participate in choices over funding of technologies considered harmful. Finally, they must ensure that government can access the advanced skills (e.g. in the field of AI) and data needed to design appropriate regulations and policies, ensuring that new technologies do not harm the public interest.

A sectoral approach is required in policy areas in which sectors have different challenges and needs. This is particularly the case of data access policies, digital technology adoption and diffusion policies, and support for digital technology application development. For example, challenges in agriculture often relate to data sharing and integration, while in retail ensuring data privacy is a rising concern.

Table 4.2. Major changes to innovation policies called for by digitalisation, by policy domain

Policy domain	Changes required
Access to data	• Ensure access to data for innovators, taking into account the diversity of data and preserving rights and incentives[1] • Explore the development of markets for data
Business innovation and entrepreneurship	• Ensure that policies are anticipatory, responsive and agile • Support service innovation that implements digital technologies • Adapt the intellectual property (IP) system • Support the development of generic (multi-purpose) digital technologies[1]
Public research, education and training	• Promote open science (access to data, publications) • Support training in digital skills for science • Support interdisciplinarity in research • Invest in digital infrastructure for science • Facilitate co-creation between industry, science and civil society • Ensure that skills needed for digital innovation are being developed (in collaboration with education and labour market policy authorities) • Support education and training for the development of managerial skills
Competition, collaboration and inclusiveness	• Review competition policies from the perspective of innovation in the digital age (e.g. rules regarding takeovers and standards; IP systems) • Support digital technology adoption by all firms, particularly SMEs[1] • Support social and territorial inclusiveness in digital innovation activities
Cross-cutting principles	• Frame national policies in view of global markets • Engage with citizens to address technology-related public concerns • Adopt a sectoral approach to policy making when necessary

1. These areas require a sectoral approach to innovation policy making.
Note: SMEs = small and medium-sized companies.
Source: OECD (2019), *Digital Innovation: Seizing Policy Opportunities*, https://doi.org/10.1787/a298dc87-en.

Innovation is also influenced by many policies that do not target innovation explicitly or primarily. These include education, tax, health, environmental, transportation and competition policies. Competition policy is particularly critical for innovation, as only the right competitive environment will stimulate firms to innovate and foster innovation-driven growth.

Data access policies

As data now constitute a major input to innovation, access to data – and to the tools that gather and help interpret data – will influence who participates in digital innovation, and in what ways. Therefore, a specific policy agenda around data access needs to be developed (OECD, 2015a). The main objective of data access policies should balance two elements. On the one hand, policies should ensure the broadest possible access to data and knowledge (incentivising sharing and reuse) to favour competition and innovation. On the other, they should respect constraints regarding data privacy, ethical considerations, economic costs and benefits (i.e. incentives to produce the data) and intellectual property rights (IPRs).

Policies should consider the diversity of data types, which imply differences in terms of access and other challenges associated with their generation, access and exploitation. Access to public research data, in particular, allows the reproduction and testing of the validity of scientific research, as well as reuse in further research (OECD, 2015b; Dai, Shin and Smith, 2018). Some governments establish open access to data generated by public services (e.g. weather monitoring, urban transportation, etc.) to foster data-driven innovation. For instance, the United Kingdom's open data portal (data.gov.uk) publishes data from the central government, local authorities and other public bodies. It produces data on a variety of fields to create new opportunities for organisations to build innovative digital goods and services. Other examples include the open data portals of Canada (open.canada.ca), France (data.gouv.fr), Japan (data.go.jp) and the United States (data.gov).

Appropriate conditions should also be created to allow for the emergence of data markets. Trading data may facilitate innovation, as well as put a price on data generation and curation for future use – thus facilitating the generation of more data. Markets in data also facilitate entry for start-ups that are data-poor but which require data as part of their business model.

There are, however, major challenges to the development of markets for data (and knowledge). Data, for example, are often adapted to a specific context. Outside of this context, they may have little or no value, thereby limiting their transferability. Other key challenges relate to appropriability, difficulties in evaluating the true market value and quality of data, and privacy and safety concerns affecting personal data.

Policies to support innovation and entrepreneurship

Ensure that policies are anticipatory, responsive and agile

The policy instruments needed for the digital age should be anticipatory, responsive and agile. The innovation agenda is shifting quickly and difficult to predict in certain fields. Therefore, government needs to become more flexible and alert to change, while keeping (prudential) rules of engagement when it comes to specific policy instruments.

Approaches to ensuring this policy responsiveness includes the deployment and monitoring of small-scale policy experiments. These experiments can help assess their relevance and efficiency in a context of high uncertainty, based on which they could be easily scaled up, down or abandoned.

In a context of rapid change, it is also critical to streamline application procedures for innovation support instruments. For example, the Pass French Tech programme offers fast-growing start-ups simplified and quick access to services to help them expand. These include services in areas such as financing, access to new markets, innovation and business development (French Tech, n.d.).

Using digital tools to design innovation policy and monitor policy targets is another option to spur faster and more effective decision making. For example, semantic analysis can identify policy trends and anticipate technology trends by exploring large quantities of textual information (e.g. innovation policy documents, patents, scientific articles) (OECD, 2018). While still experimental, semantic analysis has tracked strengths in specific research fields based on text information contained in publications. It is also used by innovation and research funding agencies to build better connections between recipients of support, based on information on their research activities.

Instruments that do not target a specific technology can also increase flexibility. Mission-oriented programmes that set a goal, but do not impose the means to reach it, can help. Such programmes may provide more autonomy and agility to choose the proper technological avenues to achieve a stated policy objective. The drawbacks of instruments without a specific target must be considered against the advantage of greater flexibility.

Certain environments, including public procurement with specific requirements such as data security, leave no choice of technology. In these cases, designing public institutions connected to technology developments in the private sector can prove useful. These institutions inform governments about the latest technology developments, as well as their potential benefits and harmful impacts. Data61 in Australia and the Digital Catapult in the United Kingdom are examples.

Support service innovation to lever the potential of digital technologies

Many innovation policies have been conceived for innovation in manufacturing. This has specific characteristics, such as being intensive in research and development (R&D), and often resulting in patents. In a context where services are becoming a key focus of innovation, policy initiatives should ensure that services innovation is considered. Initiatives that include emerging needs in services innovation may include support for projects. These could develop entirely new services using digital technologies, such as the Smart and Digital Services Initiative in Austria (FFG, n.d.) It could also support manufacturing SMEs to develop services related to their products. Service design vouchers for manufacturing SMEs in the Netherlands is one such example (RVO, 2018).

Adapt intellectual property systems

Digitalisation is transforming the IP system, which was designed for tangible inventions embodied in physical products and processes. With digitalisation, the IP system is confronted with new questions that require policy responses. These include how to incentivise data generation in a context of increasingly open data systems. Other questions revolve around ownership of patentable inventions produced by AI and the risk of counterfeiting the intangible components of products.

New digital technologies may also help enforce IP rights. Eventually, blockchain-enhanced IP on a range of intangible goods (e.g. photographs, music, movies), and even some tangible goods (such as 3D-printed items with unique digital identifiers), may help to create a more easily enforced system of IP.

Support development of generic digital technologies to respond to societal challenges

Policies need to ensure that multi-purpose digital technologies are developed to serve both commercial purposes, as well as social and environmental purposes. Public research is often driving advances in these areas, while an increasing number of non-profit organisations and private firms also pursue such objectives. Many examples exist of AI applications that tackle social and environmental challenges. Satellite imagery and deep learning techniques, for example, identify illegal fishing vessels and monitor changes in coral reefs to inform conservation interventions. Audio sensors can detect illegal logging. And face detection, social network analysis and natural language processing can identify victims of sexual exploitation on the Internet (Chui et al., 2018).

More engagement and debate with the public is also needed to demonstrate the characteristics of these technologies and appropriately address public concerns (e.g. privacy protection, development of applications for the public good). A lack of engagement with society creates the risk of a future backlash. This could have negative impacts on the development and deployment of important technologies.

Public research and education policies

Promote the digitalisation of public research

Strengthening researchers' digital skills would ensure that new digital tools are integrated into public research processes (e.g. ML techniques). Specific training and capacity building activities, for example, could be offered. This tactic is a key objective of the digitalisation strategy for the higher education sector in Norway (2017-21) (Government of Norway, 2018).

Such measures should be accompanied by investments in digital tools and infrastructures critical for research (e.g. platforms for data sharing, supercomputing facilities for AI). Japan, for example, is investing more than USD 120 million annually to build a HPC infrastructure. It will be accessible to universities and public research centres for R&D purposes in a range of fields (HPCI, n.d.).

Stimulating interdisciplinary research (e.g. cross-departmental research projects) and the engagement in partnerships with other research institutions and with industry is also a larger priority in a context of digital innovation. Specifically, data science applications provide for new opportunities across academic disciplines. With regards to partnerships between industry and science, physical spaces remain important for more collaborative innovation. However, digital platforms can complement physical space and allow for new types of collaboration across geographic boundaries.

Build digital skills, including in the field of data analytics

Education and research authorities play a key role in building the digital skills needed across the economy. Innovation authorities should collaborate with them towards several goals. First, they help identify the new skills needed in a context of digital transformation. Second, they should provide inputs for university and vocational training programmes to fill critical talent shortages (e.g. data scientists) that often requires more interdisciplinary curricula. Innovation authorities also facilitate training for SMEs from traditional sectors to ensure they leverage the potential of digital technologies.

Foster competitive, collaborative and inclusive innovation ecosystems

Ensure that innovation ecosystems remain competitive

Dialogue between competition authorities and agencies responsible for innovation policy should address key questions. These include the use of data as a source of market power and the contestability of markets in which digital innovation is an important feature. Such markets are subject to rapid innovation (a source of contestability) and various sorts of scale economies (a source of persistent concentration) (Guellec and Paunov, 2018). New policies need to recognise the importance and prevalence of economies of scale, while ensuring equal access to markets and resources. As competition in digital markets is global, greater co-operation across jurisdictions may also be needed (OECD, 2019).

Do innovation policy instruments and regulations (e.g. support for R&D, IPRs) have an asymmetric impact on market players? Policy makers should consider this question. While such instruments are accessible to all in principle, this may not be the case in practice. For example, firms may lack capacity to defend their IP rights in courts, to co-operate effectively with public labs or to access public procurement.

Support collaboration for innovation

In the digital context, innovation policies will have to continue supporting collaborative innovative ecosystems. New policy approaches to foster collaborative innovation include the use of crowdsourcing and open challenges, as well as the creation of living labs. These approaches can help find innovative solutions to pressing challenges and foster co-creation between various actors.

Intermediary organisations, such as the Fraunhofer Institutes in Germany and the Catapult Centres in the United Kingdom, have become central players in innovation ecosystems. They provide services such as matching firms that need technology solutions with potential suppliers. New research and innovation centres, often public-private partnerships, have also been created. These centres provide spaces for multidisciplinary teams of public researchers and businesses to work together on specific technology challenges. They often stand out for their innovative organisational structures. Examples include Data61 in Australia and Smart Industry Fieldlabs in the Netherlands (Box 4.1).

Box 4.1. Approaches to fostering collaboration for innovation in the digital age

Intermediary organisations

Intermediary organisations connect different actors in innovation ecosystems (innovators, big firms, SMEs, investors, etc.) and facilitate their matching and collaboration for research and innovation. The Catapult Centres in the United Kingdom are a network of ten not-for-profit, independent physical centres that connect businesses with the United Kingdom's research and academic communities. Each focuses on a strategic technology area in which the United Kingdom has great potential for growth. They offer a space with the facilities and expertise to enable businesses and researchers to collaboratively solve key problems and develop new products and services on a commercial scale. They also support firms' access to foreign markets, create and retain high value jobs and attract inward investments from global technology businesses.

Collaborative research and innovation centres

Several countries have created (networks of) research centres. Multidisciplinary teams of public researchers and businesses work together at these centres on specific technology challenges. The centres provide spaces for collaboration and co-creation and often stand out for their innovative organisational structures.

CSIRO's Data 61, Australia's largest digital R&D centre, aims to put Australia at the forefront of data-driven innovation. To that end, it pursues new-to-the-world fundamental and applied research and works collaboratively with others in the innovation ecosystem, including universities, government and industry. To increase agility and attract digital talent, Data61 has adopted a "start-up culture" or "market pull" approach. Organisational structures are flatter (i.e. with less middle-management and higher autonomy of staff). Research leaders are encouraged to experiment with new ideas and take risks, while maintaining alignment with the strategic goals of the organisation. A "challenge model" also stimulates multidisciplinary teams to address large-scale social and business challenges.

Smart Industry Fieldlabs in the Netherlands are public-private partnerships to create physical or digital spaces for member companies and research institutions. Together, they develop, test and implement new smart industry technological solutions in various fields. These include automation, zero-defect manufacturing, flexible production, value creation based on big data, 3D printing and robotics. The 32 field labs, which have flat structures and follow a project-based approach, typically include users of such solutions, (potential) suppliers and knowledge institutes. They are active in collaborative research, concept validation, prototyping, testing and validation.

Crowdsourcing, open challenges and living labs

Various countries are harnessing the power of crowdsourcing, open challenges and living labs to drive innovation. The US government designed Citizenscience.gov to enhance use of crowdsourcing to engage the public in addressing social needs and accelerate innovation. The Social Challenges Innovation Platform (SocialChallenges.eu) encourages social innovators and entrepreneurs to post innovative solutions to social and environmental challenges that public authorities, private firms or non-governmental organisations aim to solve. Pit Stops, organised by the Digital Catapult (United Kingdom), encourage open innovation by bringing together large firms, SMEs, start-ups and academics to solve specific technology challenges. Disruptive technology start-ups and other actors able to solve such challenges are identified via online open calls.

Living labs are localised areas of experimentation within urban environments in which stakeholders collaboratively develop new technology-enabled solutions. For instance, Antwerp (Belgium) is developing a "City of Things" (IMEC, n.d.) through installation of a dense network of smart sensors and wireless gateways in buildings, streets and objects. Companies can use collected data to build innovative applications.

Source: OECD (2019), *Digital Innovation: Seizing Policy Opportunities*, https://doi.org/10.1787/a298dc87-en.

Support digital technology adoption by all firms, particularly SMEs

Firms (particularly SMEs) face important challenges to adapt to digital transformation. Such adaptation requires much more than simply purchasing new computers and software: it is about changing business processes, and often business models. This frequently implies new strategic capabilities, new skills, investments in new technologies and significant restructuring, all of which can carry risk. The failure of many SMEs that do not digitalise would mean the loss of much industry and market-specific know-how, which constitutes unique intangible capital. It is therefore in the public interest to support adaptation of SMEs selectively; less competitive firms would not be saved to avoid hampering the competitive process.

Innovative policy to foster diffusion focuses on helping test new digital technology applications by, for instance, creating test beds and regulatory sandboxes. Innovative initiatives also enhance early adoption of advanced digital technologies. To that end, they help innovators access state-of-the-art facilities and expertise (e.g. in the fields of AI or supercomputing). SMEs are revisiting traditional instruments to foster technology adoption such as awareness-raising campaigns, innovation vouchers, technical assistance and training. They are adapting these instruments to specific challenges of the digital age, and often use digital tools themselves (Box 4.2).

Box 4.2. Demonstration and testing of new digital technologies

Demonstration and testing facilities for SMEs

Some countries have established new facilities for demonstration and testing of digital technologies to increase adoption. For instance, the SME 4.0 Competence Centres in Germany offer SMEs access to demonstrations of Industry 4.0 technologies and sector-specific applications (e.g. 3D printing, sensors). These demonstration facilities are often located at universities and allow simulation of business and production processes in a similar to real-world environment.

The Industry Platform 4 FVG, established in the Italian region of Friuli Venezia Giulia, offers access to testing equipment, prototyping tools and demonstration labs. Several Austrian universities (TU Wien, TU Graz and Johannes Kepler University Linz) have also set up pilot factories, where SMEs have the chance to test new technologies and production processes without affecting production in their own facilities.

Experimenting with new technology applications

Countries are also exploring novel approaches to fostering testing of, and experimentation with, new digital technologies and applications in a near to real-world environment:

- **Test beds** provide environments where new technology developments can be tested in controlled but near to real-world conditions. Such testing environments are critical for research and innovation in certain areas, such as autonomous driving, and help accelerate adoption of new digital technologies. Finland is establishing a number of test beds for the open development of transport and mobility solutions, including automated driving, mobility-as-a-service and intelligent traffic infrastructures. In the United Kingdom, the National Health Service (NHS) introduced a Test Beds Programme in partnership with industry. This allows testing of innovations such as combinations of new digital devices such as sensors, monitors and wearables with data analysis. It also permits testing of new approaches to service delivery facilitated by digital technologies. Successful innovations are then made available to the NHS and care organisations around the country.
- **Regulatory sandboxes** provide a limited form of regulatory waiver or flexibility for firms to test new products or business models with reduced regulatory requirements. At the same time, they preserve some safeguards such as ensuring appropriate consumer protection. Sandboxes help identify and better respond to regulatory breaches and enhance regulatory flexibility. They are particularly relevant in highly regulated industries, such as financial services, transport, energy and health. The United Kingdom's Financial Conduct Authority pioneered this approach with the launch of the Fintech regulatory sandbox, which encourages innovation in financial technology. The sandbox provides a controlled environment for businesses to test innovative products and services without incurring the regulatory consequences of pilot projects.

Source: OECD (2019), *Digital Innovation: Seizing Policy Opportunities*, https://doi.org/10.1787/a298dc87-en.

Support social and territorial inclusiveness

Innovation policies play a role in enhancing participation of disadvantaged individuals in digital innovation activities. Policy instruments address social inclusiveness challenges in the digital age in various ways. Some aim to build capacities through, for example, digital skills and entrepreneurship education. Others address discrimination and stereotypes through role models and mentoring programmes, among other approaches. Still others address barriers to entrepreneurship faced by disadvantaged groups. For example, they facilitate access to finance through microcredit or equity financing, provide tailored business development support and promote entrepreneurs' insertion in business and research networks. Planes-Satorra and Paunov (2017) provides a wide range of examples of inclusive innovation policy.

Digital transformation also seems to favour further concentration of innovation activities in innovation hotspots (often urban areas). This calls for policies favouring territorial inclusiveness. "Excellence-based policies", even if blind to location, tend to favour geographical concentration, since excellence is concentrated. These indirectly widen the gap between leading and lagging regions. Excellence-based policies should therefore be complemented by policies favouring geographical inclusiveness and diversity. They should focus on fostering innovation at the local/regional level, and building on specific regional strengths and comparative assets (e.g. the Smart Specialisation approach in the European Union).

Cross-cutting policy principles

Set national policies in view of developments in global markets

Digitalisation facilitates the circulation of knowledge, including across national borders, reducing government's ability to restrict the benefits of policies to its own country. That raises a challenge for national policy makers. How can they ensure their own citizens (and taxpayers) benefit from national policies? Furthermore, how can they ensure that most of the benefits (e.g. income generated, productivity gained or jobs created) do not leak abroad? One related question concerns the sharing of benefits generated by the exploitation of national data (e.g. from the public health system) by foreign multinationals. Co-operative solutions are needed to share benefits arising from international flows of data and knowledge linked to national policies among countries. The OECD activity on base erosion and profit shifting is a step in this direction.[1]

Engage with citizens to address technology-related public concerns

The digital transformation has captured much attention in the press and with the public. In some cases, people fear leakages of personal data and the threat of robots taking jobs. Government and other actors must engage with stakeholders about these technologies and allay concerns through, for example, enhanced data privacy protection. Consultations with the public during the development of digital transformation strategies and other related policies can contribute. Without such public engagement, there is a risk of backlash in the future. This could have potentially negative impacts on the development and deployment of these technologies and their related benefits (OECD, 2015b; Winickoff, 2017; Dai, Shin and Smith, 2018).

Adopt a sectoral approach to policy making when necessary

Three policy domains require a sectoral approach when designing new initiatives, as the challenges and needs faced by sectors in these areas vary significantly:

- **Data access policies** should consider the diversity of data types needed for innovation in different sectors, given differences to access and other challenges associated with data generation, exploitation and ownership. For instance, precision agriculture draws mainly on sensor and satellite data. Conversely, the retail sector exploits consumer purchasing and social media data to personalise services. In agriculture, challenges often relate to data sharing and integration, while in retail ensuring data privacy is a rising concern.

- **Digital technology adoption and diffusion policies** should be tailored to the specific needs of the sector and/or type of actor (notably SMEs). These policies could involve awareness-raising, training and education, demonstration and testing of new technologies, and the operation of intermediary institutions. Diffusion is more challenging in some sectors than in others due to different production structures. For instance, many small firms in a sector may be geographically dispersed or a few larger ones may be geographically close. Other challenges include the landscape of intermediary institutions and/or the availability of digital capacities.
- **Policies to develop sectoral applications of digital technologies** should be supported where market conditions have inhibited the development of private sector-led solutions. This will ensure that such technologies provide benefits across the economy. The gap between future digital technology opportunities and current applications differs across sectors. This frustrates adoption of digital technologies for firms operating in certain sectors where applications do not yet exist (e.g. in the field of AI). Public research could support building more applications and help adoption across the economy where private business does not have the incentives to produce them.

Designing effective and tailored support to sectors operating in the digital context requires, as a first step, establishing mechanisms to strengthen policy intelligence. These may include roadmaps or sectoral plans for strategic sectors, in collaboration with industry stakeholders and social partners. Examples include the Sector Competitiveness Plans developed by six sector-specific Industry Growth Centres in Australia (Government of Australia, 2017).

Conclusion

Chapter 4 discusses the impacts of the digital transformation on innovation processes and outcomes. The chapter highlights general trends across the economy and factors behind sector-specific dynamics. In view of such impacts, it evaluates how policy support to innovation should adapt and in what directions, providing examples of novel approaches to innovation policy.

The chapter shows that four pervasive trends characterise innovation in the digital age. First, data are becoming key inputs for innovation. Second, innovation activities increasingly focus on the development of services enabled by digital technologies. Third, innovation cycles are accelerating. Virtual simulation, 3D printing and other digital technologies are providing opportunities for more experimentation and versioning in innovation processes. Fourth, innovation is becoming more collaborative, given the growing complexity and interdisciplinary needs of digital innovation.

Impacts of the digital transformation differ significantly across (and within) sectors. This reflects differences in the scope of opportunities for innovation in products, processes and business models that digital technologies offer. It reflects differences in the types of data needed for innovation and thus the challenges faced for their exploitation. And it reflects different conditions for digital technology adoption and diffusion.

The effective development of innovation in the digital age requires that governments adopt policy mixes that respond to the changing context created by the digital transformation. The changes called for by digitisation affect the entire innovation policy spectrum, but to varying degrees across policies. Access to data has become a major new theme in all policy domains relating to innovation, such as innovation support, public research and competition. It has also become a policy domain in itself, subject for instance to confidentiality and privacy issues that also directly impact innovation.

New challenges for policy making that need to be addressed include ensuring greater responsiveness and agility of policies; setting national policies in view of global markets; and engaging with citizens on new technologies. Equally, policy makers must ensure that government can access advanced skills, such as in the field of AI, and data needed to design appropriate regulations and policies. Finally, they must ensure that new technologies and applications do not harm the public interest. A sectoral approach is also needed in some policy areas.

This chapter is a first step in understanding the changing characteristics of innovation in the digital age. An important priority for policy research in this field involves gathering cross-country information on adoption rates of most advanced digital technologies at the firm level. Such data need to consider and capture ongoing technology trends. They would allow adoption trends to be measured across sectors, and across types of firms and locations. This, in turn, would help better identify the specific factors spurring and restraining digital innovation.

References

Amazon (n.d.), "Amazon Go", webpage, www.amazon.com/b?node=16008589011 (accessed 16 April 2018).

Bughin, J. et al. (2017), "Artificial intelligence: The next digital frontier?" *Discussion Paper*, June, McKinsey Global Institute, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

Calvino, F. et al. (2018), "A taxonomy of digital intensive sectors", *OECD Science, Technology and Industry Working Papers*, No. 2018/14, OECD Publishing, Paris, https://doi.org/10.1787/f404736a-en.

Chui, M., et al. (2018), "Notes from the AI frontier: Applying AI for social good", *Discussion Paper*, December, McKinsey Global Institute.

Dai, Q., E. Shin and C. Smith (2018), "Open and inclusive collaboration in science: A framework", *OECD Science, Technology and Industry Working Papers*, No. 2018/07, OECD Publishing, Paris, https://dx.doi.org/10.1787/2dbff737-en.

Ezell, S. (2018), "Why manufacturing digitalization matters and how countries are supporting it", Information Technology and Innovation Foundation, Washington, DC, www2.itif.org/2018-manufacturing-digitalization.pdf (accessed 21 January 2019).

FFG (n.d.), "Service innovations – Funding, guidelines", webpage, https://www.ffg.at/en/programme/service-innovations (accessed 26 June 2019).

French Tech (n.d.), "Pass French Tech", webpage, www.lafrenchtech.com/en-action/pass-french-tech (accessed 25 May 2018).

Gawer, A. and M. Cusumano (2014), "Industry platforms and ecosystem innovation", *Journal of Product Innovation Management*, Vol. 31/3, Product Development and Management Association, St. Paul, United States, pp. 417-433, http://dx.doi.org/10.1111/jpim.12105.

General Electric (2017), *What is FastWorks?*, GE Reports Canada, www.ge.com/reports/tag/fastworks/ (accessed 25 April 2018).

Government of Australia (2017), *Industry Growth Centres Initiative: Sector Competitiveness Plans Overview*, September, Department of Industry, Innovation and Science, www.industry.gov.au/sites/g/files/net3906/f/May%202018/document/pdf/industry_growth_centres_initiative_-_sector_competitiveness_plans_overview.pdf.

Government of Norway (2018), "Digitalisation Strategy for the Higher Education Sector 2017-2021", webpage, www.regjeringen.no/en/dokumenter/digitalisation-strategy-for-the-higher-education-sector-2017-2021/id2571085/sec5 (accessed 29 May 2018).

Guellec, D. and C. Paunov (2018), "Innovation policies in the digital age", *OECD Science, Technology and Industry Policy Papers*, No. 59, OECD Publishing, Paris, https://dx.doi.org/10.1787/eadd1094-en.

Hewlett-Packard (2017), "Hirotec: From smart manufacturing, to smart factory – to smart enterprise", webpage, https://www.ptc.com (accessed 23 April 2018).

HPCI (n.d.), High Performance Computer Infrastructure website, www.hpci-office.jp/folders/english (accessed 20 April 2019).

IMEC (n.d.), "City of things", webpage, www.imec-int.com/en/cityofthings (accessed 20 April 2019).

Kern, J. and P. Wolff (2019), "The digital transformation of the automotive supply chain – an empirical analysis with evidence from Germany and China: Case study contribution to the OECD TIP Digital and Open Innovation project", OECD, Paris, www.innovationpolicyplatform.org/www.innovationpolicyplatform.org/system/files/imce/AutomotiveSupplyChain_GermanyChina_TIPDigitalCaseStudy2019_1/index.pdf.

Lely (2016), "Sensor technology helps farmers with even more precision and relevant information", 17 June, www.lely.com/au/news/2016/06/17/sensor-technology/.

Lund, S., J. Manyika and K. Robinson (2016), "Managing talent in a digital age", *McKinsey Quarterly*, March, McKinsey & Company, www.mckinsey.com/industries/high-tech/our-insights/managing-talent-in-a-digital-age.

OECD (2019), *Digital Innovation: Seizing Policy Opportunities*, OECD Publishing, Paris, https://doi.org/10.1787/a298dc87-en.

OECD (2018), "CSTP-TIP workshop: Semantic analysis for innovation policy – Summary", OECD, Paris, www.innovationpolicyplatform.org/www.innovationpolicyplatform.org/system/files/imce/OECD%20TIP_SemanticAnalysis_summary_FINAL_0/index.pdf (accessed 13 November 2018).

OECD (2017), *The Next Production Revolution: Implications for Governments and Business,* OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-en.

OECD (2015a), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264229358-en.

OECD (2015b), *The Innovation Imperative: Contributing to Productivity, Growth and Well-Being*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264239814-en.

Pandolph, S. (2017), "Sephora leads the way in personalization", *Business Insider France*, 22 September, www.businessinsider.fr/us/sephora-leads-the-way-in-personalization-2017-9.

Paunov, C. and S. Planes-Satorra (2019), "How are digital technologies changing innovation? Evidence from the agriculture, automotive and retail sectors"*, OECD Science, Technology and Industry Policy Papers*, No. 74, OECD Publishing, Paris, https://doi.org/10.1787/67bbcafe-en.

Planes-Satorra, S. and C. Paunov (2017), "Inclusive innovation policies: Lessons from international case studies"*, OECD Science, Technology and Industry Working Papers*, No. 2017/2, OECD Publishing, Paris, https://dx.doi.org/10.1787/a09a3a5d-en.

PSC/CAR (2017), *Planning for Connected and Automated Vehicles*, prepared by Public Sector Consultants and Center for Automotive Research for Greater Ann Arbor Region, www.cargroup.org (accessed 29 October 2018).

Rogers, E. (2003), *Diffusion of Innovations, 5th Edition*, The Free Press, New York.

RVO (2018), "Service Design-Vouchers", webpage (in Dutch), Netherlands Enterprise Agency, www.rvo.nl/subsidies-regelingen/service-design-vouchers (accessed 23 October 2018).

The Economist/Capgemini (n.d.), "At a crossroads: Supply chains and digital transformation", webpage, https://learnmore.economist.com/story/capgemini-supply-chains-and-digital-transformation (accessed 17 April 2018).

Tripoli, M. and J. Schmidhuber (2018), "Emerging opportunities for the application of blockchain in the agri-food industry agriculture", *Issue Paper*, Food and Agriculture Organization of the United Nations and International Centre for Trade and Sustainable Development, Rome and Geneva, www.fao.org/3/ca1335en/CA1335EN.pdf.

VDA (2015), "Automation: From driver assistance systems to automated driving", webpage, www.vda.de/en/topics/innovation-and-technology/automated-driving/automated-driving.html (accessed 29 October 2018).

Winickoff, D. (2017), "Public acceptance and emerging production technologies", in *The Next Production Revolution: Implications for Governments and Business,* OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-en.

Wolfert, S. et al. (2017), "Big data in smart farming – A review", *Agricultural Systems*, Vol. 153, Elsevier, Amsterdam, pp. 69-80, http://dx.doi.org/10.1016/J.AGSY.2017.01.023.

Zhu, K., K. Kraemer and S. Xu (2006), "The process of innovation assimilation by firms in different countries: A technology diffusion perspective on e-business", *Management Science*, Vol. 52/10, InformsPubsOnLine, pp. 1557-1576, http://dx.doi.org/10.1287/mnsc.1050.0487.

Note

[1] www.oecd.org/tax/beps/.

5 Artificial intelligence, digital technology and advanced production

Alistair Nolan

Directorate for Science, Technology and Innovation, OECD

Chapter 5 examines a selection of policies to enable the use of digital technology in advanced production. The first part looks at individual technologies, their uses and specific policy implications, namely artificial intelligence (AI), blockchain and 3D printing, as well as new materials and nanotechnology (development of which involves complex digital processes). The second part addresses cross-cutting policy issues relevant to digital technology and production. These are technology diffusion, connectivity and data, standards-setting processes, digital skills, access to and awareness of high-performance computing, intellectual property systems and public support for research and development. With respect to public research, particular attention is given to research on computing and AI, as well as the institutional mechanisms needed to enhance the impact of public research.

Introduction

New digital technologies are essential to raising living standards and countering the declining labour productivity growth in many OECD countries that has occurred over recent decades. Rapid population ageing – the dependency ratio in the OECD is set to double over the next 35 years – makes raising labour productivity more urgent. Digital technologies can increase productivity in many ways. For example, they can reduce machine downtime, as intelligent systems predict maintenance needs. They can also perform work more quickly, precisely and consistently with the deployment of increasingly autonomous, interactive and inexpensive robots. New digital technologies in production will also benefit the natural environment in several ways by, for instance, making zero-defect production a possibility in some industries.

Digital production technologies: Recent developments and policy implications

Artificial intelligence in production

The Oxford English Dictionary defines artificial intelligence (AI) as "the theory and development of computer systems able to perform tasks normally requiring human intelligence". Expert systems – a form of AI drawing on pre-programmed expert knowledge – have been used in industrial processes for close to four decades (Zweben and Fox, 1994). The development of deep learning using artificial neural networks[1] has been the main source of recent progress in the field. As a result, AI can be applied to most industrial activities – from optimising multi-machine systems to enhancing industrial research (Box 5.1). Furthermore, the use of AI in production will be spurred by automated ML processes that can help businesses, scientists and other users employ the technology more readily. With respect to AI that uses deep learning techniques and artificial neural networks, the greatest commercial potential for advanced manufacturing is expected in supply chains, logistics and process optimisation (Chui et al., 2018). Survey evidence also suggests that the transportation and logistics, automotive and technology sectors lead in terms of the share of early AI-adopting firms (Küpper et al., 2018).

Box 5.1. Recent applications of artificial intelligence in production

A sample of recent uses of AI in production illustrates the breadth of the industries and processes involved:

- In pharmaceuticals, AI is set to become the "primary drug-discovery tool" by 2027, according to Leo Barella, Global Head of Enterprise Architecture at AstraZeneca. AI in preclinical stages of drug discovery has many applications. They range from compound identification and managing genomic data to analysing drug safety data and enhancing in-silico modelling (AI Intelligent Automation Network, 2018).
- In aerospace, Airbus deployed AI to identify patterns in production problems when building its new A350 aircraft. A worker might encounter a difficulty that has not been seen before, but the AI, analysing a mass of contextual information, might recognise a similar problem from other shifts or processes. Because the AI immediately recommends how to solve production problems, the time required to address disruptions has been cut by one-third (Ransbotham et al., 2017).
- In semiconductors, an AI system can assemble circuitry for computer chips, atom by atom (Chen, 2018); Landing.ai has developed machine-vision instruments to identify defects in manufactured products – such as electronic components – at scales that are invisible to the unaided eye.
- In the oil industry, General Electric's camera-carrying robots inspect the interior of oil pipelines, looking for microscopic fissures. If laid side by side, this imagery would cover 1 000 square kilometres every year. AI inspects this photographic landscape and alerts human operators when it detects potential faults (Champain, 2018).

- In mining, AI is used to explore for mineral deposits and optimise use of explosives at the mine face (even considering the cost of milling larger chunks of unexploded material). It is also used to operate autonomous drills, ore sorters, loaders and haulage trucks. In July 2017, BHP switched to completely autonomous trucks at a mine in Western Australia (Walker, 2017).
- In construction, generative software uses AI to explore every permutation of a design blueprint. It suggests optimal building shapes and layouts, including the routing of plumbing and electrical wiring. Furthermore, it can link scheduling information to each building component.
- AI is exploring decades of experimental data to radically shorten the time needed to discover new industrial materials, sometimes from years to days (Chen, 2017).
- AI is enabling robots to take plain-speech instructions from human operators, including commands not foreseen in the robot's original programming (Dorfman, 2018).
- Finally, AI is making otherwise unmanageable volumes of Internet of Things (IoT) data actionable. For example, General Electric operates a virtual factory, permanently connected to data from machines, to simulate and improve even highly optimised production processes. To permit predictive maintenance, AI can process combined audio, video and sensor data, and even text on maintenance history. This can greatly surpass the performance of traditional maintenance practices.

Beyond its direct uses in production, the use of AI in logistics is enabling real-time fleet management, while significantly reducing fuel consumption and other costs. AI can also lower energy consumption in data centres (Sverdlik, 2018). In addition, AI can assist digital security. For example, the software firm Pivotal has created an AI system that recognises when text is likely to be part of a password, helping to avoid accidental online dissemination of passwords. Meanwhile, Lex Machina is blending AI and data analytics to radically alter patent litigation (Harbert, 2013). Many social-bot start-ups also automate tasks such as meeting scheduling (X.ai), business-data and information retrieval (butter.ai), and expense management (Birdly). Finally, AI is being combined with other technologies – such as augmented and virtual reality – to enhance workforce training and cognitive assistance.

AI could also create entirely new industries based on scientific breakthroughs enabled by AI, much as the discovery of deoxyribonucleic acid (DNA) structure in the 1950s led to a revolution in industrial biotechnology and the creation of vast economic value – the global market for recombinant DNA technology has been estimated at USD 500 billion.[2]

Adopting AI in production: main challenges

To date, despite AI's potential, its adoption in manufacturing has been limited. By one estimate, even among AI-aware firms, only around 20% use one or more AI technologies in core areas of business or at scale (Bughin et al., 2017). A more recent survey of 60 US manufacturers with annual turnovers of between USD 500 million and USD 10 billion yielded still more striking evidence of the limited diffusion of AI, finding that:

> *"Just 5% of respondents have mapped out where AI opportunities lie within their company and developing a clear strategy for sourcing the data AI requires, while 56% currently have no plans to do so." Atkinson and Ezell (2019).*

The challenges in using AI in production relate to its application in specific systems and the collection and development of high-quality training data.[3] The highest-value uses of AI often combine diverse data types, such as audio, text and video. In many uses, training data must be refreshed monthly or even daily (Chui et al., 2018). Furthermore, many industrial applications are still somewhat new and bespoke, limiting data availability. By contrast, sectors such as finance and marketing have used AI for a longer time (Faggella, 2018). Without large volumes of training data, many AI models are inaccurate. A deep learning supervised algorithm may need 5 000 labelled examples per item and up to 10 million labelled examples to match human performance (Goodfellow, Bengio and Courville, 2016).

In the future, research advances may make AI systems less data-hungry. For instance, AI may learn from fewer examples, or generate robust training data (Simonite, 2016). In 2017, the computer program AlphaGo Zero famously learned to play Go using just the rules of the game, without recourse to external data. In rules-based games such as chess and Go, however, high performance can be achieved based on simulated data. But for industry, training data come from real-world processes and machines.[4]

Data scientists usually cite data quality as the main barrier to successfully implementing AI. Industrial data might be wrongly formatted, incomplete, inconsistent, or lack metadata. Data scientists will often spend 80% of their time cleaning, shaping and labelling data before AI systems can be put to work. The entire process requires skilled workers, and may have no *a priori* guarantee of success. Data might have to be drawn and unified from data silos in different parts of a company. Customer data, for instance, may be held separately from supply-chain data. Connecting data silos could also require complementary ICT investments. Moreover, some processes may simply lack the required volumes of data.

Adding to the challenge, manufacturers might have accuracy requirements for AI systems much greater than those in other sectors. For instance, degrees of error acceptable in a retailer's AI-based marketing function would likely be intolerable in precision manufacturing. Furthermore, implementing AI projects involves a degree of experimentation. Consequently, it may be difficult to determine a rate of return on investment (ROI) *a priori*, especially by comparison with more standardised investments in ICT hardware. Generally, SMEs are less able to bear risk than larger firms, so uncertainty about the ROI is a particular hindrance to AI uptake in this part of the enterprise population.

The considerations described above highlight the importance of skills for firms attempting to adopt AI. However, AI skills are everywhere scarce. Even leading tech companies in Silicon Valley report high vacancy rates in their research departments, owing to acute competition for AI-related talent. High salaries paid to capable AI researchers reflects the demand for such skills: OpenAI, a non-profit, paid its top researcher more than USD 1.9 million in 2016. AI talent is also mobile, and highly concentrated across countries. A recent estimate suggests that half of the entire AI workforce in Europe is found in just three countries: the United Kingdom, France and Germany (Linkedin Economic Graph, 2019). Furthermore, AI projects often require multidisciplinary teams with a mix of skills, which can be challenging to find. And because many talented graduates in data science and ML are drawn to work on novel AI applications, or at the research frontier, retaining talent in industrial companies can be another difficulty. Skills shortages are unlikely to disappear in the near term, given the many years needed to fully train AI specialists (Bergeret, 2019).

Companies face the question of how best to access the expertise needed to advance AI use. For many companies, turning to universities or public research organisations might not be a first choice. Uncertainties about the match in understanding of business needs, ownership of intellectual property (IP), operational timeframes, or other concerns, can make this route unattractive to some firms. Firms might turn to providers of management consultancy services, but for SMEs these services could be excessively expensive, and might give rise to concerns regarding dependence on the service provider. Some mid-sized and larger industrial companies have decided to create their own in-house AI capabilities, but this path is generally limited to companies with significant financial and other resources. This overall environment highlights the importance of public, or public-private, institutions to help accelerate technology diffusion (see section "Technology diffusion" below).

AI: Specific policies

Perhaps the two most important areas where governments can assist in the uptake of AI concern support for the development of skills, and the funding and operational practices of institutions for technology diffusion. Both of these policy areas are discussed later in this chapter. This subsection focuses on issues relating to training data, and measures to address hardware constraints. Later subsections also refer to relevant issues in connection with rules for IP, and research support. Many other policies – not addressed here – are most relevant to the (still uncertain) consequences of AI. These include policies for competition; economic and social policies that mitigate inequality; and measures that affect public perceptions of AI. Well-designed policies for AI are likely to have high returns because AI can be widely applied and accelerate innovation (Cockburn, Henderson and Stern, 2018). Some of the policies concerned – such as those affecting skills – are also relevant to any new technology.

Governments can take steps to help firms generate value from their data

Many firms hold valuable data, but don't use them effectively. They may lack in-house skills and knowledge, a corporate data strategy or data infrastructure, among other reasons. This can be the case even in firms with enormous financial resources. For example, by some accounts, less than 1% of the data generated on oil rigs are used (The Economist, 2017).

However, non-industrial sources of expertise – including many AI start-ups, universities and other institutions – could create value from data held by industrial firms. To help address this mismatch, governments can act as catalysts and honest brokers for data partnerships. Among other measures, they could work with relevant stakeholders to develop voluntary model agreements for trusted data sharing. For example, the US Department of Transportation has prepared the draft "Guiding Principles on Data Exchanges to Accelerate Safe Deployment of Automated Vehicles". The Digital Catapult in the United Kingdom also plans to publish model agreements for start-ups entering into data-sharing agreements (DSAs).

Government agencies can co-ordinate and steward DSAs for AI purposes

Government agencies can co-ordinate and steward DSAs for AI purposes. DSAs operate between firms, and between firms and public research institutions. In some cases, all data holders would benefit from data sharing. However, individual data holders are often reluctant to share data unilaterally – it might be of strategic importance to a company, for instance – or remain unaware of potential data-sharing opportunities. For example, 359 offshore oil rigs were operational in the North Sea and the Gulf of Mexico as of January 2018. AI-based prediction of potentially costly accidents on oil rigs would be improved if this statistically small number of data holders were to share their data. In fact, the Norwegian Oil and Gas Association asked all members to have a data-sharing strategy in place by the end of 2018. In such cases, government action could be helpful. Another example where DSAs might be useful relates to data in supply chains. Suppliers of components to an original equipment manufacturer (OEM) might improve a product using data on how the product performs in production. Absent a DSA, the OEM might be reluctant to share such data, even if doing so could benefit both parties.

The Digital Catapult's Pit Stop open-innovation activity, which complements its model DSAs, is an example of co-ordination between data holders and counterparts with expertise in data analysis. Pit Stop brings together large businesses, academic researchers and start-ups in collaborative problem-solving challenges around data and digital technologies. The Data Study Group at the Turing Institute, also in the United Kingdom, enables major private- and public-sector organisations to bring data science problems for analysis. The partnership is mutually beneficial. Institute researchers work on real-world problems using industry datasets, while businesses have their problems solved and learn about the value of their data.

Governments can promote open data initiatives

Open data initiatives exist in many countries, covering diverse public administrative and research data. To facilitate AI applications, disclosed public data should be machine-readable. In addition, in certain situations, copyright laws could allow data and text mining. The laws would need to prevent that use of AI does not lead to substitution of the original works or unreasonably prejudice legitimate interests of the copyright owners. Governments can also promote the use of digital data exchanges[5] that share public and private data for the public good. Public open data initiatives usually provide access to administrative and other data that are not directly relevant to AI in industrial companies. Nevertheless, some data could be of value to firms, such as national, regional or other economic data relevant to demand forecasts. Open science could also facilitate industrial research (see Chapter 3).

Technology itself may offer novel solutions to use data better for AI purposes

Governments should be alert to the possibilities of using AI technology in public open data initiatives. Sharing data can require overcoming a number of institutional barriers. Data holders in large organisations, for example, can face considerable internal obstacles before receiving permission to release data. Even with a DSA, data holders worry that data might not be used according to the terms of an agreement, or that client data will be shared accidentally. In addition, some datasets may be too big to share in practical ways: for instance, the data in 100 human genomes could consume 30 terabytes (30 million megabytes).

Uncertainty over the provenance of counterpart data can hinder data sharing or purchase, but approaches are being developed to address this concern and incentivise secure data exchange. For example, Ocean Protocol, created by the non-profit Ocean Protocol Foundation, combines blockchain and AI (Ocean Protocol, n.d.). Data holders can obtain the benefits of data collaboration, with full control and verifiable audit. Under one use case, data are not shared or copied. Instead, algorithms go to the data for training purposes, with all work on the data recorded in the distributed ledger. Ocean Protocol is building a reference open-source marketplace for data, which users can adapt to their own needs to trade data services securely.

Governments can also help resolve hardware constraints for AI applications

AI entrepreneurs might have the knowledge and financial resources to develop a proof-of-concept for a business. However, they may lack the necessary hardware-related expertise and hardware resources to build a viable AI company. To help address such issues, Digital Catapult runs the Machine Intelligence Garage programme. It works with industry partners such as GPU manufacturer NVidia, intelligent processing unit-producer Graphcore and cloud providers Amazon Web Services and Google Cloud Platform. Together, they give early-stage AI businesses access to computing power and technical expertise. Policies addressing hardware constraints in start-ups might not directly affect industrial companies, but they could positively shape the broader AI ecosystem in which industrial firms operate.

Blockchain in production

Blockchain – a distributed ledger technology (DLT) – has many potential applications in production (Box 5.2). Blockchain is still an immature technology, and many applications are only at the proof-of-concept stage. The future evolution of blockchain involves various unknowns, including with respect to standards for interoperability across systems. However, similar to the “software as a service” model, companies such as Microsoft, SAP, Oracle, Hewlett-Packard, Amazon and IBM already provide “blockchain as a service”. Furthermore, consortia such as Hyperledger and the Ethereum Enterprise Alliance are developing open-source DLTs in several industries (Figueiredo do Nascimento, Roque Mendes Polvora and Sousa Lourenco, 2018).

Adopting blockchain in production creates several challenges: blockchain involves fundamental changes in business processes, particularly with regard to agreements and engagement among actors in a supply chain. When many computers are involved, the transaction speeds may also be slower than some alternative processes (however, fast protocols operating on top of blockchain are under development). Blockchains are most appropriate when disintermediation, security, proof of source and establishing a chain of custody are priorities (Vujinovic, 2018). A further challenge is that much blockchain development remains atomised. Therefore, the scalability of any single blockchain-based platform – be it in supply chains or financial services – will depend on whether it can operate with other platforms (Hardjano, Lipton and Pentland, 2018).

Blockchain: Possible policies

Regulatory sandboxes help governments better understand a new technology and its regulatory implications. At the same time, they enable industry to test new technology and business models in a live environment. Evaluations of the impacts of regulatory sandboxes are sparse; one exception is FCA (2017), even if this assessment covers only the first year of a scheme in the United Kingdom. Blockchain regulatory sandboxes mostly focus on Fintech. They are being developed in countries as diverse as Australia, Canada, Indonesia, Japan, Malaysia, Switzerland, Thailand and the United Kingdom (Figueiredo do Nascimento, Roque Mendes Polvora and Sousa Lourenco, 2018). The scope of sandboxes could be broadened to encompass blockchain applications in industry and other non-financial sectors. The selection of participants needs to avoid benefiting some companies at the expense of others.

By using blockchain in the public sector, governments could raise awareness of blockchain's potential, when it improves on existing technologies. However, technical issues need to be resolved, such as how to trust the data placed on the blockchain. Trustworthy data may need to be certified in some way. Blockchain may also raise concerns for competition policy. Some large corporations, for example, may mobilise through consortia to establish blockchain standards, e.g. for supply-chain management.

Box 5.2. Blockchain: Potential applications in production

By providing a decentralised, consensus-based, immutable record of transactions, blockchain could transform important aspects of production when combined with other technologies. Several examples are listed below:

- A main application of blockchain is tracking and tracing in supply chains. One consequence could be less counterfeiting. In the motor-vehicle industry alone, firms lose tens of billions of dollars a year to counterfeit parts (Williams, 2013).
- Blockchain could replace elements of enterprise resource-planning systems. The Swedish software company IFS has demonstrated how blockchain can be integrated with enterprise resource-planning systems in the aviation industry. Commercial aircraft have millions of parts. Each part must be tracked, and a record kept of all maintenance work. Blockchain could help resolve failures in such tracking (Mearian, 2017).
- Blockchain is being tested as a medium permitting end-to-end encryption of the entire process of designing, transmitting and printing three-dimensional (3D) computer-aided design (CAD) files. The goal is that each printed part embody a unique digital identity and memory (Figueiredo do Nascimento, Roque Mendes Polvora and Sousa Lourenco, 2018). If successful, this technology could incentivise innovation using 3D printing, protect IP and help address counterfeiting.
- By storing the digital identity of every manufactured part, blockchain could provide proof of compliance with warranties, licences and standards in production, installation and maintenance (Figueiredo do Nascimento, Roque Mendes Polvora and Sousa Lourenco, 2018).
- Blockchain could induce more efficient use of industrial assets. For example, a trusted record of the usage history for each machine and piece of equipment would help develop a secondary market for such assets.
- Blockchain could help monetise the IoT, authenticating machine-based data exchanges and implementing associated micro-payments. In addition, recording machine-to-machine exchanges of valuable information could lead to "data collateralisation". This could give lenders the security to finance supply chains and help smaller suppliers overcome working-capital shortages (Maerian, 2017). By providing verifiably accurate data across production and distribution processes, blockchain could also enhance predictive analytics.
- Blockchain could further automate supply chains through the digital execution of "smart contracts", which rely on pre-agreed obligations being verified automatically. Maersk, for example, is working with IBM to test a blockchain-based approach for all documents used in bulk shipping. Combined with ongoing developments in the IoT, such smart contracts might eventually lead to full transactional autonomy for many machines (Vujinovic, 2018).

3D printing

3D printing is expanding rapidly, thanks to falling printer and materials prices, higher-quality printed objects and innovation in methods. For example, 3D printing is possible with novel materials, such as glass, biological cells and even liquids (maintained as structures using nanoparticles). Robot-arm printheads allow objects to be printed that are larger than the printer itself, opening the way for automated construction. Touchless manipulation of print particles with ultrasound allows printing of electronic components sensitive to static electricity. Hybrid 3D printers combine additive manufacturing with computer-controlled machining and milling. Research is also advancing on 3D printing with materials programmed to change shape after printing.

Most 3D printing is used to make prototypes, models and tools. Currently, 3D printing is not cost-competitive at volume with traditional mass-production technologies, such as plastic injection moulding. Wider use of 3D printing depends on how the technology evolves in terms of the print time, cost, quality, size and choice of materials (OECD, 2017a). The costs of switching from traditional mass-production technologies to 3D printing are expected to decline in the coming years as production volumes grow. However, it is difficult to predict precisely how fast 3D printing will diffuse. Furthermore, the cost of switching is not the same across all industries and applications.

3D printing: Specific policies

OECD (2017a) examined policy options to enhance 3D printing's effects on environmental sustainability. One priority is to encourage low-energy printing processes (e.g. using chemical processes rather than melting material, and automatic switching to low-power states when printers are idle). Another priority is to use and develop low-impact materials with useful end-of-life characteristics (such as compostable biomaterials). Policy mechanisms to achieve these priorities include:

- targeting grants or investments to commercialise research in these directions
- creating a voluntary certification system to label 3D printers with different grades of sustainability across multiple characteristics, which could also be linked to preferential purchasing programmes by governments and other large institutions.

Ensuring legal clarity around intellectual property rights (IPRs) for 3D printing of spare parts that are no longer manufactured could also be environmentally beneficial. For example, a washing machine that is no longer in production may be thrown away because a single part is broken. A CAD file for the required part could keep the machine in operation. However, most CADs are proprietary. One solution would be to incentivise rights for third parties to print replacement parts for products, with royalties paid to the original product manufacturers.

Government can help develop the knowledge needed for 3D printing at the production frontier

Bonnin-Roca et al. (2016) observe many potential uses for metals-based additive manufacturing (MAM) in commercial aviation. However, MAM is a relatively immature technology. The fabrication processes at the technological frontier have not yet been standardised, and aviation requires high quality and safety standards. The aviation sector would benefit if the mechanical properties of printed parts of any shape, using any given feedstock on any given MAM machine, could be accurately and consistently predicted. This would also help commercialise MAM technology. Government could help develop the necessary knowledge. Specifically, the public sector could support the basic science, particularly by funding and stewarding curated databases on materials' properties. It could broker DSAs across users of MAM technology, government laboratories and academia. It could support the development of independent manufacturing and testing standards. And it could help quantify the advantages of adopting new technology by creating a platform documenting early users' experiences.

Bonnin-Roca et al. (2016) suggest such policies for the United States, which leads globally in installed industrial 3D manufacturing systems and aerospace production. However, the same ideas could apply to other countries and industries. These ideas also illustrate how policy opportunities can arise from a specific understanding of emerging technologies and their potential uses. Indeed, governments should strive to develop expertise on emerging technologies in relevant public structures. Doing so will also help anticipate possible but hard-to-foresee needs for technology regulation.

New materials and nanotechnology

Scientists are studying materials in more detail than ever before. This is due to advances in scientific instrumentation, such as atomic-force microscopes, and developments in computational simulations. Today, materials with entirely novel properties are emerging. Solids have been created with densities comparable to the density of air, for example. Composites can be super-strong and lightweight. Some materials remember their shape, repair themselves or assemble themselves into components, while others can respond to light and sound (The Economist, 2015).

The era of trial and error in material development is also ending. Powerful computer modelling and simulation of materials' structure and properties can indicate how they might be used in products. Desired properties, such as conductivity and corrosion resistance, can be intentionally built into new materials. Better computation is leading to faster development of new and improved materials, more rapid insertion of materials into new products, and improved processes and products. In the near future, engineers will not only design products, but also the materials from which products are made (Teresko, 2008). Furthermore, large companies will increasingly compete in terms of materials development. For example, a manufacturer of automotive engines with a superior design could enjoy longer-term competitive advantage if it also owned the material from which the engine is built.

Closely related to new materials, nanotechnology involves the ability to work with phenomena and processes occurring at a scale of 1 to 100 nanometres (nm) (a standard sheet of paper is about 100 000 nm thick). Control of materials on the nanoscale – working with their smallest functional units – is a general-purpose technology with applications across production (Friedrichs, 2017). Advanced nanomaterials are increasingly used in manufacturing high-tech products, e.g. to polish optical components.

New materials and nanotechnology: Specific policies

No single company or organisation will be able to own the entire array of technologies associated with materials innovation. Accordingly, a public-private investment model is warranted, particularly to build cyber-physical infrastructure and train the future workforce (McDowell, 2017).

New materials will raise new policy issues and give renewed emphasis to a number of longstanding policy concerns. New digital security risks could arise. For example, a computationally assisted materials "pipeline" based on computer simulations could be hackable. Progress in new materials also requires effective policy in already important areas, often related to the science-industry interface. For example, well-designed policies are needed for open data and open science. Such policies could facilitate sharing or exchanges of modelling tools and experimental data, and simulations of materials' structures, among other possibilities.

Professional societies are developing a materials-information infrastructure to provide decision support to materials-discovery processes (Robinson and McMahon, 2016). This includes databases of materials' behaviour, digital representations of materials' microstructures and predicted structure-property relations, and associated data standards. International policy co-ordination is needed to harmonise and combine elements of cyber-physical infrastructure across a range of European, North American and Asian investments and capabilities. It is too costly (and unnecessary) to replicate resources that can be accessed through web services. A culture of data sharing – particularly pre-competitive data – is required (McDowell, 2017).

Sophisticated and expensive tools are also needed for research in nanotechnology. State-of-the-art equipment costs several million euros and often requires bespoke buildings. It is almost impossible to gather an all-encompassing nanotechnology research and development (R&D) infrastructure in a single institute, or even a single region. Consequently, nanotechnology requires inter-institutional and/or international collaboration to reach its full potential (Friedrichs, 2017). Publicly funded R&D programmes should allow involvement of academia and industry from other countries. They should also enable flexible collaborations between the most suitable partners. The Global Collaboration initiative under the European Union's Horizon 2020 programme is an example of this approach.

Support is needed for innovation and commercialisation in small companies. Nanotechnology R&D is mostly conducted by larger companies for three reasons. First, they have a critical mass of R&D and production. Second, they can acquire and operate expensive instrumentation. Third, they are better able to access and use external knowledge. Policy makers could improve access to equipment of small and medium-sized enterprises (SMEs) by increasing the size of SME research grants; subsidising or waiving service fees; and/or providing SMEs with vouchers for equipment use.

Regulatory uncertainties regarding risk assessment and approval of nanotechnology-enabled products must also be addressed, ideally through international collaboration. These uncertainties severely hamper the commercialisation of nano-technological innovation. Policies should support the development of transparent and timely guidelines for assessing the risk of nanotechnology-enabled products. At the same time, they should strive for international harmonisation in guidelines and enforcement. In addition, more needs to be done to properly treat nanotechnology-enabled products in the waste stream (Friedrichs, 2017).

Selected cross-cutting policy issues

This section addresses cross-cutting policies relevant to all the digital technologies described above. The issues examined are technology diffusion, connectivity and data, standards-setting processes, digital skills, access to and awareness of high-performance computing (HPC), IP systems and public support for R&D.

Technology diffusion

Most countries, regions and companies are primarily technology users, rather than technology producers. For them, technology diffusion and adoption should be priorities. Even in the most advanced economies diffusion can be slow or partial. For example, a survey of 4 500 German businesses in 2015 found that only 4% had implemented digitalised and networked production processes or planned to do so (ZEW-IKT, 2015). Similarly, a survey of SME manufacturers in the United States in 2017 found that 77% had no plans to deploy the IoT (Sikich, 2017).

Policies that broaden technology diffusion not only help to raise labour productivity growth, they might also lower inequality in rates of wage growth. Policy makers tend to acknowledge the critical importance of technology diffusion at a high level. However, they may overlook technology diffusion in the overall allocation of attention and resources (Shapira and Youtie, 2017).

Certain features of new digital technologies could make diffusion more difficult. Potential technology users must often evaluate large and growing amounts of information on rapidly changing technologies and the skills and other inputs they require. Even the initial step of collecting sensor data can be daunting. A typical industrial plant, for example, might contain machinery of many vintages from different manufacturers. In turn, these could have control and automation systems from different vendors, all operating with different communication standards. And whereas many prior digital production technologies enhanced pre-existing processes, blockchain could entail a more challenging redesign of business models.

Diffusion in SMEs involves particular difficulties

An important issue for diffusion-related institutions is that small firms tend to use key technologies less frequently than larger firms. In Europe, for example, 36% of surveyed companies with 50-249 employees use industrial robots, compared to 74% of companies with 1 000 or more employees (Fraunhofer, 2015). Only 16% of European SMEs share electronic supply-chain data, compared to 29% of large enterprises. This discrepant pattern of technology use directly reflects the availability of skills. For instance, only around 15% of European SMEs employ information and communication technology (ICT) specialists, compared to 75% of large firms (EC, 2017) (Box 5.3).

Box 5.3. Diffusing technology to SMEs: Some key considerations

Various steps can be taken to help diffuse technology to SMEs, including the following:

It is important to systematise key information for SMEs. A number of countries have developed tools to help SMEs transform technologically. Germany's Industry 4.0 initiative has documented over 300 uses cases of applications of digital industrial technologies. It also includes contacts to experts (www.plattform-i40.de). And the United Kingdom's 2017 Mayfield Commission led to the creation of an online self-assessment tool. It gives firms a benchmark against best practice, with guidelines on supporting actions (www.bethebusiness.com). Information provided through such initiatives also needs to encompass AI.

Particularly useful is information on the expected return on investment (ROI) in new technologies, as well as information on essential complementary organisational and process changes. One international survey asked 430 professionals working across industry sectors what could help them implement intelligent business strategy in their organisation. More than half (56%) wanted more information linking initiatives to ROI (AI Intelligent Automation Network, 2018). But careful thinking and exposition of this information is needed. Ezell (2018) notes that an ROI may be hard to calculate when the technology frontier is expanding. ROIs for some AI projects may be particularly hard to determine *a priori*, in part because data cleaning – which involves an element of art – is key to the outcomes of most AI investments. Investment decisions may also have to include strategic considerations such as the need to remain viable in future supply chains.

Because the skills to absorb information are scarce in many SMEs, simply providing information on technology is not enough. Providing signposts to reliable sources of SME-specific expertise can help. For example, as part of its SMEs Go Digital Programme, Singapore's TechDepot provides a list of pre-approved digital technology and service solutions suited to SMEs. Targeted skills development is also useful. For instance, Tooling U-SME – an American non-profit organisation owned by the Society of Manufacturing Engineers – provides online industrial manufacturing training and apprenticeships.

Test beds can also provide SMEs with facilities to test varieties and novel combinations of digital and other equipment. In this way, they can de-risk prospective investments.

Diffusion requires conditions to support the creation of growth-oriented start-ups and efficient allocation of economic resources

By ensuring conditions such as timely bankruptcy procedures and strong enforcement of contracts, governments can support the creation of businesses. Increasing new-firm entry and growth is important for diffusion. OECD research has highlighted the role of new and young firms in net job creation and radical innovation. Unconstrained by legacy systems, start-ups often introduce forms of organisation that new technologies require. Electric dynamos, for example, were first commercialised in the mid-1890s during the second industrial revolution. It took almost four decades, and a wave of start-up and investment activity in the 1920s, before suitably reorganised factories became widespread and productivity climbed (David, 1990).

Recent OECD analysis of micro-economic allocation processes highlights the importance for leading-edge production of conducive economic and regulatory framework conditions. These conditions include competitive product markets and flexible labour markets. Low costs for starting and closing a business are also important. Furthermore, openness to foreign direct investment and trade provides a vehicle for technology diffusion and an incentive for technology adoption. Such conditions all facilitate efficient resource allocation. Efficient resource allocation helps incumbent firms and start-ups adopt new technologies and grow. Andrews, Criscuolo and Gal (2016) estimate that more liberal markets, especially in services, could avoid up to half of the difference in multi-factor productivity between "frontier" and "laggard" firms, and accelerate diffusion of new organisational models.

Several additional factors can aid diffusion. These include openness to internationally mobile skilled labour, and the strength of knowledge exchange within national economies. A key such exchange is the interaction between scientific institutions and businesses.

Institutions for diffusion can also be effective if well designed

In addition to enabling framework conditions, effective institutions for technology diffusion are also important. Innovation systems invariably contain multiple sources of technology diffusion, such as universities and professional societies. Shapira and Youtie (2017) provide a typology of diffusion institutions. It ranges from applied technology centres (e.g. the Fraunhofer Institutes in Germany) to open technology mechanisms (e.g. the Bio-Bricks Registry of Standard Biological Parts). Some institutions involved, such as technical extension services, tend to receive low priority in the standard set of innovation support measures. But they can be effective if well designed. For example, the United States' Manufacturing Extension Partnership has recently been estimated to return USD 14.5 per dollar of federal funding.

New diffusion initiatives are emerging, some of which are still experimental. For instance, alongside established applied technology centres, such as the Fraunhofer Institutes, partnership-based approaches are increasing. An example is the US National Network for Manufacturing Innovation (NNMI). The NNMI uses private non-profit organisations as the hub of a network of company and university organisations to develop standards and prototypes in areas such as 3D printing and digital manufacturing and design.

Technology diffusion institutions need realistic goals and time horizons

Upgrading the ability of manufacturing communities to absorb new production technologies takes time. More effective diffusion is likely when technology diffusion institutions are empowered and resourced to take longer-term perspectives. Similarly, evaluation metrics should emphasise longer-run capability development rather than incremental outcomes and revenue generation.

Introducing new ways to diffuse technology takes experimentation. Yet many governments want quick and riskless results. Policy making needs better evaluation evidence and a readiness to experiment with organisational designs and practices. Concerns over governmental accountability combined with ongoing public austerity in many economies could mean that institutions will be reluctant to risk change, slowing the emergence of next-generation institutions for technology diffusion (Shapira and Youtie, 2017).

Policies on connectivity and data

Broadband networks are essential to Industry 4.0. They reduce the cost of accessing information and expand the means for sharing data and knowledge. In this way, they help develop new goods, services and business models and facilitate research. Policy priorities in this area include furthering access to high-speed broadband networks, including in rural and remote areas, and overhauling laws governing the speed and coverage of communication services (OECD, 2017b). Fibre-optic cable is of particular importance for Industry 4.0 (Box 5.4).

Policies to promote competition and private investment, as well as independent and evidence-based regulation, have helped to extend coverage. When market forces cannot fulfil all policy objectives, governments can respond with a series of tools. These could include competitive public tenders for infrastructure deployment, legal obligations on operators, and subsidies for national and municipal broadband networks.

Other measures include fostering open access arrangements and initiatives to reduce deployment costs. “Dig once” practices, for example, mandate installation of fibre conduits in publicly funded road projects (OECD, 2018b). Technological developments are also likely to expand opportunities for providing services in underserved areas. For example, broadband could be delivered through “White Spaces”, the gaps in radio spectrum between digital terrestrial television channels.

Box 5.4. The importance of fibre-optic cable for Industry 4.0

Fibre-optic connectivity is important for Industry 4.0, and has numerous advantages over copper-cable based Internet. Fibre-optic cable provides faster speed, with a current upper range of 100 gigabytes per second. It provides faster access to cloud-hosted information, along with greater reliability, signal strength and bandwidth. Its lower latency is important for many digitally controlled machines, for collaboration among employees and for accommodating new technologies such as haptics (which remotely replicate a sense of touch). It improves security because the signal is lost during breaches of fibre-optic cable. It resists interference, stemming, for example, from proximity to machinery. Moreover, 5G networks rely on fibre connectivity.

Enhancing trust in digital services is critical to data sharing and the uptake of broadband. Industry 4.0 also creates risks that could erode the perceived benefits of digital technologies. While challenging to measure, digital security incidents appear to be increasing in terms of sophistication, frequency and influence (OECD, 2017b). In one 2014 incident, hackers breached the office computers of a German steel mill and overrode the shut-off mechanisms on the steel mill’s blast furnace (Long, 2018).

Such incidents affect firms’ reputations and competitiveness. They also impose significant costs on the economy as a whole, restricting ICT adoption and business opportunities. New digital security solutions are emerging. In homomorphic encryption, for example, data are always encrypted, even when being computed on in the cloud. But the technological race between hackers and their targets is continuous. And SMEs, in particular, need to introduce or improve their digital security risk management practices.

Restricting cross-border data flows should be avoided

Research is beginning to show that restricting data flows can lead to lost trade and investment opportunities, higher costs of cloud and other information technology services, and lower economic productivity and gross domestic product growth (Cory, 2017). Manufacturing creates more data than any other sector of the economy. Cross-border data flows are expected to grow faster than growth in world trade. Restricting such flows, or making them more expensive, for instance by obliging companies to process customer data locally, can raise firms’ costs and increase the complexity of doing business, especially for SMEs.

A prospective policy issue: Legal data portability rights for firms?

In April 2016, the European Union’s General Data Protection Regulation established the right to portability for personal data. A number of companies, such as Siemens and GE, are vying for leadership in online platforms for the IoT. As digitalisation proceeds, such platforms will become increasingly important repositories of business data. If companies had portability rights for non-personal data, competition among platforms could grow, and switching costs for firms could fall.

A prospective policy issue: Frameworks to protect non-personal sensor data

The protection of machine-generated data is likely to become a growing issue as Industry 4.0 advances. This is because sensors are becoming ubiquitous, more capable, increasingly linked to embedded computation, and used to stream large volumes of often critical machine data. Single machines may contain multiple component parts made by different manufacturers, each equipped with sensors that capture, compute and transmit data. These developments raise legal and regulatory questions. For instance, are special provisions needed to protect data in value chains from third parties? Which legal entities should have ownership rights of machine-generated data under what conditions? And, what rights to ownership of valuable data should exist in cases of business insolvency?

Increasing trust in cloud computing

Cloud computing is another technology where policy might be needed. Cloud use can bring efficiency gains for firms. And Industry 4.0 will require increased data sharing across sites and company boundaries.[6] Consequently, machine data and data analytics and even monitoring and control systems will increasingly be situated in the cloud. The cloud will also enable independent AI projects to start small, and scale up and down as required. Indeed, Google’s Chief AI scientist, Fei-Fei Li, recently argued that cloud computing will democratise AI.[7]

Governments can act to increase trust in the cloud and stimulate cloud adoption. The use of cloud computing in manufacturing varies greatly across OECD countries. In Finland, 69% of manufacturers use the cloud, for example, compared to around 15% in Germany. Firms in countries where cloud use is low often cite fears over data security and uncertainty about placing data in extra-territorial servers. However, cloud use can bring increased data security, especially for SMEs. For example, Amazon Web Services, a market leader, reportedly provides more than 1 800 security controls. This affords a level of data security beyond what most firms could themselves provide. Government could take steps, for example, to help SMEs better understand the technical and legal implications of cloud service contracts. This could include providing information on the scope and content of certification schemes relevant for cloud computing customers.

Developing digital skills

Digital technologies create new skills needs. Occupational titles like “industrial data scientist”, and “bioinformatics scientists” are recent, reflecting technology-driven changes in skills demand. Individuals need the necessary basic skills to adopt new digital technologies. The lack of generic analytic skills and advanced skills is hindering technology adoption. For instance, surveys show that a shortage of skilled data specialists is a main impediment to the use of data analytics in business (OECD, 2017b).

Concern is widespread regarding possible labour market disruptions from automation driven by digital technology. Data from the OECD Programme for International Assessment of Adult Competencies highlight a lack of ICT skills in low-skilled adult populations in semi-skilled occupations. This means this demographic group is at high risk of losing jobs to automation.

Forecasting skills needs is hazardous. Just a few years ago, few would have foreseen that smartphones would disrupt, and in some cases end, a wide variety of products and industries, from notebook computers and personal organisers to niche industries making musical metronomes and hand-held magnifying glasses (functions now available through mobile applications).

Because foresight is imperfect, governments must establish systems that draw on the collective information and understanding available regarding emerging needs for skills. In that regard, businesses, trade unions, educational institutions and learners can all contribute. Students, parents and employers must have access to information with which to judge how well educational institutions perform and assess the career paths of graduates of different programmes. In turn, educational and training systems must be organised such that resources flow efficiently to courses and institutions that best cater to the demand for skills. Institutions like Sweden's job security councils, or the SkillsFutureSingapore agency, play such roles (Atkinson, 2018). And business and government must work together to design training schemes, with public authorities ensuring the reliability of training certification.

How learning is delivered matters greatly

Policies for improving skills for Industry 4.0 typically include fostering ICT literacy in school curricula. This literacy ranges from use of basic productivity software such as word processing programmes and spreadsheets, to coding and even digital security courses. Throughout formal education, more multidisciplinary programmes and greater curricular flexibility are often required. For instance, students should be able to select a component on mechanical engineering and combine this with data science, bio-based manufacturing, or other disciplines.

In a comprehensive review of science, technology, engineering and mathematics (STEM) education, Atkinson and Mayo (2010) identify a series of priorities. These emphasise helping students follow their interests and passions; respecting the desire of younger students to be active learners; and giving greater opportunity to explore a wide variety of STEM subjects in depth. Equally important are increasing the use of online, video-game and project-based learning, and creating options to take tertiary-level STEM courses at secondary level. Japan's Kosen schools have proven the efficacy of many of these ideas since the early 1960s (Schleicher, 2018).

Many governments are implementing forward-looking programmes to match ICT training priorities with expected skills needs. In Belgium, for example, the government carries out prospective studies on the expected impact of the digital transformation on occupations and skills in a wide variety of fields. The results are then used to select training courses to be reinforced for emerging and future jobs (OECD, 2017b). Estonia and Costa Rica have also changed school curricula based on where they estimate jobs will be in the future.

Lifelong learning must be an integral part of work

Advancing automation and the birth of new technologies also mean that lifelong learning must be an integral part of work. Each year, inflows to the labour force from initial education represent only a small percentage of the numbers in work, who in turn will bear much of the cost of adjustment to new technologies. Both considerations underscore the importance of widespread lifelong learning. Disruptive changes in production technology highlight the importance of strong and widespread generic skills, such as literacy, numeracy and problem solving. These foundation skills are the basis for subsequent acquisition of technical skills, whatever they turn out to be. In collaboration with social partners, governments can help spur development of new training programmes, such as conversion courses in AI for those already in work.

Digital technology will itself affect how skills are developed

Digital technology is creating opportunities to develop skills in novel ways. For example, in 2014, Professor Ashok Joel, and graduate students, at Georgia Tech University, created an AI teaching assistant – Jill Watson – to respond to online student questions. For months students were unaware that the responses were non-human (Korn, 2016). iTalk2Learn is a European Union project to develop an open-source intelligent mathematics tutoring platform for primary schools. Closer to the workplace, researchers at Stanford University are developing systems to train crowdworkers using machine-curated material generated by other crowdworkers. And Upskill (www.upskill.io) provides wearable technology to connect workers to the information, equipment, processes and people they need in order to work more efficiently. Among other potential benefits, in a world where lifelong learning will be essential, AI could help learners understand the idiosyncrasies of how they learn best.

Participation in standards-setting processes

Advanced production operates in a vast matrix of technical standards. The semiconductor industry, for example, uses over 1 000 standards (Tassey, 2014). Standards development relevant to Industry 4.0 is underway in many fields. These range from machine-to-machine communication and data transmission to 5G (a global standard for which is expected by 2019), robotics and digital identifiers for objects. Over 100 standards initiatives exist today for the IoT and Industry 4.0 (Ezell, 2018).

Countries and firms that play primary roles in setting international standards can enjoy advantages if new standards align with their own national standards and/or features of their productive base. The public sector's role should be to encourage industry, including firms of different sizes, to participate at early stages in international (and in some cases national) standards setting. Dedicated support could be given to include under-represented groups of firms in processes to develop standards.

The development of AI standards – particularly technical standards – is at a very early stage so far. Most national AI strategies refer to the development of AI ethics standards. But this oversight dimension of standards, around ethics and corporate governance, also needs technical standards (a term like "algorithmic transparency" doesn't yet have a technical definition). The timing of standards setting – too soon or too late – is always an issue raised when assessing how standards affect innovation. In the past, often, just a few main players negotiated standards. But now there are large numbers of developers working on Open Source projects that will also find standards solutions. In some areas of AI, who defines a standard first may be less important than with previous technologies.

Improving access to high-performance computing

HPC is increasingly important for firms in industries ranging from construction and pharmaceuticals to the automotive sector and aerospace. Airbus, for instance, owns 3 of the world's 500 fastest supercomputers. Two-thirds of US-based companies that use HPC say that: "increasing performance of computational models is a matter of competitive survival" (US Council on Competitiveness, 2014). How HPC is used in manufacturing is also expanding, going beyond applications such as design and simulation to include real-time control of complex production processes. Financial rates of return to HPC use are high. By one estimate, each EUR 1 invested generates, on average, EUR 69 in profits (EC, 2016). A 2016 review observed that

> *"(m)aking HPC accessible to all manufacturers in a country can be a tremendous differentiator, and no nation has cracked the puzzle yet" (Ezell and Atkinson, 2016).*

Box 5.5. Getting supercomputing to industry: Possible policy actions

- Raise awareness of industrial use cases, with quantification of their costs and benefits.
- Develop a one-stop source of HPC services and advice for SMEs and other industrial users.
- Provide low-cost, or free, limited experimental use of HPC for SMEs, with a view to demonstrating the technical and commercial implications of the technology.
- Establish online software libraries or clearing houses to help disseminate innovative HPC software to a wider industrial base.
- Give incentives for HPC centres with long industrial experience, such as the Hartree Centre in the United Kingdom, or TERATEC in France, to advise centres with less experience.
- Modify eligibility criteria for HPC projects, which typically focus on peer review of scientific excellence, to include criteria of commercial impact.
- Engage academia and industry in the co-design of new hardware and software, as has been done in European projects such as Mont Blanc (Mont Blanc, n.d.).
- Include HPC in university science and engineering curricula.
- Explore opportunities for co-ordinating the purchase of commercially provided computing capacity.

As Industry 4.0 becomes more widespread, demand for HPC will rise. But like other digital technologies, the use of HPC in manufacturing falls short of potential. One estimate is that 8% of US firms with fewer than 100 employees use HPC. However, half of manufacturing SMEs could use HPC (for prototyping, testing and design) (Ezell and Atkinson, 2016). Public HPC initiatives often focus on the computation needs of “big science”. Greater outreach to industry, especially SMEs, is frequently needed. Ways forward – a number of which are described in EC (2016) – are set out in Box 5.5.

Even developing countries may be advised to have a backbone network of high-performance computers. Initially, a low-income economy may have few sophisticated industrial uses for HPC. However, high-performance computers can find initial applications in research and science, and then later be applied in industry. Cloud-based supercomputing cannot meet all supercomputing needs. This is only viable when applications are needed occasionally. If industry or scientific applications are regular or continuous, then a cloud-based service may be too expensive.

Intellectual property systems

Digital technologies are raising new challenges for IP systems. 3D printing, for example, might create complications in connection with patent eligibility. For instance, if 3D printed human tissue improves upon natural human tissue, it may be eligible for patenting, even though naturally occurring human tissue is not. More fundamentally, new patenting frameworks may be needed in a world where machines have the ability to invent. AI systems have already created patentable inventions (OECD, 2017a).

AI raises many complex challenges for IP systems, such as identifying infringements of patent laws. These laws will be complicated by AI systems that automatically – and unpredictably – learn from many publicly available sources of information (Yanisky-Ravid and Liu, 2017). An overarching policy challenge is to balance the needs around IP. On the one hand, IP is necessary for incentivising certain types of innovation. On the other, it should not hamper diffusion of technologies such as AI and 3D printing.

Public support for R&D

The complexity of many emerging production technologies exceeds the research capacities of even the largest firms. In such cases public-private research partnerships may be needed. Microelectronics, new materials and nanotechnology, among others, have arisen because of advances in scientific knowledge and instrumentation. Publicly financed basic research has often been critical. For decades, for example, public funding supported progress in AI, including during unproductive periods of research, to the point where AI today attracts huge private investment (National Research Council, 1999). Recent declines in public support for research in some major economies is a concern.

Many possible targets exist for government R&D and commercialisation efforts. As discussed below, these range from quantum computing (Box 5.6), to advancing AI.

An overarching research challenge relates to computation itself

Processing speeds, memory capacities, sensor density and accuracy of many digital devices are linked to Moore’s Law. This asserts that the number of transistors on a microchip doubles about every two years (Investopedia, n.d.). However, atomic-level phenomena and rising costs constrain further shrinkage of transistors on integrated circuits.

Many experts believe a limit to miniaturisation will soon be reached. At the same time, applications of digital technologies across the economy rely on increasing computing power. For example, the computing power needed for the largest AI experiments is doubling every three-and-a-half months (OpenAI, 16 May 2018). By one estimate, this trend can be sustained for at most three-and-a-half to ten years, even assuming public R&D commitments on a scale similar to the Apollo or Manhattan projects (Carey, 10 July 2018).

Much, therefore, depends on achieving superior computing performance (including in terms of energy requirements). Many hope that significant advances in computing will stem from research breakthroughs in optical computing (using photons instead of electrons), biological computing (using DNA to store data and calculate) and/or quantum computing (Box 5.6).

Box 5.6. A new computing regime: The race for quantum computing

Quantum computers function by exploiting the laws of subatomic physics. A conventional transistor flips between on and off, representing 1s and 0s. However, a quantum computer uses quantum bits (qubits), which can be in a state of 0, 1 or any probabilistic combination of both 0 and 1 (for instance, 0 with 20% and 1 with 80% probability). At the same time, qubits interact with other qubits through so-called quantum entanglement (which Einstein termed "spooky action at a distance").

Fully developed quantum computers, featuring many qubits, could revolutionise certain types of computing. Many of the problems best addressed by quantum computers, such as complex optimisation and vast simulation, have major economic implications. For example, at the 2018 CogX Conference, Dr Julie Love, Microsoft's director of quantum computing, described how simulating all the chemical properties of the main molecule involved in fixing nitrogen – nitrogenase – would take today's supercomputers billions of years. Yet this simulation could be performed in hours with quantum technology. The results of such a simulation would directly inform the challenge of raising global agricultural productivity and limiting today's reliance on the highly energy-intensive production of nitrogen-based fertiliser. Rigetti Computing has also demonstrated that quantum computers can train ML algorithms to a higher accuracy, using fewer data than with conventional computing (Zeng, 22 February 2018).

Until recently, quantum technology has mostly been a theoretical possibility. However, Google, IBM and others are beginning to trial practical applications with a small number of qubits (Gambetta, Chow and Teffen, 2017). For example, IBM Quantum Experience (IBM, n.d.) offers free online quantum computing. However, no quantum device currently approaches the performance of conventional computers.

By one estimate, fewer than 100 people globally possess the skills to write algorithms specifically for quantum computers. Azhar (2018) calculates that companies involved in any aspect of quantum computing employ fewer than 2 000 people globally. Skill constraints may be lessened by Google's release of Cirq. This software toolkit allows developers without specialised knowledge of quantum physics to create algorithms for quantum machines (Giles, 2018a). Zapata Computing, a start-up, aims to offer a range of ready-made software that firms can use on quantum computers (Giles, 2018b).

The further development of robust, scalable quantum computing involves major research and engineering challenges. Global annual public investment in quantum computing could range from EUR 1.5 billion to EUR 1.9 billion. While relatively small, venture capital funding is growing, led by D-Wave (USD 175 million), Rigetti (USD 70 million), Cambridge Quantum Computing (USD 50 million) and IonQ (USD 20 million) (Azhar, 2018). The People's Republic of China is scheduled to open a National Laboratory for Quantum Information Sciences in 2020, with a projected investment of USD 10 billion. Chinese scientists are making major research advances. In July 2018, for instance, they broke a record for the number of qubits linked to one another through quantum entanglement (Letzer, 2018).

A need for more – and possibly different – research on AI

Public research funding has been key to progress in AI since the origin of the field. The National Research Council (1999) shows that while the concept of AI originated in the private sector – in close collaboration with academia – its growth largely results from many decades of public investments. Global centres of AI research excellence (e.g. at Stanford, Carnegie Mellon and the Massachusetts Institute of Technology) arose because of public support, often linked to US Department of Defense funding. However, recent successes in AI have propelled growth in private sector R&D for AI. For example, earnings reports indicate that Google, Amazon, Apple, Facebook and Microsoft spent a combined USD 60 billion on R&D in 2017, including an important share on AI. By comparison, total US federal government R&D for non-defence industrial production and technology amounted to around USD 760 million in 2017 (OECD, 2019).

Many in business, government and among the public believe AI stands at an inflection point, ready to achieve major improvements in capability. However, some experts emphasise the scale and difficulties of the outstanding research challenges. Some AI research breakthroughs could be particularly important for society, the economy and public policy. However, corporate and public research goals might not fully align. Jordan (2018) notes that much AI research is not directly relevant to the major challenges of building safe intelligent infrastructures, such as medical or transport systems. He observes that unlike human-imitative AI, such critical systems must have the ability to deal with:

> *"…distributed repositories of knowledge that are rapidly changing and are likely to be globally incoherent. Such systems must cope with cloud-edge interactions in making timely, distributed decisions and they must deal with long-tail phenomena whereby there is (sic) lots of data on some individuals and little data on most individuals. They must address the difficulties of sharing data across administrative and competitive boundaries." (Jordan, 2018)*

Other outstanding research challenges relevant to public policy relate to making AI explainable; making AI systems robust (image-recognition systems can easily be misled, for instance); determining how much prior knowledge will be needed for AI to perform difficult tasks (Marcus, 2018); bringing abstract and higher-order reasoning, and "common sense", into AI systems; and inferring and representing causality. Jordan (2018) also identifies the need to develop computationally tractable representations of uncertainty. No reliable basis exists for judging when – or whether – research breakthroughs will occur. Indeed, past predictions of timelines in the development of AI have been extremely inaccurate.

Research and industry can often be linked more effectively

Government-funded research institutions and programmes should be free to combine the right partners and facilities to address challenges of scale-up and interdisciplinarity. Investments are often essential in applied research centres and pilot production facilities to take innovations from the laboratory into production. Demonstration facilities such as test beds, pilot lines and factory demonstrators are also needed. These should provide dedicated research environments with the right mix of enabling technologies and the technicians to operate them. Some manufacturing R&D challenges may need expertise from manufacturing engineers and industrial researchers, as well as designers, equipment suppliers, shop-floor technicians and users (O'Sullivan and López-Gómez, 2017).

More effective research institutions and programmes in advanced production may also need new evaluation indicators. These would go beyond traditional metrics such as numbers of publications and patents. Additional indicators might also assess such criteria as successful pilot line and test-bed demonstration, training of technicians and engineers, consortia membership, the incorporation of SMEs in supply chains and the role of research in attracting FDI.

Conclusion

New digital technologies are key to the next production revolution. Realising their full potential requires effective policy in wide-ranging fields, including skills, technology diffusion, data, digital infrastructure, research partnerships, standards and IPRs. Typically, these diverse policy fields are not closely connected in government structures and processes. Governments must also adopt long-term time horizons, for instance, in pursuing research agendas with possible long-term payoffs. Public institutions must also possess specific understanding of many fast-evolving digital technologies. One leading authority argues that converging developments in several technologies are about to yield a “Cambrian explosion” in robot diversity and use (Pratt, 2015). Adopting Industry 4.0 poses challenges for firms, particularly small ones. It also challenges governments’ ability to act with foresight and technical knowledge across multiple policy domains.

References

AI Intelligent Automation Network (2018), “AI 2020 : The global state of intelligent enterprise”, webpage, https://www.aiia.net/events-intelligentautomation-chicago/downloads/ai-2020-the-global-state-of-intelligent-enterprise (accessed 9 July 2019).

Andrews, D., C. Criscuolo and P.N. Gal (2016), “The global productivity slowdown, technology divergence and public policy”, *Hutchins Center Working Paper* # 24, September. https://www.brookings.edu/wp-content/uploads/2016/09/wp24_andrews-et-al_final.pdf.

Atkinson, R.D. and Ezell, S. (2019), “The manufacturing evolution: how AI will transform manufacturing and the workforce of the future”, Information Technology and Innovation Foundation, Washington, DC: https://itif.org/publications/2019/08/06/manufacturing-evolution-how-ai-will-transform-manufacturing-and-workforce.

Atkinson, R.D. (2018), “How to reform worker-training and adjustment policies for an era of technological change”, Information Technology and Innovation Foundation, Washington, DC: http://www2.itif.org/2018-innovation-employment-workforce-policies.pdf.

Atkinson, R.D. and M. Mayo (2010), “Refueling the U.S. innovation economy: Fresh approaches to science, technology, engineering and mathematics (STEM) Education”, Information Technology and Innovation Foundation, Washington, DC: https://www.itif.org/files/2010-refueling-innovation-economy.pdf.

Azhar, A. (2018), “Exponential view: Dept. of quantum computing”, The Exponential View, 15 July, www.exponentialview.co/evarchive/#174.

Bergeret, B. (2019), “AI and Europe’s medium-sized firms: How to overcome an Achilles heel”, *OECD Observer*, http://oecdobserver.org/news/fullstory.php/aid/6259/AI_and_Europe_92s_medium-sized_firms:_How_to_overcome_an_Achilles_heel.html.

Bonnin-Roca, J. et al. (2016), “Policy needed for additive manufacturing”, *Nature Materials*, Vol. 15, Nature Research, Springer, pp. 815-818, https://doi.org/10.1038/nmat4658.

Bughin, J. et al. (2017), “Artificial intelligence: The next digital frontier?” *Discussion Paper*, June, McKinsey Global Institute, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

Carey, R. (10 July 2018), “Interpreting AI compute trends”, AI blog, https://aiimpacts.org/interpreting-ai-compute-trends/.

Champain, V. (2018), “Comment l’intelligence artificielle augmentée va changer l’industrie”, *La Tribune*, Paris, 27 March, www.latribune.fr/opinions/tribunes/comment-l-intelligence-artificielle-augmentee-va-changer-l-industrie-772791.html.

Chen, S. (2018), “Scientists are using AI to painstakingly assemble single atoms”, *Science*, American Association for the Advancement of Science, Washington, DC, 23 May, www.wired.com/story/scientists-are-using-ai-to-painstakingly-assemble-single-atoms/.

Chen, S. (2017), “The AI company that helps Boeing cook new metals for jets”, *Science*, American Association for the Advancement of Science, Washington, DC, 12 June, www.wired.com/story/the-ai-company-that-helps-boeing-cook-new-metals-for-jets/.

Chui, M. et al. (2018), “Notes from the AI frontier: Insights from hundreds of use cases”, *Discussion Paper*, April, McKinsey & Company, New York, April, www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning.

Cockburn, I., R. Henderson and S. Stern (2018), "The impact of artificial intelligence on innovation", in A. Agrawal, J. Gans and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.

Cory, N. (2017), "Cross-border data flows: Where are the barriers and what do they cost ?", Information Technology and Innovation Foundation, Washington, DC, https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost.

David, P.A. (1990), "The Dynamo and the computer: An historical perspective on the modern productivity paradox", *American Economic Review*, Volume 80, Issue 2, pp.355-361.

Digital Catapult (2018), "Machines for machine intelligence: Providing the tools and expertise to turn potential into reality", *Machine Intelligence Garage, Research Report 2018,* London, www.migarage.ai.

Dorfman, P. (2018), "3 Advances changing the future of artificial intelligence in manufacturing", *Autodesk Newsletter*, 3 January, www.autodesk.com/redshift/future-of-artificial-intelligence/.

EC (2017), "Europe's digital progress report 2017", https://ec.europa.eu/digital-single-market/en/news/europes-digital-progress-report-2017.

EC (2016), "Implementation of the Action Plan for the European High-Performance Computing Strategy", *Commission Staff Working Document*, SWD(2016)106, European Commission, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016SC0106.

Ezell, S. (2018), "Why Manufacturing Digitalization Matters and How Countries Are Supporting It", Information Technology and Innovation Foundation, Washington, DC, www2.itif.org/2018-manufacturing-digitalization.pdf (accessed 21 January 2019).

Ezell, S.J. and R.D. Atkinson (2016), "The vital importance of high-performance computing to US competitiveness", Information Technology and Innovation Foundation, Washington, DC, www2.itif.org/2016-high-performance-computing.pdf.

Faggella, D. (2018), "Industrial AI applications – How time series and sensor data improve processes", 31 May, Techemergence, San Francisco, www.techemergence.com/industrial-ai-applications-time-series-sensor-data-improve-processes/.

FCA (2017), "Regulatory sandbox lessons learned report", Financial Conduct Authority, London, www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf.

Figueiredo do Nascimento, S., A. Roque Mendes Polvora and J. Sousa Lourenco (2018), "#Blockchain4EU: Blockchain for industrial transformations", Publications Office of the European Union, Luxembourg, https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/blockchain4eu-blockchain-industrial-transformations.

Fraunhofer (2015), "Analysis of the impact of robotic systems on employment in the European Union", https://ec.europa.eu/digital-single-market/news/fresh-look-use-robots-shows-positive-effect-automation.

Friedrichs, S. (2017), "Tapping nanotechnology's potential to shape the next production revolution", in *The Next Production Revolution: Implications for Governments and Business,* OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-8-en.

Gambetta, J.M., J.M. Chow and M. Teffen (2017), "Building logical qubits in a superconducting quantum computing system", *npj Quantum Information*, Vol. 3/2, Nature Publishing Group and University of New South Wales, London and Sydney, https://doi.org/10.1038/s41534-016-0004-0.

Giles, M. (2018a), "Google wants to make programming quantum computers easier", *MIT Technology Review*, 18 July, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/611673/google-wants-to-make-programming-quantum-computers-easier/.

Giles, M. (2018b), "The world's first quantum software superstore – or so it hopes – is here", *MIT Technology Review*, 17 May, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/611139/the-worlds-first-quantum-software-superstore-or-so-it-hopes-is-here/.

Goodfellow, I., Y. Bengio and A. Courville (2016), *Deep Learning*, MIT Press, Cambridge, Massachusetts.

Harbert, T. (2013), "Supercharging patent lawyers with AI: How Silicon Valley's Lex Machina is blending AI and data analytics to radically alter patent litigation", *IEEE Spectrum*, 30 October, Institute of Electrical and Electronics Engineers, New York, https://spectrum.ieee.org/geek-life/profiles/supercharging-patent-lawyers-with-ai.

Hardjano, T., A. Lipton and A.S. Pentland (2018), "Towards a design philosophy for interoperable blockchain systems", 7 July, Massachusetts Institute of Technology, Cambridge, https://hardjono.mit.edu/sites/default/files/documents/hardjono-lipton-pentland-p2pfisy-2018.pdf.

House of Lords (2018), "AI in the UK: Ready, willing and able?", *Select Committee on Artificial Intelligence – Report of Session 2017-19*, HL Paper No. 100, Authority of the House of Lords, London, https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.

IBM (n.d.), "Quantum", webpage, www.research.ibm.com/quantum (accessed 9 July 2019).

Investopedia (n.d.), "Moore's Law", webpage, www.investopedia.com/terms/m/mooreslaw.asp (accessed 10 May 2019).

Jordan, M. (2018), "Artificial intelligence – The revolution hasn't happened yet", *Medium*, 19 April, San Francisco, https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7.

Korn, M. (2016), "Imagine discovering that your teaching assistant really is a robot", *The Wall Street Journal*, 6 May, https://www.wsj.com/articles/if-your-teacher-sounds-like-a-robot-you-might-be-on-to-something-1462546621.

Küpper, D. et al. (2018), "AI in the factory of the future: The ghost in the machine", 18 April, Boston Consulting Group, www.bcg.com/publications/2018/artificial-intelligence-factory-future.aspx.

Letzer, R. (2018), "Chinese researchers achieve stunning quantum entanglement record", *Scientific American*, 17 July, Springer Nature, www.scientificamerican.com/article/chinese-researchers-achieve-stunning-quantum-entanglement-record/.

Linkedin Economic Graph (2019),"Talent in the European Labour Market", Linkedin Economic Graph, November, https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/reference-cards/research/2019/AI-Talent-in-the-European-Labour-Market.pdf.

Long, L. (2018), "The most pressing challenge modern manufacturers face? Cybersecurity", Engineering.com, 30 April, https://www.engineering.com/AdvancedManufacturing/ArticleID/16856/The-Most-Pressing-Challenge-Modern-Manufacturers-Face-Cybersecurity.aspx.

Marcus, G. (2018), "Innateness, AlphaZero and artificial intelligence", arxiv.org, Cornell University, Ithaca, United States, https://arxiv.org/ftp/arxiv/papers/1801/1801.05667.pdf.

McDowell, D.L. (2017), "Revolutionising product design and performance with materials innovation", in *The Next Production Revolution: Implications for Governments and Business,* OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-10-en.

Mearian, L. (2017), "Blockchain integration turns ERP into a collaboration platform", *Computerworld*, 9 June, IDG, Framingham, United States, www.computerworld.com/article/3199977/enterprise-applications/blockchain-integration-turns-erp-into-a-collaboration-platform.html.

Mont Blanc (n.d.), "Mont Blanc project", webpage, http://montblanc-project.eu/ (accessed 9 July 2019).

National Research Council (1999), *Funding a Revolution: Government Support for Computing Research*, The National Academies Press, Washington, DC, https://doi.org/10.17226/6323.

Ocean Protocol (n.d.), Ocean Protocol website, www.oceanprotocol.com (accessed 9 July 2019).

OECD (2019), *OECD Main Science and Technology Indicators* (database), http://oe.cd/msti (accessed 10 July 2019).

OECD (2018a), *Going Digital in a Multilateral World,* Interim Report of the OECD Going Digital Project, Meeting of the OECD Council at Ministerial Level, Paris, 30-31 May 2018, OECD, Paris, www.oecd.org/going-digital/C-MIN-2018-6-EN.pdf.

OECD (2018b), "Bridging the rural digital divide", *OECD Digital Economy Papers*, No. 265, OECD Publishing, Paris, http://dx.doi.org/10.1787/852bd3b9-en.

OECD (2017a), *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-en.

OECD (2017b), *OECD Digital Economy Outlook 2017,* OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264276284-en.

OpenAI (16 May 2018), "AI and compute", OpenAI blog, San Francisco, https://blog.openai.com/ai-and-compute/.

O'Sullivan, E. and C.López-Gómez (2017), "An international review of emerging manufacturing R&D priorities and policies for the next production revolution", in *The Next Production Revolution: Implications for Governments and Business,* OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-14-en.

Pratt, G.A. (2015), "Is a Cambrian explosion coming for robotics?", *Journal of Economic Perspectives*, Volume 29/3, American Economic Association, Pittsburgh, pp. 51-60, https://doi.org/10.1257/jep.29.3.51.

Ransbotham, S. et al. (2017), "Reshaping business with artificial intelligence: Closing the gap between ambition and action", *MIT Sloan Management Review*, Massachusetts Institute of Technology, Cambridge, https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/.

Robinson, L. and K. McMahon (2016), "TMS launches materials data infrastructure study," *JOM*, Vol. 68/8, Springer, New York, pp. 2014-2016, https://doi.org/10.1007/s11837-016-2011-1.

Schleicher, A. (2018), "How Japan's schools are creating a new generation of innovators", 13 March, http://oecdeducationtoday.blogspot.com/2018/03/japan-kosen-school-innovation-technology.html.

Shapira, P and J. Youtie (2017), "The next production revolution and institutions for technology diffusion", in *The Next Production Revolution: Implications for Governments and Business,* OECD Publishing, Paris, https://doi.org/10.1787/9789264271036-11-en.

Sikich (2017), "2017 manufacturing report", www.sikich.com/wp-content/uploads/2017/09/SKCH-Manufacturing-Report-2017-08-17.pdf.

Simonite, T. (2016), "Algorithms that learn with less data could expand AI's power", *MIT Technology Review*, 24 May, Massachusetts Institute of Technology, Cambridge, www.technologyreview.com/s/601551/algorithms-that-learn-with-less-data-could-expand-ais-power/.

Sverdlik, Y. (2018), "Google is switching to a self-driving data center management system", 2 August, Data Center Knowledge, www.datacenterknowledge.com/google-alphabet/google-switching-self-driving-data-center-management-system.

Tassey, G. (2014), "Competing in advanced manufacturing: The need for improved growth models and policies", *Journal of Economic Perspectives*, Vol. 28, No.1, pp 27-48.

Teresko, J. (2008), "Designing the next materials revolution", *IndustryWeek*, 8 October, Informa, Cleveland, www.industryweek.com/none/designing-next-materials-revolution.

The Economist (2017), "Oil struggles to enter the digital age", *The Economist*, 6 April, London, www.economist.com/business/2017/04/06/oil-struggles-to-enter-the-digital-age.

The Economist (2015), "Material difference", Technology Quarterly, 12 May, *The Economist*, London, www.economist.com/technology-quarterly/2015-12-05/new-materials-for-manufacturing.

US Council on Competitiveness (2014), "The Exascale effect: The benefits of supercomputing for US industry", US Council on Competitiveness, Washington, DC, www.compete.org/storage/images/uploads/File/PDF%20Files/Solve_Report_Final.pdf.

Vujinovic, M. (2018), "Manufacturing and blockchain: Prime time has yet to come", *CoinDesk*, 24 May, www.coindesk.com/manufacturing-blockchain-prime-time-yet-come/.

Walker, J. (2017), "AI in mining: Mineral exploration, autonomous drills, and more", *Techemergence*, 3 December, www.techemergence.com/ai-in-mining-mineral-exploration-autonomous-drills/.

Williams, M. (2013), “Counterfeit parts are costing the industry billions”, *Automotive Logistics*, 1 January, Ultima Media, London, https://automotivelogistics.media/intelligence/16979.

Wissner-Gross, A. (2016), “Datasets over algorithms”, *Edge.org*, Edge Foundation, Seattle, www.edge.org/response-detail/26587.

Yanisky-Ravid, S. and X.Liu (2017), “When artificial intelligence systems produce inventions: The 3A era and an alternative model for patent law”, *39 Cardozo Law Review*, 2215-2263 (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931828.

Zeng, W. (22 February 2018), “Forest 1.3: Upgraded developer tools, improved stability, and faster execution”, Rigetti Computing blog, https://medium.com/rigetti/forest-1-3-upgraded-developer-tools-improved-stability-and-faster-execution-561b8b44c875.

ZEW-IKT (2015), “Industrie 4.0: Digitale (R)Evolution der Wirtschaft” [Industry 4.0: Digital (r)evolution of the economy], ZEW, Mannheim, http://ftp.zew.de/pub/zew-docs/div/IKTRep/IKT_Report_2015.pdf.

Zweben, M. and M.S. Fox (1994), *Intelligent Scheduling*, Morgan Kaufmann Publishers, San Francisco.

Notes

[1] Deep learning with artificial neural networks is a technique in the broader field of machine learning (ML) that seeks to emulate how human beings acquire certain types of knowledge. The word “deep” refers to the numerous layers of data processing. The term “artificial neural network” refers to hardware and/or software modelled on the functioning of neurons in a human brain.

[2] AI will, of course, have many economic and social impacts. In relation to labour markets alone, intense debates exist on AI’s possible effects on labour displacement, income distribution, skills demand and occupational change. However, these and other considerations are not a focus of this chapter.

[3] In the development of improved forms of AI, increased data availability has been critical. Over the past 30 years, the length of time between data creation and the most publicised AI breakthroughs has been much shorter than between algorithmic progress and the same breakthroughs (Wissner-Gross, 2016). Using a variant of an algorithm developed 25 years earlier, for example, Google’s GoogLeNet software achieved near-human level object classification in 2014. But the software was trained on ImageNet, a huge corpus of labelled images and object categories that had become available just four years earlier (at its peak, ImageNet reportedly employed close to 50 000 people in 167 countries, who sorted around 14 million images [House of Lords, 2018]).

[4] Many tools that firms employ to manage and use AI exist as free software in open source (i.e. their source code is public and modifiable). These include software libraries such as TensorFlow and Keras, and tools that facilitate coding such as GitHub, text editors like Atom and Nano, and development environments like Anaconda and RStudio. Machine learning-as-a-service platforms also exist, such as Michelangelo – Uber’s internal system that helps teams build, deploy and operate ML solutions.

[5] An example of a data exchange is datacollaboratives.org.

[6] For example, Ezell (2018) reports that “BMW has set a goal of knowing the real-time status of all major production equipment at each company that produces key components for each of its vehicles”.

[7] See Professor Li’s full remarks at the 2017 Global StartupGrind Conference: https://www.startupgrind.com/blog/cloud-will-democratize-ai/.

6 Digitalisation in the bioeconomy: Convergence for the bio-based industries

James Philp

Directorate for Science, Technology and Innovation, OECD

Chapter 6 focuses on digitalisation and the bio-based industries that are starting to make impacts in the chemicals and materials sectors. As a result of next-generation genome sequencing, biology and biotechnology have become data-rich. Developing bioprocesses has often been hampered at the biological stage – the efficiency of the production strain or biocatalyst. The new discipline of synthetic biology or engineering biology is ushering in an era of more precise control of construction of DNA parts, genes, and all the way to production strains. Engineering biology needs digitalisation and vice versa. The bioeconomy is wider than biotechnology, however. There are many other ways that converging technologies and digitalisation can be applicable to the bioeconomy.

Introduction

In essence, the bioeconomy is about using renewable feedstocks to produce everyday goods and services. The bioeconomy concept has expanded well beyond the boundaries set in the OECD (2009) publication The Bioeconomy to 2030: Designing a policy agenda. It now encompasses a wide range of sectors and activities including chemicals, food, agriculture, dairy, forestry, pulp and paper, waste management and others. The bioeconomy is not just concerned with biotechnology. It is now seen as a new means of production that will gradually replace fossil-based production and be consistent with the concept of a circular economy (Philp and Winickoff, 2018).

Synthetic biology is an interdisciplinary field that aims to design and make biological parts and systems. It has to become an engineering discipline to take its place in future advanced manufacturing. If synthetic biology goes beyond the domain of science, many of the potential impacts linked to successful manufacturing will be achieved.

There is optimism in the future of synthetic biology. Biology has gone from being a data-poor discipline to being data-rich, which makes biology amenable to much greater computational analysis. And where there are algorithms, there is the possibility for automation. Automation brings faster “design-build-test” cycles, which will go a long way to conquering two of the long-term challenges of biotechnology, namely the lack of reproducibility and reliability.

The whole bioeconomy business cycle is ripe for digitalisation. This includes extraction and procurement of materials, as well as logistics and distribution of intermediate goods. It also comprises the retail of final products to consumers, including, as envisioned in a circular economy, the reuse, repair and recycling of products and materials.

At the heart of the bioeconomy’s future is the need for a different kind of workforce with multi- and interdisciplinary skills. Among several other key attributes, professionals in future bio-based industries will need to be much more familiar with digital skills such as programming and data science. This chapter illustrates there is still much to do, even in educating a future “biomechatronics-ready” workforce to drive this far-reaching but still-to-be-achieved manufacturing sector.

The great convergence

This section is concerned with examples of how digitalisation and biotechnology can work together. Together, they can provide solutions to major bioeconomy policy goals that could not be tackled by either alone. This can be seen as a form of convergence, which OECD defines as the coming together of different technologies to solve problems that cannot be addressed by a single technology.

The combination of digital and biological transformation may greatly change the design and handling of production processes and their products. A workshop of the Global Bioeconomy Summit of 2018 in Berlin was entitled “The great convergence: Digitalisation, biologicalisation and the future of manufacturing”. It described how “bio-intelligent value adding” could be disruptive in future manufacturing.

While this form of convergence is usually considered a future potential, a form of convergence of special interest to the bioeconomy is already functioning. This is the mix of industrial biotechnology with green chemistry (Philp, Ritchie and Allan, 2013). This chapter addresses some aspects of this convergence and provides examples.

Why is convergence necessary?

This subsection explores the need for convergence. Box 6.1 summarises the challenges and policies needed to bring synthetic or engineering biology into advanced manufacturing. Convergence is becoming a necessity for business survival. Sean Ward, Chief Technology Officer of Synthace in the United Kingdom has said, "As working with the physical world is becoming increasingly digital, every company that is out there is discovering that they either are a technology business or they are dead. And that is what is happening with biology: it is becoming a technology business" (Quaglia, 20 February 2017).

Throughout the history of biology, experimentation has been difficult due to a scarcity of data. That situation has changed dramatically this century as technological improvements in experimental high-throughput (HT) measurement have made biology data-rich. This has created a need for tools to facilitate the analysis and interpretation of biological data (Fong, 2014). In the data-rich age, predictive design and rapid evaluation are at the core of any engineering (synthetic) biology approach. These accompany assembly of new materials through laboratory automation, HT characterisation and post-production processing.

In the earliest years of bio-based production, it took 50-300 person years and many millions of dollars to bring a metabolically engineered product to market (Hong and Nielsen, 2012; Carlson, 2018). Even recently, it took on average over seven years to launch a bio-based product (Il Bioeconomista, 10 June 2015). The earliest commercial successes of such products were achieved without the full advantages of rich data. If a deluge of metabolically engineered microorganisms producing bio-based chemicals was subsequently expected, then that deluge has not arrived. Follow-on commercial successes have been few (e.g. Van Dien, 2013). Some progress has been made, however. For example, commercial scale production of 1,4-BDO (an organic compound) was performed less than five years after the first detectable amount of BDO was produced in an engineered *E. coli* strain (Burgard et al., 2016).

Box 6.1. Concepts to unite manufacturing and materials-discovery communities to harness opportunities emerging from synthetic biology

The unifying concepts are:

- platform technologies to support the delivery of synthetic biological materials
- a highly trained interdisciplinary workforce
- academic/industry/government co-development that can implement and innovate these technologies
- standardisation and interoperability of biological parts for new materials
- sustainable materials manufacturing and management
- a common language and vision that places synthetic biology at the nexus of other disciplines, especially materials science, chemistry, computer science and engineering.

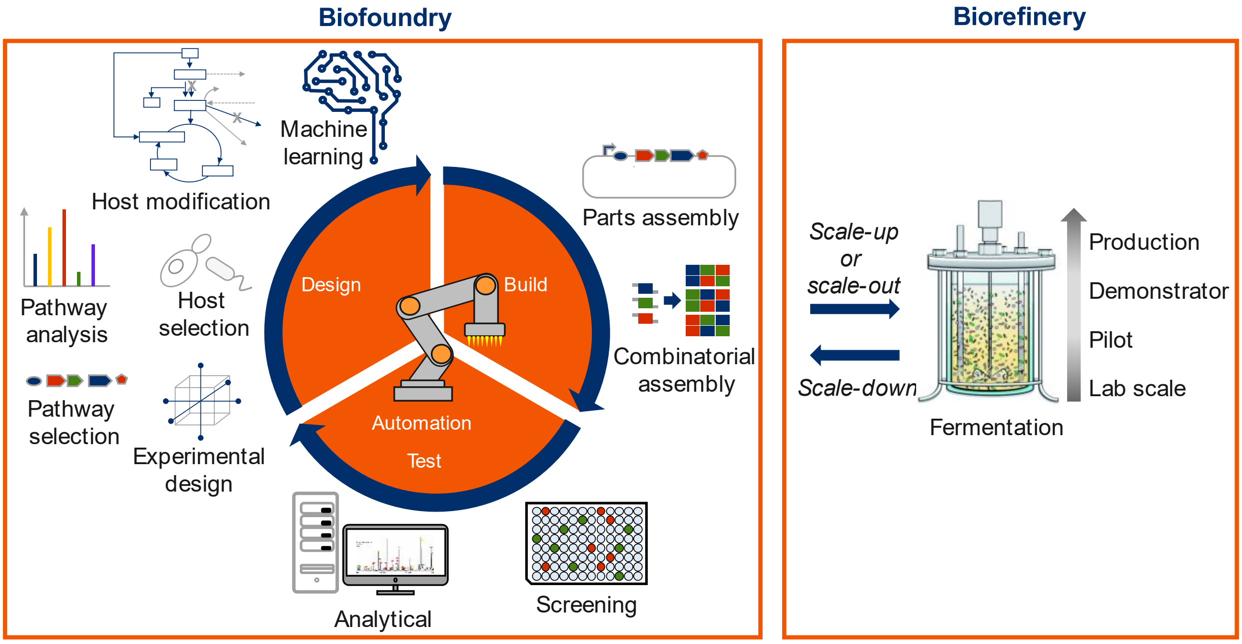
Source: Le Feuvre and Scrutton (2018), "A living foundry for synthetic biological materials: A synthetic biology roadmap to new advanced materials".

Overarching view: Greater integration of biotechnology with the engineering design cycle

At a fundamental level, most biotechnology as yet fails to meet some of the specific criteria of engineering. Essential differences between the scientific method (test a hypothesis through experimentation) and engineering design (design a solution to a problem and test the outcome) must be addressed. Concepts such as interoperability, separation of design from manufacture, standardisation of parts and systems, all of which are central to engineering disciplines, have been largely absent from biotechnology (OECD, 2014). Therefore, weaknesses can be expected at the level of the engineering cycle, depicted in a generic way in Figure 6.1.

Many variants on the engineering cycle exist, but Figure 6.1 shows the basic elements through phases of initial design, building and testing of a part/system/device. No one expects an optimal design on the first attempt. Thereafter, the process is iterated as often as is necessary to meet the engineering specifications.

Figure 6.1. The engineering design cycle



Source: Kitney et al. (2019), "Enabling the advanced bioeconomy through public policy supporting biofoundries and engineering biology", https://www.cell.com/trends/biotechnology/pdf/S0167-7799(19)30076-9.pdf.

The test phase is the current bottleneck

Within the engineering design cycle, the test phase is the primary bottleneck – a challenge that will only be solved through biology and automation of the iterative processes. Evaluation of an organism's phenotype – its observable physical properties – is a major rate-limiting step in metabolic engineering (Wang, 2014). When constructing production strains for biofuels or bio-based chemicals, design success will be measured by the amount of product formed. This may require separation of individual strains and determination of the concentration of the chemical of interest produced by each. If so, the process of multiplexing (bringing many input streams into one) in design and build has been defeated. In effect, this results in demultiplexing (breaking one input stream into many) (Rogers and Church, 2016).

This is where an important bottleneck remains – orders of magnitude fewer constructs can be tested than can be designed and built. The throughput is limited to hundreds of thousands of design evaluations per day. Improving this throughput by mechanical or electronic automation will be limited as the orders of magnitude of improvement needed are so high. The needed advances must come from biology itself (e.g. Rogers et al., 2015; Xiao et al., 2016), but are ultimately linked to automation of the iterative processes.

An integrated technology platform could unlock the potential

Integrating engineering design with biotechnology could unlock commercial potential, especially when combined with digitalisation and automation.

Genomatica, an American company, is a leader in producing bio-based chemicals from metabolically engineered strains. In its view, the key to removing bottlenecks is: "an integrated technology platform encompassing metabolic modelling, HT pathway and strain construction, quantitative small-scale screening, and systems biology, all of which are intimately linked to fermentation and process engineering" (Burgard et al., 2016).

This agrees with the view of Lee and Kim (2015). They believe one reason the process is so challenging is “that researchers often fail to consider a fully integrated industrial bioprocess when developing microbial strains with new activities”. They refer to this as the “systems metabolic engineering framework”.

The integration of technologies, especially to enable multiple iterations of design and construction of strains, could typically benefit from digitalisation and automation. The incorporation of artificial learning and artificial intelligence (AI) would remove the need for laborious, time-consuming human intervention between iterations. For example, the large number of metabolic engineering studies could provide an invaluable database. This source could capture information on titre (the concentration), yield and productivity in response to genetic and fermentation conditions. These data, in turn, could be built into machine-learning models, which increasingly remove human involvement in the design-build-test cycle. The day should come when the results of one round of “test” iteration should inform the next round of “design” without human intervention.

Accelerated discovery of natural biological materials (see section 3.9 on the Earth BioGenome Project) is required to explore the diversity of materials and provide access to new materials properties that are lacking. The further development of next-generation deoxyribonucleic acid (DNA) sequencing and DNA synthesis is vital to such efforts. Research programmes could embrace these new technologies to give access to the potential power of vast libraries of biological materials (natural and synthetic) to create the materials and composites of the future.

Regarding these libraries of biological materials, Hadadi et al. (2016) used computational tools to construct a database of all theoretical biochemical reactions based on known biochemical principles and compounds. This database complements projects such as the Earth BioGenome, which would open up the many “unknowns”. The database includes more than 130 000 hypothetical enzymatic reactions that connect two or more metabolites through novel enzymatic reactions. These reactions have never been reported in living organisms. Through the database, users can search for all possible routes from any substrate compound to any product.

Reproducibility is a continuing problem

The essence of the reproducibility problem is that design tools for process-based research and development (R&D) are inadequate. Increasingly, life science, and chemical and food product development have become a global supply chain of people, instruments, organisations, knowledge and data. This supply chain must be orchestrated to deliver an increasingly complex portfolio of products, while meeting intensifying cost and regulatory pressures. Integrating software therefore needs to go far beyond integrating in the laboratory: integration across the entire business is the best way to reduce errors.

A recent survey identified reproducibility as an issue for design for a majority of respondents. The survey concerning scientific reproducibility achieved responses of 1 576 researchers, of which 703 were biologists. More than half pointed to insufficient replication in the lab, poor oversight or low statistical power. Physicists and chemists were the most confident of the reproducibility of their scientific literature. When respondents were asked how best to address the reproducibility issue, nearly 90% – more than 1 000 people – ticked “more robust experimental design”, “better statistics” and “better mentorship” (Baker, 2016).

Early in the history of synthetic biology, Kwok (2010) highlighted reproducibility as a challenge and it remains so (e.g. Hayden, 2015; Beal et al., 2016). This challenge has to be conquered for bio-based manufacturing to become a credible manufacturing platform of the future.

Many researchers have called for completely new computational languages for biotechnology. They argue that variants of natural languages such as English are too imprecise and ambiguous to tackle the highly complex systems of biology and biotechnology. Antha is perhaps the first bona fide attempt to create a programming language for general-purpose computation in biology (Sadowski, Grant and Fell, 2016). It is built on Google’s Go programming language, but incorporates domain-specific features, such as liquid handling planning. Antha is claimed to enable experiments of an entirely new level of complexity. It embraces the departure from experimenting by changing one-factor-at-a-time enshrined in the scientific method, by enabling detection of interactions between different experimental factors.

The creator of Antha, Synthace of London, exemplifies the challenge of reproducibility. Synthace worked with Merck to create a new microbial manufacturing platform for bio-therapeutics. They examined the interactions between 27 factors to integrate strain construction with process development. This is far too complex and time-consuming to address with a screening approach. Even screening a billion assays a second would result in impossible time periods to investigate every permutation of these 27 genetic and process factors. Using multifactorial methods, the system navigated this space, revealing key factor interactions in a small fraction of the time.

Reliability, predictability and reproducibility

Reliability and predictability are two other facets affecting reproducibility. The challenge of designing a fully predictable gene network is preventing engineering biology from realising its full economic impact. Many areas of engineering have confronted and solved similar challenges. Key to resolving this situation is the automation of the design-build-test bio-based engineering cycle.

Robust and predictable scale-up is also necessary for success in biological manufacturing. Scale-up presents new and different challenges compared to laboratory-scale design. For example, a microbial production strain needs to be robust to function in an industrial-scale fermentation process: what works in a laboratory has every chance of failing in a 10 000 litre fermenter.

Only a few examples have deliberately employed synthetic biology to increase robustness in bio-based production. To this end, the United States' Defense Advanced Research Projects Agency (DARPA) has a research programme on Biological Robustness in Complex Settings (BRICS). BRICS is pursuing the fundamental understanding and component technologies needed to transition engineering biology systems from well-defined laboratory environments into more complex settings. In this new environment, they can achieve greater biomedical, industrial, and strategic potential.

Automation can help address test-phase obstacles

Automation in synthetic biology promises to clear bottlenecks in the test phase, but it needs engineering standards to facilitate data exchange. New HT evaluation and metrology methods are needed to overcome the test-phase bottleneck. These often involve bio-imaging methods and informatics workflows that are generally automated. They depend on sophisticated software for acquisition and management of both qualitative and quantitative data.

The pursuit of automation in synthetic biology has been termed bio-design automation (BDA) (Densmore, 2012). This approach is predicated upon solving small parts of a larger problem one piece at a time. After all the necessary pieces are defined and solved, solutions for each sub-problem can be automated, connected and reused to solve larger problems (Appleton et al., 2017). This process can arguably increase abstraction and reuse, and create greatly scaled systems, in size and complexity.

One of the greatest challenges to realising BDA is the lack of engineering standards and documentation needed for repeatedly engineering these systems. All stages of the design cycle have opportunities to store and exchange data on genetic designs. Standards facilitate these data exchanges. Two of the most common standards in synthetic biology for these purposes are the Synthetic Biology Open Language and the Systems Biology Markup Language, the latter supported by more than 250 different software tools. Other standards are reviewed by Appleton et al. (2017), who also describe future needs, several of which call for open-source approaches to software development.

Similarly, most research-based pharmaceutical companies use HT screening methods. This allows simultaneous tests of hundreds of thousands of compounds against a specific model of disease. Automation with robots has been necessary to achieve levels of throughput not feasible with humans. Now, a new generation of automation is tackling even more complex functions. Known as intelligent automation, it is based on robotic process automation systems that combine process automation software and AI (KPMG, 2018).

Manufacturing in the modern economy works because design and testing software can talk to manufacturing hardware via multiple layers of application programming interfaces. This points to the need for biotechnology to have its own high-level programming language(s) and software to transform the engineering design, testing and learning cycle.

Industrial biotechnology and green chemistry convergence

This subsection explores the convergence between industrial biotechnology and green chemistry. For Le Feuvre and Scrutton (2018),

> *"(t)he conflation of synthetic biology and (combinatorial) synthetic chemistry, and exploration of potential connections with contemporary manufacturing platforms such as Additive Manufacturing (3D printing), defines a new era in the exploration of new advanced materials…"*

Digitalisation can hasten the convergence of green chemistry and industrial biotechnology. Green chemistry involves designing environmentally benign chemical processes. As such, it is one of the most important and practical tools to integrate principles of sustainable economic development into chemistry and the chemical industry (Makarova et al., 2017). Industrial biotechnology is largely about using biotechnology to produce chemicals of various types. The policy objectives of industrial biotechnology and green chemistry are, then, effectively the same. Both are "wet" sciences or technologies, and each discipline can serve the other. These shared qualities create a natural evolution towards convergence. To speed that evolution requires more than serendipity; there are clear ways in which digitalisation can hasten product development.

Chemistry can help overcome a key technical challenge that undermines production of bio-based equivalents of high-volume chemicals. Three key metrics of bioprocesses are often poorer than in petrochemistry: titre, yield and productivity. These metrics are often too low to be scalable because most natural microbial processes are incompatible with an industrial process (e.g. Harder, Bettenbrock and Klamt, 2016; Maiti et al., 2016). Chemistry can improve these metrics. In the case of ethanol, the titres and yields from fermentation are adequate. For many other chemicals this is not the case.

Some bio-based chemicals are best made from biomass using a purely chemical process. In the end, the desired result is the same. Unsustainable chemicals and materials are eventually replaced with bio-based equivalents that are sustainable and renewable. This is not simply about using chemical tools to aid biology or biology tools to aid chemistry. Rather, it is a genuine co-operation to make a better result.

Industrial biotechnology converges with chemistry and with information technology/computing

There is plenty of scope for digitalisation to enhance the production advantages of combining industrial biotechnology and green chemistry. For example, Gerbaud et al. (2017) have proposed computer-aided molecular design (CAMD) for bio-based commodity molecules. They discussed coupling CAMD tools with computer-aided organic synthesis tools for two purposes. First, they could propose enhanced bio-sourced molecules, which could be synthesised using eco-friendly pathways. Second, they could analyse their sustainability.

Data analysis and storage as bottlenecks

Conquering the challenges of the test phase and convergence will push the bottleneck into data analysis and storage. This subsection looks at using DNA to avoid the storage problem, and how policy makers can support this process.

A fully multiplexed design-build-test cycle that links phenotype to DNA sequence will enable the evaluation of millions of designs per cycle. However, this will also create an unprecedented amount of data. This, in turn, may move the production bottleneck to data storage.

In the age of ML, data should ultimately inform the next iteration of design in the absence of humans (Rogers and Church, 2016). For example, AutoBioCAD promises to design genetic "circuits" for *E. coli* with virtually no human user input (Rodrigo and Jaramillo, 2013). Thus, algorithms are needed that incorporate ML to correlate data from different data sets. The aim is to link genes, proteins and pathways without a priori knowledge (Wurtzel and Kutchan, 2016).

Is DNA storage the answer?

A crisis in data storage is looming in the next two decades as silicon-based storage methods struggle to keep pace with demand. Long-term storage is perhaps the fastest growing segment of the data storage market. In 2015 and 2016 combined, more data were created than in all of preceding history (Service, 2017). By 2040, if all data were stored for instant access, the archive would consume 10 to 100 times the expected supply of microchip-grade silicon (Zhirnov et al., 2016). Without radical change, a data crunch may be unavoidable.

DNA as a storage medium may offer a way to prevent a storage crisis. It seems far-fetched to store digital data in DNA, but it is already possible to translate digital information into genetic information. In 2016, researchers at Microsoft and the University of Washington broke the record for storing digital data in DNA. They managed to store and retrieve 200 megabytes (MB) of information (including high-definition video, multiple books and articles as well as a database) using DNA provided by Twist Bioscience (Ogunnaike, 2016). In 2018, they doubled their record to 400 MB of data on DNA. Their breakthroughs could pave the way to exabyte storage (Tung, 2018).

As an example of the possibilities of DNA storage, Shipman et al. (2017) encoded real information (images) and optimised the method of delivery, nucleotide content of the sequences and reconstruction method. They used a population of bacteria.

The storage potential of DNA vastly exceeds that of all other media. One estimate suggests all the world's data could be stored in 1 kilogramme of DNA (Extance, 2016). Another proposes that 215 petabytes (PB) (215 million gigabytes) – roughly all the information on the Internet – could be stored in a single gramme of DNA (Service, 2017).

DNA storage is much too expensive as a storage medium, as the technology is only in its formative stages. While the cost of DNA sequencing has become trivial, DNA synthesis (writing), despite reduced costs, is still too expensive for mass exploitation. It remains orders of magnitude higher than sequencing costs.[1] What needs to be done, in general terms, to commercialise DNA storage, is the following:

- Develop better algorithms to translate digital information into biological information and to enable fast, accurate and cost-efficient retrieval of information.
- Invent and advance new chemistries to enable cheap DNA synthesis.
- Incorporate more automation in production workflows to achieve cost reductions.

Public policy can help achieve all of these goals, especially in research subsidy, support for small and medium-sized enterprises and spin-outs, and policies to support technology transfer. In particular, support for automation through public foundries would be important. Reducing transaction costs by identifying fruitful public-private partnerships would also hasten progress: a leading partnership between Microsoft and the University of Washington could be a model. Research programmes that target industry-academia collaboration would be one way to build such partnerships.

Blockchain for benefit sharing and protecting sensitive information

Blockchain, which uses a highly secure, distributed database technology, holds a number of advantages for different types of life sciences projects and companies. It is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way" (Iansiti and Lakhani, 2017). The technology, with its high level of encryption and security, is at the heart of Bitcoin and other virtual currencies.

The Earth BioGenome Project (EBP) aims to sequence all the plants, animals and single-celled organisms on Earth (the eukaryotic species) within ten years to help unlock the vast economic potential of biodiversity (EBP, n.d.). As one hurdle for such an ambitious project, data sharing must balance two goals. On the one hand, it must ensure a permanent, freely available resource for future scientific discovery. On the other, it must respect the access and benefit sharing guidelines of the Nagoya Protocol[2] (Lewin et al., 2018).

The EBP aims to address the challenge of data storage. The completed project will generate around 200 PB of data. This will require new architectures, algorithms and software for improved quality, efficiency and cost-effectiveness, as well as data analysis, big data visualisation and sharing. The project is expected to promote these tools for equitable worldwide sharing of data, analytic tools and data mining resources.

Blockchain could also support traceability for benefits sharing and prevention of bio-piracy. By registering biological and biomimetic intellectual property (IP) assets on the blockchain, code banks could record the provenance, rights and obligations associated with nature's assets. This could help track their provenance and use (World Economic Forum, 2018).

Blockchain may help tackle the quite different challenges applying to the health and pharmaceuticals industry, especially around sensitive patient data. This branch of the life services industry is generating an increasing amount of sensitive data and transactions. Some have proposed that blockchain will become essential in dealing with these growing data (KPMG, 2018). Blockchain is well suited for managing areas such as supply chain, privacy, transaction processing, contracts and licensing, and sensitive medical records.

Digital security

All life sciences, whether public or private, are vulnerable to cyber-attacks. Bio-based industries that help produce chemicals and materials have similar concerns for cybersecurity as the chemicals industry. Bio-production relies heavily on data, on IP and research, all of which need protection for firms to reap the financial benefits of their investments.

The health and pharmaceuticals sector of the life sciences face these and other more specific issues, such as patient privacy. A recent survey indicated that companies are elevating cybersecurity to a strategic imperative. However, the pace of protection lags behind their desire to adopt digital technologies to drive innovation (KPMG, 2018). There are many ways to launch a cyber-attack on a bio-production company.

Many different types of organisation are involved in bio-production security. They range from feedstock suppliers and customers to information technology (IT) professionals from law firms and IP offices. Cybersecurity is only as strong as the weakest link in the overall system of protection.

Cloud computing

Life sciences companies in health and pharmaceuticals are increasingly using cloud computing to optimise complex processes with a view to reducing business costs. For example, user-based pricing models are paving the way to lower capital investment and operational costs (KPMG, 2018). Cloud-based solutions can make data available for clinical trials while meeting security and regulatory requirements. Further, the cloud enables complex data analysis from Internet of Things and real-time devices. For such reasons, cloud technology is one of the top priorities in enhancing internal efficiency.

Frontiers in bio-production

This section examines three of many different and intersecting future bio-production strategies: biofoundries, bio-based three-dimensional (3D) printing and cell-free synthetic biology. The three are described in ascending order of their expected deployment. First, as "design, build, test" iteration facilities, biofoundries are expected to drastically reduce the time and effort needed to go from idea to product. Second, bio-based 3D printing can capture the complexity of a biological entity (e.g. cell, tissue or higher form of biological specialisation such as an organ). It requires an intimate marriage of genetic and digital code to guarantee the high levels of accuracy needed. Third, cell-free synthetic biology expresses much of the control at the digital level to create cell-free biomanufacturing processes.

Biofoundries

This subsection concentrates on the need for biofoundries to be created within public research organisations. Biofoundries can integrate tools, technologies and overall process analysis into a platform to enable more efficient biological engineering. Through reduced cycle times and increased capacity, biofoundries might help achieve sustainability goals.

A biofoundry develops and integrates industrially relevant production microbes; advanced tools for biological engineering and data analysis; and robust, scaled-up processes for integrated biomanufacturing. In a traditional biorefinery, fermentation science and engineering may dominate at a large industrial scale. Biorefineries, conversely, are seldom discussed in terms of production strain manufacture and biological engineering. The biofoundry might also be viewed as a much smaller facility for HT iterative processes. These processes are driven by robotics and automation prior to scale-up in a larger facility such as a biopharmaceutical production plant or an industrial biorefinery. Ultimately, the streamlining of both into a single industrial workflow could be possible.

Box 6.2. Examples of public biofoundries

The Edinburgh Genome Foundry (EGF). The EGF claims to be the only fully automated DNA design, assembly and test facility in the United Kingdom. The EGF hosts CUBA, a collection of free public apps to assist with various DNA design and manufacturing tasks. It also has graphical frameworks and computational libraries for DNA design and manufacturing. It is creating EMMA-DB, a new web platform to manage genetic parts for the EMMA assembly standard, and to design new constructs from these parts (EGF, n.d.).

National University of Singapore biofoundry. The aim is to drive foundational science towards translational clinical and industrial biotechnology applications. The foundry is equipped with a robotic system that interfaces with various HT analytical instruments. This enables the biofoundry to systematically (re)design, build, test and learn to make an efficient, automated manufacturing platform. The Singapore biofoundry aims to become a central hub for synthetic biology research in Asia.

The MIT-Broad Foundry. Faced with uncertainties about the technology, this biofoundry was tasked with building organisms to produce ten molecules in three months without the biofoundry staff knowing the molecules in advance. The foundry produced the desired molecule, or a closely related one, for six out of ten targets and advanced towards production of the others (Casini et al., 2018; MIT-Broad Foundary, n.d.).

Alternatively, the classic hallmark of engineering may be envisioned with design in a biofoundry and manufacture in a separate plant at a different, even international, location. Instead of biotechnology companies owning and running their own laboratories, biofoundries in the future could do this for them (The Economist, 2018). The earliest biofoundries have already arrived (Box 6.2).

McClymont and Freemont (2017) argue that existing or new automation technologies can enable reproducible research. For this to happen, the technologies must be present in both individual research groups and centralised DNA foundries that can be accessed using cloud-based applications. They envisage that individual laboratories with in-house, low-cost automation work cells can access biofoundries via the cloud to carry out more complex experimental workflows. Technology companies exist to start enabling this process. Individual researchers and organisations can send experimental designs to foundries and return output data to the researchers.

McClymont and Freemont (2017) contend this strategy of individual, decentralised researcher/organisation and centralised biofoundry linked to the cloud via technology companies has tremendous potential. They believe it should “shift a growing proportion of molecular, cellular and synthetic biology into a fully quantitative and reproducible era”.

3D bio-printing

In 3D bio-printing, layer-by-layer precise positioning of biological materials, biochemicals and living cells is used to fabricate 3D structures (Murphy and Atala, 2014). Much of the literature concentrates on printing tissues and organs. Work has already started on 3D printing of bacteria for various bio-production purposes, although the field is still in its infancy. Previous work on producing chassis strains for production has focused on making minimal cells that act as the chassis, to which other functionalities are subsequently added (e.g. Kim et al., 2016).

Alternatively, Kyle (2018) discussed 3D printing for applications as diverse as bioremediation, environmental biosensors, oil spill filters and wound dressings. A particularly enticing prospect is to use bacteria to couple materials production with 3D printing technology. There are many challenges. Apart from the sheer volume of technical work, the future of the field will have to reconcile many issues before 3D bio-printing of bacteria can become “the next frontier in biofabrication” (Kyle, 2018). These issues include reusability, scalability, faster printing times and the environmental impact of 3D bacterial printing systems.

Cell-free synthetic biology

For now, the most relevant application of cell-free synthetic biology relates to metabolic engineering for the production of fuels, chemicals and materials. Naturally, it also applies to other bio-production processes and products. Directly related to the presence of the cell itself, various problems arise when using microbes as living chemical factories. Even in a simple bacterium, cellular metabolism is complicated and hard to control. The desired product, if it accumulates within the cell, is often toxic to the cell.

Alternatively, cell-free systems present several critical advantages. These include fast synthesis rates, direct reaction control and tolerance to toxic substrates or products. Also, cell-free systems circumvent the oft-quoted problem of scale-up because they are inherently industrially scalable (Zawada et al., 2011). The “inefficiencies” of fermentation processes (yield, titre and productivity) can be overcome in the absence of the cell. Therefore, cell-free systems provide a better possibility to produce the substance of interest at maximal yield to improve the bio-production process (Lu, 2017).

For the policy maker and risk assessment community, cell-free synthetic biology in environmental applications generates certain benefits. In bioremediation, for example, it would allow deployment of gene networks and metabolic pathways without risk of unrestrained replication and spread of new microbial strains (Karig, 2017). This would therefore circumvent the need to assess the risk from genetically modified organisms (e.g. OECD, 2015). Nevertheless, any potential risks from cell-free synthetic biology would still require science-based risk assessment.

Predictably, many difficult technical challenges remain. To broaden the applications, cell-free synthetic biology needs to be integrated with other technologies, such as 3D printing and AI. Thus, the need for greater convergence with chemistry and information technologies is evident.

Skills and education for the bioeconomy workforce

This section looks at the need for greater inter- and multidisciplinary education that would equip graduates with sufficient depth and breadth to drive the bioeconomy workforce. Delebecque and Philp (2018) looked at skills and education gaps from the production workforce to R&D. They concluded that higher education is not ready for a revolution in manufacturing that includes bio-based production. Time is limited to address the challenges: the Netherlands alone will soon require an estimate 10 000 bioeconomy experts (Langeveld, Meesters and Breure, 2016).

At the nub of the issue is the need for much greater inter- and multidisciplinary education. This training must combine biology and engineering fields with sufficient depth so as not to trivialise them. At the same time, these graduates need sufficient breadth to be truly problem-solving pioneers of engineering biology.

Backcasting: Mechatronics revisited to shape the education of the future engineering biologist

Mechatronics, already central to the modern global economy, could yield lessons for educating future engineering biologists. A translation of the French standard NF E 01-010 (Norme Française, 2008) defines mechatronics as “an approach aiming at the synergistic integration of mechanics, electronics, control theory and computer science within product design and manufacturing, in order to improve and/or optimise its functionality”.

Historically, the rise of the mechatronics engineer depended on uniting the principles of mechanics, electronics and computing to generate simpler, more economical and reliable systems. Education was refined over decades to optimise the undergraduate curriculum. This helped create the mechatronics engineers that have revolutionised manufacturing. Such an education necessitated multi- and inter-disciplinarity in critical fields such as mechanical, electrical, electronic, computer and control engineering.

The experience of mechatronics studies could inform an approach to educating a workforce for the bioeconomy. The integration of various fields has resulted in mechatronics engineers who can both solve design problems and manufacture. This is exactly the mix required by engineering biology.

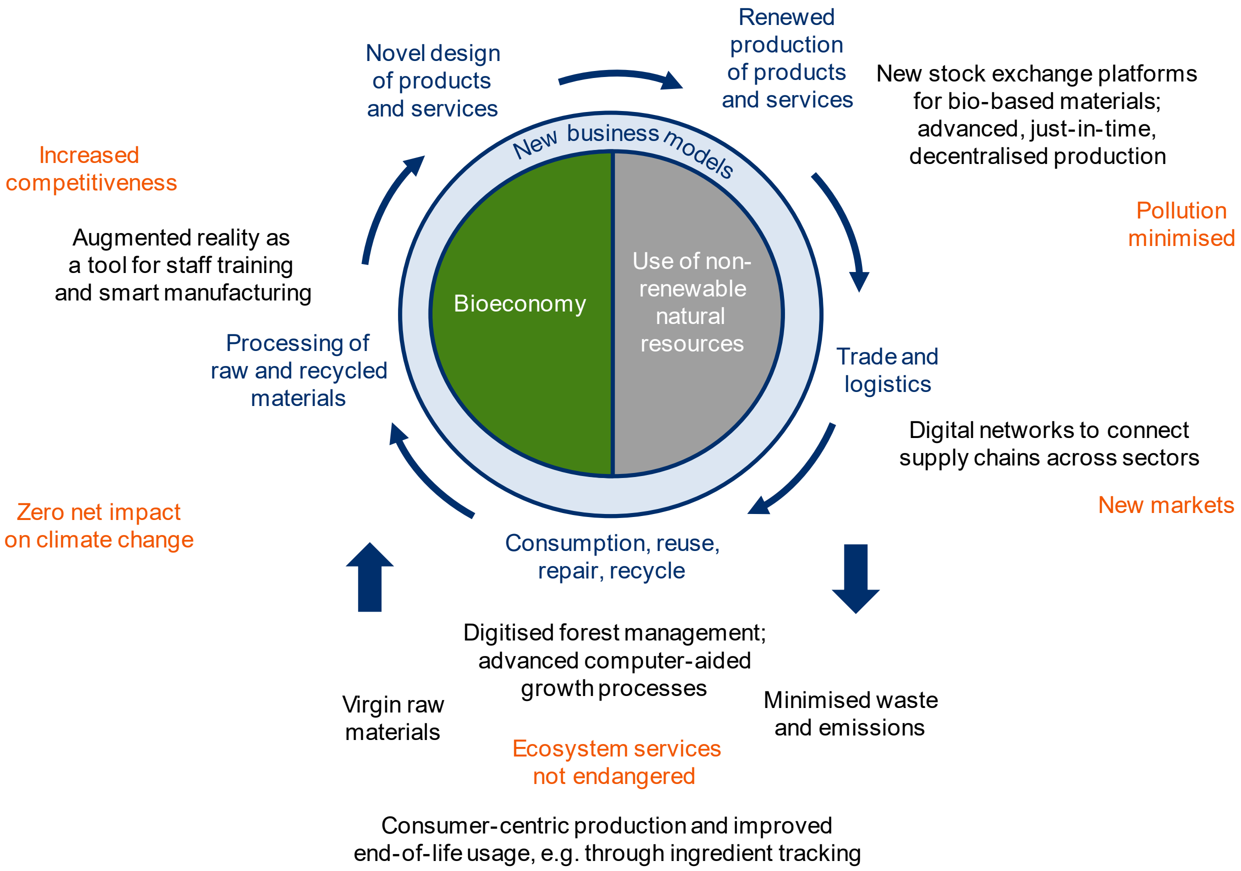
The transition from an orientation based on research to production will require a paradigm shift in biotechnology education (Delebecque and Philp, 2018). Universities will need to attract school-leavers with a more mathematical background into biotechnology. Students who graduate will need to be equally capable in DNA engineering and computation.

Digitalisation of the forestry bioeconomy

Digitalisation could offer forestry solutions that add value to the bioeconomy. Many countries with a significant forestry industry have large numbers of forest owners and few forests. Europe alone has some 16 million forest owners (Hetemäki, 2014). Compare this with the oil industry, where over 80% of the world’s proven crude oil reserves are located in the 13 OPEC countries.[3] Forestry biorefineries, by comparison to oil refineries, are expected to be small to medium facilities with local production and perhaps local consumption, a classic example of distributed manufacturing (Srai et al., 2016).

A local forest bioeconomy ecosystem and value chain could include hundreds of thousands of forest owners, entrepreneurs and companies specialising in forest service, harvesting, transport and logistics, and the production of forest products or energy. Managing this complexity requires IT tools such as apps, websites, consumer platforms and databases. Consumers are using several IT tools to both steer demand and extend their influence throughout the value chain (MISTRA, 2017). With the circular forestry bioeconomy in mind, Figure 6.2 shows a concept of how digital solutions can add value to the bioeconomy.

Figure 6.2. Digitalisation and the circular forestry bioeconomy



Source: Adapted from MISTRA (2017), "Bioeconomy and digitalisation".

Satellite technology in the forest bioeconomy

Satellite technology may be a critical tool for the forest bioeconomy, both to monitor biodiversity and to combat illegal logging. National forest monitoring systems need to deliver cost-effective and quality-controlled information across the three pillars of the bioeconomy (social, economic and environmental). Most recently, climate change has become a driving force for forest monitoring, especially concerning forest degradation and deforestation (Asner, 2009; Mitchell, Rosenqvist and Mora, 2017). The mitigation of climate change through forest management by storing carbon in the forest ecosystem is likely to become an economic and financial tool for forestry (Holmgren and Marklund, 2007). But without robust statistics, understanding the loss of biodiversity and reduction of carbon sequestration capacity from deforestation and forest degradation becomes much more difficult.

Forest monitoring is no easy task. In the past, foresters would use field and aerial surveys to collect forest cover data and aerial photography to analyse forest stocks. All of these methods were slow, laborious and expensive.

Satellite monitoring may be the only feasible future method of forest monitoring (Lynch et al., 2013). In an interesting development, a Finnish company combines machine-vision software and light detection and ranging technology (Arbonaut, n.d.). At an altitude of around 2 kilometres, laser beams can generate three-dimensional point data on an object as small as a single tree. Knowing the diameter of the crown of the tree can predict its volume (MEAE, 2017).

Making such forestry inventories supports sustainable forestry management (Crowther et al., 2015). The technology can also be used to assess carbon stocks in tropical forests. It can calculate the amount of carbon dioxide (CO_2) removed from the atmosphere, entitling a country to payments for carbon capture via forests under the Paris Agreement.

Another major issue for a burgeoning forest bioeconomy is illegal logging. This practice already costs nations tens of billions of dollars annually, and contributes some 12% of total anthropogenic CO_2 emissions globally (Lynch et al., 2013). Illegal logging is linked to warlordism, land grabbing and violent crime (Nuwer, 2016). It is also, of course, in violation of national regulations.

A satellite-based alert system could prove a potent weapon in the fight against deforestation through illegal logging. Less than eight hours after it detects that trees are being cut, a system can send e-mails to warn that an area is endangered. That rapid response could enable environmental managers to catch illegal loggers before they damage large swathes of forest (Popkin, 2016). The traditional methods of forest monitoring are far too slow to be useful against illegal logging, as speed is essential.

Examples of the potential for future bio-based materials

Many examples could illustrate the potential for bio-based materials. The three selected all have high economic and societal value, but differ in terms of engineering biology and IT or chemistry convergence (Box 6.3).

Box 6.3. Selected bio-based materials of high economic and social value

Angiotensin converting enzyme inhibitors

Captopril was the first marketed angiotensin converting enzyme (ACE) inhibitor. Its effects on blood pressure mechanisms mimicked those of a peptide discovered in the Brazilian pit viper *Bothrops jararaca* (Mladic et al., 2017). The viper uses an ACE molecule to make its prey faint from a rapid drop in blood pressure. The discovery heralded major changes in the approach to treatment of hypertension and heart failure.

ACE inhibitors have been credited with saving millions of lives. The ACE market, valued at USD 11.7 billion in 2015, was expected to reach USD 12.45 billion by 2024. The search continues for new ACE inhibitors due to the prevalence of hypertension as the human population ages.

It is unlikely that much benefit reverts back to the genetic origins of the initial discovery. A goal of the Nagoya Protocol is to distribute wealth created from genetic discoveries more evenly. The Access and Benefit-sharing Clearing-House (ABS Clearing-House) is a key tool for monitoring the use of genetic resources along the value chain, including through the internationally recognised certificate of compliance. Blockchain technology lends itself to this task. It can record transactions between two parties efficiently in a verifiable and permanent way, thus providing secure traceability. Blockchain could also provide enhanced security of data in clinical trials.

Graphene and green chemistry

Graphene is a key material of the future. It conducts electricity better than copper and will eventually find its way into consumer electronics. Electricity conductance and physical flexibility mean that graphene has many potential applications. These range from energy storage devices to lighting and displays, solar panels, tyres, bicycle frames and fashion items (Mertens, 2018). For example, deformable batteries with flexible, foldable and/or stretchable capabilities are ideal for wearable and portable electronics (Ye et al., 2018). Graphene may be the material of choice for 3D printable batteries. Estimating the market value of graphene is complicated as the range of uses cannot yet be fully explored. It is mainly limited to research applications due to high costs. The 2015 price was some USD 500 per gramme.

Researchers in Australia have created a new method of graphene synthesis. It involves heating soybean oil in air until it breaks down into carbon building units that are essential for the synthesis of graphene (Seo et al., 2017). Moreover, the researchers demonstrated the versatility of the method by using other renewable carbon-containing materials such as butter.

While soybean oil has other valuable uses, lignin is generated in large quantities. However, it is difficult to valorise in any value-added process due to its complexity. Lignin is produced in large quantities in the pulp and paper industry, and often burned for power generation. However, Liu, Chen and Gao (2017) described a method for converting lignin into graphene.

Spider silk

Spider silks are the toughest known biological materials. They are lightweight and virtually invisible to the human immune system, and thus have "revolutionary potential for medicine and industry" (Babb et al., 2017). Among newer applications of spider silk being considered are microphones in hearing aids and cell phones. Stronger than steel, tougher than Kevlar, the range of applications is large. For example, the US army has been testing protective garments for soldiers made from spider silk. An *E. coli* variant of spider silk could replace Kevlar in air bags as it is both strong and flexible. And in 2017, Adidas unveiled a spider silk shoe using Biosteel fibres from AMSilk. Called the Adidas Futurecraft Biofabric, the shoes are reported to be biodegradable in less than 36 hours in the presence of an enzyme.

Biologists are attracted to the study of spider silk because of the large diversity of silks and proteins involved in their synthesis. Even after decades of research on orb-weaver spider silks, knowledge of all the proteins within an orb-weaver species is incomplete – and there are tens of thousands of spider species. Moreover, nature can also inform a production process: there are genes that encode proteins that turn liquid silk into solid silk thread. Genomics is the newest tool to unravel this complexity.

Engineering biologists are interested in spider silk as there are many candidates of genes and proteins for transgenic studies. This implies the possibility of tailor-made spider silks for different materials and applications. However, working with spiders as factories is impracticable. The expression of spider silk genes in a microorganism with subsequent fermentation processes is much more attractive. Much remains to be discovered. The sheer diversity of spiders and their silks lends itself to use of digital tools for curating knowledge, as well as for the "pick and mix" analysis for new consumer applications. Expression in microorganisms is extremely complex. Digitally assisted design, screening and automation will be needed to drastically reduce the design, build and test time.

Policy implications

Engineering biology materials have implications for policy makers with respect to platform technologies to support delivery of the materials; standardisation, interoperability and IT; sustainability; and digital sector. This section unpacks each of these implications.

Platform technologies to support the delivery of engineering biology materials

Focus R&D subsidy on achieving reproducibility of bio-production processes: Precompetitive design of R&D programmes (for laboratory-scale considerations) and near-market collaborative programmes could ensure that research proposals are only successful if they concentrate on improving reproducibility. Less fashionable near-market research issues also need to be investigated. These include robustness-in-design (e.g. DARPA's BRICS); titre, yield and productivity; bioprocess variables, such as the effects of media variability (e.g. different sources of molasses); internal gradients, such as oxygen and redox; and tolerance to shear stress that can cause cell breakage. Combining digital and biological tools is the best available way to reduce discovery time given the complexity of biology.

Platform technologies of various sorts: Governments need to support the platform technologies required (e.g. biofoundries, distributed R&D networks, digital platforms, data curation and digital/genetic data storage). This is the case because investment risks are too high for the private sector, and the imperatives for private action may be missing (e.g. a clear route to market). This goes beyond R&D subsidy. Innovative forms of public-private partnership are needed. These would enable both public and private actors to gain fair access to equipment, services and data (see suggestions below on IP and licences).

Academic/industry/government co-development that can implement and innovate these technologies: Implementation that involves both public and private actors could involve national action plans and roadmaps. In the United Kingdom, for example, a "leadership council" is constituted to ensure that deadlines and milestones for implementation are met. This council can easily report at ministerial level to maintain an appropriate political focus and vision.

A highly trained interdisciplinary workforce: For too long, the life sciences have been compartmentalised by discipline, such as microbiology, biochemistry and molecular biology. A greater focus on problem solving, using interdisciplinarity and including soft skills, is more appropriate to graduating biologists seeking careers in manufacturing (Delebecque and Philp, 2018). In a related issue, policy makers should prioritise identifying a common language and vision, both computing and spoken. It should place engineering biology at the nexus of other disciplines, especially materials science, automation engineering, chemistry, computer science and engineering. Both chemistry and biology benefit from greater levels of digitalisation, and the extremely important synergy of engineering biology with green chemistry should be a specific focus.

Standardisation, interoperability and intellectual property

Standardisation, interoperability and IP: Standardisation and interoperability policies can be seen throughout the history of the microprocessor industry and, more recently, in information and communication technology (ICT). The issues for engineering biology are similar, but the modern context highlights some differences. In particular, policy makers need to consider carefully the ongoing debate about open access versus IP protection to satisfy the desires of academia and the need for sufficient protection to motivate private investment.

Legal issues are inextricably linked with standards that enable product and process interoperability. Rules may be required that licences be either royalty-free or royalty-bearing on terms that are "fair, reasonable and non-discriminatory", a system used extensively in the ICT sector (Contreras, Rai and Torrance, 2015). If patents on standards are obtained, what rules will govern the terms on which they will be made available to the community? Best outcomes for engineering biology will likely result from simultaneous consideration of technical standards and IP issues, with lessons to be learned from the ICT sector.

The use of materials transfer agreements (MTAs) provide an example of potential difficulties. MTAs underlie the legal frameworks within which biotechnology practitioners define the terms and conditions for sharing biomaterials. However, MTA legal arrangements pre-date the widespread adoption of the Internet, engineering biology, genome sequencing and gene synthesis. As such, they can place restrictions on the redistribution and commercial use of biomaterials. Moreover, they are not aligned with changes in the social objectives of science.

In response, Kahl et al. (2018) suggested a new model, the Open Materials Transfer Agreement (OMTA). This would relax restrictions and support widespread adoption within automated and semi-automated administration systems. Benefits of electronic platforms are various. Incorporation of the OMTA within electronic platforms could enable less restrictive options for sharing biomaterials as appropriate. Technology transfer offices could still review and approve such transfers, but electronic communications could replace paperwork and individual negotiations. Such electronic platforms could also offer provenance tracking, which may be a sustainability consideration.

Sustainability

Sustainable materials manufacturing and management: There are roles for digital technologies in judging sustainability. Sustainability standards should be an intense focus in the bioeconomy generally, and specifically in engineering biology and biomanufacturing. Issues such as the provenance of feedstocks could be explored using blockchain technology. Automated, digitalised protocols for sustainability assessment would decrease the financial burden on small companies tasked with proving the sustainability of their products and processes. For example, it could compare greenhouse gas emissions savings associated with products and primary fossil energy savings of the manufacturing processes with costs of fossil counterparts.

Digital security

Digital security: Individual facilities, whether publicly or privately held, could develop and validate methods and protocols for facility staff or external service providers to fortify the facility (Murch et al., 2018). This has special applicability to public-private partnerships as public research organisations are notoriously co-operative and “leaky”.

Governments could encourage the sharing of timely cyberthreat information by providing protections related to lawsuits, public disclosure and antitrust concerns, as well as safeguarding privacy and civil liberties. Cybercrimes should be prosecuted vigorously. Perpetrators should be held responsible for harm to operating systems, for stealing IP and trade secrets, or for unlawfully obtaining personal information for financial gain.

Governments could encourage cybersecurity awareness-building and co-operation. One example could be to encourage public sector actors to run cyber-attack simulations and to share the lessons learned. Efforts to enhance cybersecurity should be recognised through, for example, voluntary standards, regulations, industry programmes and information-sharing frameworks.

Conclusion

This chapter attempts to draw the needs associated with digitalisation to the attention of policy makers. This starts with education. In the near term, engineering biology needs successes. For the public policy maker, there is nothing better than success stories to show that taxpayers’ money is being wisely spent. But policy must also give the private sector confidence that governments realise that the era of bio-based production has arrived.

Due to foundations laid down in previous decades, biology, quite suddenly in this decade, finds itself in a data-rich era. This trend will undoubtedly continue and has implications for biotechnology and the emerging engineering biology. Literally hundreds of engineering (synthetic) biology start-ups are receiving investments. However, engineering biology needs a large increase in its quantitative precision to qualify as a manufacturing discipline. Some solutions can come from biology itself, but a greater alignment with automation, as in so much of modern manufacturing, is also needed. When married to the complexity of biology, there is an obvious need for a step-change in digitalisation.

References

Appleton, E. et al. (2017), “Design automation in synthetic biology”, *Cold Spring Harbor Perspectives in Biology* 9, Cold Spring Laboratory Press, United States, https://doi.org/10.1101/cshperspect.a023978.

Arbonaut (n.d.), Arbonaut website, www.arbonaut.com/en (accessed 17 June 2019).

Asner, G.P. (2009), “Tropical forest carbon assessment: Integrating satellite and airborne mapping approaches”, *Environmental Research Letters*, Vol. 4/3, IOP Publishing, Bristol, https://doi.org/10.1088/1748-9326/4/3/034009.

Babb, P.L. et al. (2017), “The *Nephila clavipes* genome highlights the diversity of spider silk genes and their complex expression”, *Nature Genetics*, Vol. 49, Nature Research, Springer, pp. 895-903.

Baker, M. (2016), “Is there a reproducibility crisis?”, *Nature*¸ Nature Research, Springer, Vol. 533, pp. 452-454.

Beal, J. et al. (2016), “Reproducibility of fluorescent expression from engineered biological constructs in E. coli”, *PLOS One*, Vol. 11/3, Proceedings of the National Academy of Sciences of the United States of America, e0150182, https://doi.org/10.1371/journal.pone.0150182.

Burgard, A. et al. (2016), “Development of a commercial scale process for production of 1,4-butanediol from sugar”, *Current Opinion in Biotechnology*, Vol. 42, Elsevier, Amsterdam, pp. 118-125.

Carlson, R. (2018), “Building the bioeconomy”, presentation at the second global bioeconomy summit, Berlin, 19-20 April, https://gbs2018.com/fileadmin/gbs2018/Downloads/GBS_2018_Conference-Brochure_web.pdf.

Casini, A. et al. (2018), “A pressure test to make 10 molecules in 90 days: External evaluation of methods to engineer biology”, *Journal of the American Chemical Society*, Vol. 140/12, American Chemical Society, Washington, DC, pp. 4302-4316.

Contreras, J.L., A.K. Rai and A.W. Torrance (2015), “Intellectual property issues and synthetic biology standards”, *Nature Biotechnology*, Vol. 33, Nature Research, Springer, pp. 24-25.

Crowther, T.W. et al. (2015), “Mapping tree density at a global scale”, *Nature*, Vol. 525, Nature Research, Springer, pp. 201-205.

Delebecque, C.J. and J. Philp (2018), “Education and training for industrial biotechnology and engineering biology”, *Engineering Biology*, Vol. 3/1, Institution of Engineering and Technology, Stevenage, United Kingdom, pp. 6-11, https://doi.org/10.1049/iet-enb.2018.0001.

Densmore, D. (2012), “Bio-design automation: Nobody said it would be easy”, *ACS Synthetic Biology*, Vol. 1/8, American Chemical Society, Washington, DC, pp. 296-296.

EBP (n.d.), Earth Biogenome Project website, www.earthbiogenome.org/ (accessed 17 June 2019).

EGF (n.d.), Edinburgh Genome Foundry website, www.genomefoundry.org/ (accessed 17 June 2019).

Extance, A. (2016), “How DNA could store all the world’s data”, *Nature*, Vol. 537, Nature Research, Springer, pp. 22-24.

Fong, S.S. (2014), “Computational approaches to metabolic engineering utilizing systems biology and synthetic biology”, *Computational and Structural Biotechnology Journal*, Vol. 11/18, Elsevier, Amsterdam, pp. 28-34.

Gerbaud, V. et al. (2017), “Computer-aided framework for designing bio-based commodity molecules with enhanced properties”, *Chemical Engineering Science*, Vol. 159, Elsevier, Amsterdam, pp. 177-193.

Hadadi, N. et al. (2016), “ATLAS of biochemistry: A repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies”, *ACS Synthetic Biology*, Vol. 5/10, American Chemical Society, Washington, DC, pp. 1155-1166, https://doi.org/10.1021/acssynbio.6b00054.

Harder, B-J., K. Bettenbrock and S. Klamt (2016), "Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*", *Metabolic Engineering*, Vol. 38, Elsevier, Amsterdam, pp. 29-37.

Hayden, E.C. (2015), "Synthetic biology called to order. Meeting launches effort to develop standards for fast-moving field", *Nature*, Vol. 520, Nature Research, Springer, pp. 141-142.

Hetemäki, L. (ed.) (2014), *Future of the European Forest-based Sector*, European Forest Institute, Barcelona.

Holmgren, P. and L.-G. Marklund (2007), "National forest monitoring systems – purposes, options and status" in P.H. Freer-Smith, M.S.J. Broadmeadow and J.M. Lynch (eds.), *Forestry & Climate Change*, CAB International, Wallingford, United Kingdom.

Hong, K.K. and J. Nielsen (2012), "Metabolic engineering of *Saccharomyces cerevisiae*: A key cell factory platform for future biorefineries", *Cellular and Molecular Life Sciences* Vol. 69/16, Springer-Science and Business Media, pp. 2671-2690.

Iansiti, M. and K.R. Lakhani (2017), "The truth about blockchain", *Harvard Business Review*, January-February, pp. 118-127.

Il Bioeconomista (10 June 2015), "Synbio start-ups need an average of 7.4 years to launch first chemical product", Bioeconomy News, Politics and Business blog, http://ilbioeconomista.com/2015/06/10/synbio-start-ups-need-an-average-of-7-4-years-to-launch-first-chemical-product/.

Kahl, L. et al. (2018), "Opening options for material transfer", *Nature Biotechnology*, Vol. 36, Nature Research, Springer, pp. 923-927.

Karig, D.K. (2017), "Cell-free synthetic biology for environmental sensing and remediation", *Current Opinion in Biotechnology*, Vol. 45, Elsevier, Amsterdam, pp. 69-75.

Kim, J. et al. (2016), "Properties of alternative microbial hosts used in synthetic biology: Towards the design of a modular chassis", *Essays in Biochemistry*, Vol. 60/4, Biochemical Society, London, pp. 303-313.

Kitney, R. et al. (2019), "Enabling the advanced bioeconomy through public policy supporting biofoundries and engineering biology", *Trends in Biotechnology*, September, Vol. 37, Issue 9, pp. 917-920, https://www.sciencedirect.com/journal/trends-in-biotechnology/vol/37/issue/9.

KPMG (2018), "Digitalization in life sciences", *KPMG International*, No. 135128-G, KPMG, Zug, Switzerland.

Kwok, R. (2010), "Five hard truths for synthetic biology", *Nature*, Vol. 463, Nature Research, Springer, pp. 288-290.

Kyle, S. (2018), "3D printing of bacteria: The next frontier in biofabrication", *Trends in Biotechnology*, Vol. 36/4, Elsevier, Amsterdam, pp. 340-341.

Langeveld, J.W.A., K.P.H. Meesters and M.S. Breure (2016), "The bio-based economy and the bioeconomy in the Netherlands", *Biomass Research Report*, Vol. 1601/59015257, Biomass Research, Wageningen.

Lee, S.Y. and H.U. Kim (2015), "Systems strategies for developing industrial microbial strains", *Nature Biotechnology*, Vol. 33, Nature Research, Springer, pp. 1061-1072.

Le Feuvre, R.A. and N.S. Scrutton (2018), "A living foundry for synthetic biological materials: A synthetic biology roadmap to new advanced materials", *Synthetic and Systems Biotechnology*, Vol. 3/2, ScienceDirect, pp. 105-112.

Lewin, H.A. et al. (2018), "Earth BioGenome Project: Sequencing life for the future of life", *PNAS,* Vol. 115/7, National Academy of Sciences, Washington, DC, pp. 4325-4333.

Liu, F., Y. Chen and J. Gao (2017), "Preparation and characterization of biobased graphene from kraft lignin", *BioResources*, Vol. 12/3, North Carolina State University, Raleigh, pp. 6545-6557.

Lu, Y. (2017), "Cell-free synthetic biology: Engineering in an open world", *Synthetic and Systems Biotechnology*, Vol. 2/1, ScienceDirect, pp. 23-27.

Lynch, J. et al. (2013), "Choose satellites to monitor deforestation", *Nature*, Vol. 496, Nature Research, Springer, pp. 293-294.

Maiti, S. et al. (2016), "Agro-industrial wastes as feedstock for sustainable bio-production of butanol by *Clostridium beijerinckii*", *Food and Bioproducts Processing*, Vol. 98, Elsevier, Amsterdam, pp. 217-226.

Makarova, A.S. et al. (2017), "Green chemistry for the optimum technology of biological conversion of vegetable waste", *Sustainable Production and Consumption*, Vol. 10, Elsevier, Amsterdam, pp. 66-73.

McClymont, D.W. and P.S. Freemont (2017), "With all due respect to Maholo, lab automation isn't anthropomorphic", *Nature Biotechnology,* Vol. 35, Nature Research, Springer, pp. 312-314.

Mertens, R. (2018), *The Graphene Handbook*, Lulu.com.

MEAE (2017), "Wood-based bioeconomy solving global challenges", Ministry of Economic Affairs and Employment of Finland, Helsinki.

MISTRA (2017), "Bioeconomy and digitalisation", The Swedish Foundation for Strategic Environmental Research, Stockholm.

MIT-Broad Foundry (n.d.), MIT-Broad Foundry website, http://web.mit.edu/foundry/ (accessed 17 June 2019).

Mitchell, A.L., A. Rosenqvist and B. Mora (2017), "Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+", *Carbon Balance and Management*, Vol. 12/9, Springer Nature, https://doi.org/10.1186/s13021-017-0078-9.

Mladic, M. et al. (2017), "Rapid screening and identification of ACE inhibitors in snake venoms using at-line nanofractionation LC-MS", *Analytical and Bioanalytical Chemistry*, Vol. 409/25, Springer Journal, pp. 5987-5997.

Murch, R.S. et al. (2018), "Cyberbiosecurity: An emerging new discipline to help safeguard the bioeconomy", *Frontiers in Bioengineering and Biotechnology*, Vol. 6/39, Frontiers, Lausanne, https://doi.org/10.3389/fbioe.2018.00039.

Murphy, S.V. and A. Atala (2014), "3D bioprinting of tissues and organs", *Nature Biotechnology*, Vol. 32, Nature Research, Springer, pp. 773-785.

Norme Française (2008), "Mécatronique.Vocabulaire", NF E 01-010.

Nuwer, R. (2016), "Illegal logging has become more violent than ever", *National Geographic*, 3 February.

OECD (2015), *Biosafety and the Environmental Uses of Micro-Organisms: Conference Proceedings*, Harmonisation of Regulatory Oversight in Biotechnology, OECD Publishing, Paris, https://doi.org/10.1787/9789264213562-en.

OECD (2014), *Emerging Policy Issues in Synthetic Biology*, OECD Publishing, Paris, https://doi.org/10.1787/9789264208421-en.

OECD (2009), *The Bioeconomy to 2030: Designing a Policy Agenda*, OECD Publishing, Paris, https://doi.org/10.1787/9789264056886-en.

Ogunnaike, T. (2016), "Can DNA hard drives solve our looming data storage crisis", Singularity Hub, 21 October, https://singularityhub.com/2016/10/21/can-dna-hard-drives-solve-our-looming-data-storage-crisis/#sm.001skz7or17zxe3wv9f13vv6jcy8k.

Philp, J.C., R.J. Ritchie and J.E.M. Allan (2013), "Biobased chemicals: The convergence of green chemistry with industrial biotechnology", *Trends in Biotechnology*, Vol. 31/4, Elsevier, Amsterdam, pp. 219-222.

Philp, J. and D.E. Winickoff (2018), "Realising the circular bioeconomy", OECD Science, Technology and Industry Policy Papers, No. 60, OECD Publishing, Paris, https://doi.org/10.1787/31bb2345-en.

Popkin, G. (2016), "Satellite alerts track deforestation in real time", *Nature,* Vol. 530, Nature Research, Springer, pp. 392-393.

Quaglia, D. (20 February 2017), "Antha: A platform for engineering biology", PLOS Synbio Community blog, https://blogs.plos.org/synbio/2017/02/20/antha-a-platform-for-engineering-biology/.

Rodrigo, G. and A. Jaramillo (2013), “AutoBioCAD: Full biodesign automation of genetic circuits”, *ACS Synthetic Biology*, Vol. 2/5, American Chemical Society, Washington, DC, pp. 230-236.

Rogers, J.K. and G.M. Church (2016), “Multiplexed engineering in biology”, *Trends in Biotechnology*, Vol. 34/3, Elsevier, Amsterdam, pp. 198-206.

Rogers, J.K. et al. (2015), “Synthetic biosensors for precise gene control and real-time monitoring of metabolites”, *Nucleic Acids Research*, Vol. 43/15, Oxford University Press, pp. 7648-7660.

Sadowski, M.I., C. Grant and T.S. Fell (2016), “Harnessing QbD, programming languages, and automation for reproducible biology”, *Trends in Biotechnology*, Vol. 34/3, Elsevier, Amsterdam, pp. 214-227.

Seo, D.H. et al. (2017). “Single-step ambient-air synthesis of graphene from renewable precursors as electrochemical genosensor”, *Nature Communications*, Vol. 8/14217, Nature Research, Springer, https://doi.org/10.1038/ncomms14217.

Service, R.F. (2017), “DNA could store all of the world's data in one room”, *Science*, 2 March, https://doi.org/10.1126/science.aal0852.

Shipman, S.L. et al. (2017), “CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria”, *Nature*, Vol. 547, Nature Research, Springer, pp. 345-349.

Srai, J.S. et al. (2016), “Distributed manufacturing: Scope, challenges and opportunities”, *International Journal of Production Research*, Vol. 54/23, Taylor & Francis Online, pp. 6917-6935.

The Economist (2018), “Robotic labs for high-speed genetic research are on the rise. The design of synthetic lifeforms could become a new industry”, *The Economist*, 1 March.

Tung, L. (2018), “Microsoft's DNA storage breakthrough could pave way for exabyte drives”, ZD Net, 21 February, www.zdnet.com/article/microsofts-dna-storage-breakthrough-could-pave-way-for-exabyte-drives/.

Van Dien, S. (2013), “From the first drop to the first truckload: Commercialization of microbial processes for renewable chemicals”, *Current Opinion in Biotechnology*, Vol. 24/6, Elsevier, Amsterdam, pp. 1061-1068.

Wang, B.L. (2014), “Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption”, *Nature Biotechnology*, Vol. 32/5, Nature Research, Springer, pp. 473-478.

World Economic Forum (2018), “Harnessing the fourth industrial revolution for life on land”, World Economic Forum, Geneva.

Wurtzel, E.T. and T.M. Kutchan (2016), “Plant metabolism, the diverse chemistry set of the future”, *Science*, Vol. 353/6305, American Association for the Advancement of Science, Washington, DC, pp. 1232-1236.

Xiao, Y. et al. (2016), “Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis”, *Nature Chemical Biology*, Vol. 12, pp. 339-344.

Ye, M. et al. (2018), “Graphene platforms for smart energy generation and storage”, *Joule*, Vol. 2/2, Elsevier, Amsterdam, pp. 245-268.

Zawada, J.F. et al. (2011), “Microscale to manufacturing scale-up of cell-free cytokine production: A new approach for shortening protein production development timelines”, *Biotechnology and Bioengineering*, Vol. 108/7, Wiley Online Library, pp. 1570-1578.

Zhirnov, V. et al. (2016), “Nucleic acid memory”, *Nature Materials*, Vol. 15, Nature Research, Springer, pp. 366-370.

Notes

[1] Estimates of trends in DNA sequencing and synthesis costs are available at the Bioeconomy Dashboard: www.bioeconomycapital.com/bioeconomy-dashboard/.

[2] The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity is an international agreement that aims at sharing the benefits arising from the utilisation of genetic resources in a fair and equitable way.

[3] See data provided by the Organization for Petroleum Exporting Countries, www.opec.org/opec_web/en/data_graphs/330.htm.

7 The digitalisation of science and innovation policy

Michael Keenan, Dmitry Plekhanov, Fernando Galindo-Rueda, and Daniel Ker

Directorate for Science, Technology and Innovation, OECD

This chapter is based on the OECD Committee for Scientific and Technological Policy project and its exploration of digital science and innovation policy (DSIP) and the challenges it faces. DSIP initiatives refer to the adoption or implementation by public administrations of new or reused procedures and infrastructures relying on an intensive use of digital technologies and data resources to support the formulation and delivery of science and innovation policy. The chapter focuses on three issues in particular. First, it examines the need to ensure interoperability through which diverse data sets can be linked and analysed to aid policy making. Second, it looks at preventing potential misuses of DSIP systems in research assessment practices. Third, it explores management of the roles of non-government actors, particularly the private sector, in developing and operating DSIP infrastructure components and services.

Introduction

As scientific research and innovation increasingly leave a digital "footprint", datasets are becoming ever larger, more complex and available at higher speed. At the same time, technological advances – in machine learning (ML) and natural language processing, for example – are opening new analytical possibilities. Science, technology and innovation (STI) policy can benefit from these dynamics (Box 7.1). They can harness the power of digitalisation to link and analyse datasets covering diverse areas of policy activity and impact. For example, initiatives already experiment with semantic technologies to link datasets, with artificial intelligence to support big data analytics and with interactive visualisation and dashboards to promote data use in the policy process.

Box 7.1. A short overview of the OECD DSIP project

Over 2017 and 2018, the OECD mapped the landscape of digital science and innovation policy (DSIP) initiatives in OECD countries and partner economies. The OECD DSIP project aimed to help policy makers and researchers assess the transformational potential and possible pitfalls of using digital tools and sources in science and innovation policy making. The project also sought to facilitate learning between countries that are planning, developing or using DSIP systems. The project was carried out under the supervision of the OECD Committee for Scientific and Technological Policy (CSTP) and its Working Party of National Experts on Science and Technology Indicators.

The project included a survey of DSIP initiatives that provides much of the evidence used in this chapter. The survey had three elements:

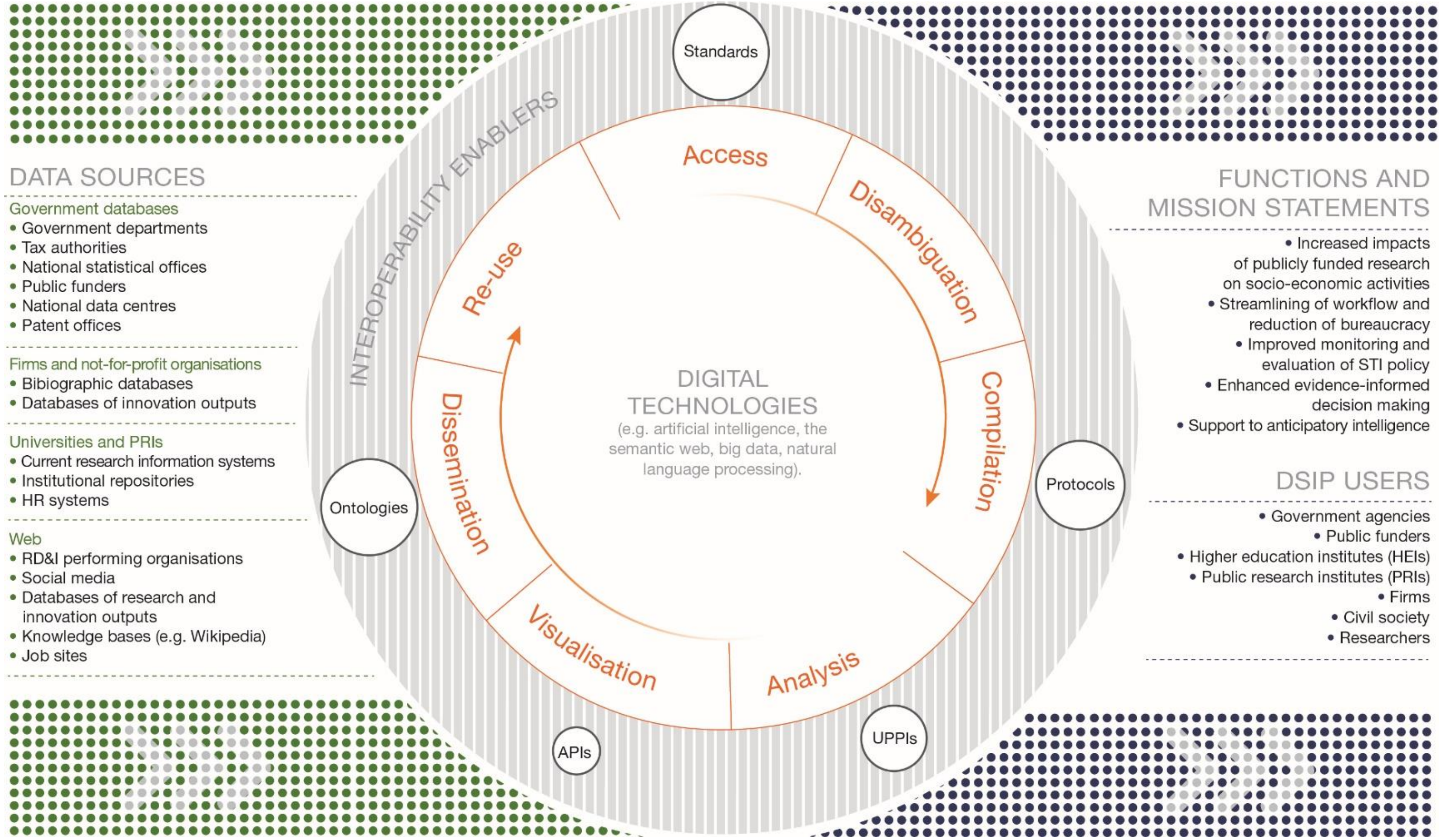
- CSTP delegates identified, and characterised to a basic level, 61 DSIP initiatives in their countries.
- Of these 61 initiatives, 39 DSIP initiative managers completed a questionnaire providing further details on the characteristics of their systems, including the data they use, the ways they link data and the main challenges they face.
- The OECD Secretariat conducted 20 follow-up interviews with DSIP initiative managers to understand better the origins and dynamics of their systems.

The OECD Secretariat carried out further interviews with leaders of not-for-profit organisations, e.g. Open Researcher and Contributor ID (ORCID) and the Common European Research Information Format. It also met with senior managers from corporate DSIP solutions providers, including Microsoft and Elsevier. The project also included a case study of Norway's DSIP landscape, as described in Box 7.4.

What is digital science and innovation policy?

Figure 7.1 provides a stylised conceptual view of a DSIP initiative and its main components. All of these elements interact in nationally specific ways, reflecting each country's history and institutional set-up. The main elements consist of various input data sources. These feed into a data cycle enabled by interoperability standards, including unique, persistent and pervasive identifiers (UPPIs). DSIP systems can perform a number of functions and are often used by a mix of users. Box 7.2 outlines several examples of DSIP initiatives from across the world.

Figure 7.1. A stylised conceptual view of a DSIP initiative and its possible main components



Note: DSIP = digital science and innovation policy; STI = science, technology and innovation; HEI = higher education institution; PRI = public research institution; API = application programming interface; UPPI = unique, persistent and pervasive identifier; HR = human resources; RD&I = research, development and innovation.
Source: OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, https://doi.org/10.1787/sti_in_outlook-2018-en.

Data are predominantly sourced from a mix of administrative data sources held by funding agencies (e.g. databases of grant awards) and organisations that perform research, development and innovation (RD&I). These include Current Research Information Systems (CRIS) in universities, and proprietary bibliometric and patent databases. Some DSIP systems have grown out of these databases. Through integration with external platforms or development of add-on services, they have evolved into infrastructures that can deliver comprehensive data analysis on research and innovation activities. Other systems have been established from the ground up. Several DSIP systems harvest data from the web to build a picture of the incidence and impacts of science and innovation activities. Web sources include, but are not limited to, company websites and social media.

DSIP infrastructures can increase the scope, granularity, verifiability, communicability, flexibility and timeliness of policy analyses. They can lead to the development of new STI indicators (Bauer and Suerdem, 2016), the assessment of innovation gaps (Kong et al., 2017), strengthened technology foresight (Kayser and Blind, 2017) and the identification of leading experts and organisations (Shapira and Youtie, 2006; Johnson, Fernholz and Fosci, 2016; Gibson et al., 2018). Furthermore, in some countries, researchers and policy makers have started to experiment with natural language processing and ML. They are using it to track emerging research topics and technologies (Wolfram, 2016; Mateos-Garcia, 6 April 2017) and to support RD&I decisions and investments (Yoon and Kim, 2012; Park, Yoon and Kim, 2013; Yoon, Park and Kim, 2013). Box 7.3 outlines the range of goals set for DSIP initiatives.

Box 7.2. Examples of DSIP systems covered by the OECD study

Databases of public funders

In **Belgium**, the Flemish department of Economy, Science and Industry, in co-operation with data providers and information technology partners, developed the Flanders Research Information Space (FRIS) in 2011. It aims to accelerate innovation, support science and innovation policy making, share information on publicly funded research with citizens and reduce the administrative burden of research reporting. The FRIS is a single window on all Flemish research. It can be used by government agencies in several ways. First, it is a tool to improve the visibility of research funding programmes. Second, it is a resource for in-depth analyses of scientific and technological trends and the development of statistical indicators on STI.

In **Brazil**, the National Council for Scientific and Technological Development established Lattes Platform with support from the Ministry of Science and Technology, the Ministry of Education and the government body "Co-ordination for the Improvement of High-Level Personnel". The platform supports policy design and formulation, management of research funding programmes and strategic planning. It is based on integrations of a variety of digital resources of Brazilian government agencies and higher education institutions (HEIs). Aside from visualising Brazilian STI datasets, Lattes Platform enables the design of add-on analytical solutions to better serve the needs and expectations of science and innovation policy makers.

In **Poland**, the Ministry of Science and Higher Education launched the POL-on system using financial support from the European Union and the technical assistance of three private companies. POL-on is an integrated information system for higher education. It supports the work of the Ministry of Science and Higher Education, as well as other ministries and institutions of science and higher education. Its main task is to create a database of scientific institutions, universities and Polish science. Information collected through the system supports the decision-making process of the Ministry of Science and Higher Education regarding Polish universities and research units. Certain parts of datasets collected by the system are made available to the public.

In **Argentina**, the Ministry of Science, Technology and Productive Innovation uses SICYTAR (Sistema de Información de Ciencia y Tecnología Argentino) to evaluate and assess STI policy initiatives, project teams and individual researchers. The system aggregates several databases, covering researchers' curriculum vitae; funded research and development (R&D) projects; information on public and private institutions performing R&D activities in Argentina; and, information on large research equipment.

Current Research Information Systems

In **Estonia**, a number of stakeholders launched the Estonian Research Information System (ETIS). These include the Ministry of Education and Research, the Estonian Science Foundation, the Scientific Competence Council, public organisations that perform RD&I, and the Archimedes Foundation. Based on multi-partner co-operation, ETIS serves as a large-scale national digital system that unites data management efforts. HEIs use ETIS as an internal system for research information management and as a tool to showcase their research. Public funders use the system to evaluate and process grant applications. ETIS is also used in national research assessments and evaluations by providing data on STI indicators, e.g. R&D revenue per research and teaching staff member and the percentage of women among scientists.

In the **Netherlands**, the National Academic Research and Collaborations Information System (NARCIS) collects data from multiple sources. These include funder databases, CRISs, institutional repositories of research performers and the Internet. Data on research outputs, projects, funding, human resources and policy documents collected by NARCIS inform policy makers on research in the Netherlands and monitor the openness of access to data. Funders also use the system to identify research gaps to improve resource planning. NARCIS also serves as an important research directory, providing researchers, journalists, and the domestic and international public with information on the status and outputs of Dutch science.

In **Norway**, the research-reporting tool Cristin collects information from research institutions, the Norwegian Centre for Research Data and ethics committees. Cristin serves as a resource for the performance-based funding model of the Ministry of Research and Education. It provides numerous users from government, industry, academia and civil society with verified information on the current status of Norwegian research.

Intelligent systems

In **Japan**, the National Graduate Institute for Policy Studies designed the SciREX Policymaking Intelligent Assistance System (SPIAS) to strengthen national evidence-informed STI policy making. SPIAS uses big data and semantic technologies to process data on research outputs and impacts, funding, R&D-performing organisations and research projects, with a view to mapping the socio-economic impacts of research. SPIAS has been used to analyse leading Japanese scientists' performance before and after receiving grants from the Japan Science and Technology Agency. It has also been used to assess the impact of regenerative medicine research in Japan, and to map emerging technologies.

In **Spain**, Corpus Viewer, developed by the State Secretariat for Information Society and Digital Agenda, processes and analyses large volumes of textual information using natural language processing techniques. Policy makers use results to monitor and evaluate public programmes, and to formulate science and innovation policy initiatives. The system is restricted to government officials.

Box 7.3. Range of typical goals of DSIP initiatives

Optimisation of administrative workflows. Digital tools can help streamline potentially burdensome administrative procedures and deliver significant efficiency gains within agencies. These benefits can also extend to those using public agencies' services, including researchers or organisations applying for (or reporting on) the use of research grants. For example, they can use interoperability identifiers to link their research profiles to grant applications. As the digital gateway to the Estonian research system, ETIS (Box 7.2) incorporates tools for grant application submissions and research reporting, thereby streamlining administrative workflows at Estonian research-performing organisations.

Improved policy formulation and design. Digitalisation offers new opportunities for more granular and timely data analysis to support STI policy; this should improve the allocation of research and innovation funding. Furthermore, DSIP systems often link data collected by different agencies. In this way, they provide greater context to policy problems and interventions, and offer possibilities for more integrated interagency policy design at the research or innovation system level. To give a country example: the Japanese Ministry of Education, Culture, Sports, Science and Technology and the National Institute of Science and Technology Policy have launched the SciREX data and information infrastructure to improve STI policy formulation and design. The system provides datasets to support STI policy studies. It aims to improve the accountability and transparency of public investments in R&D and strengthen the methodological frameworks used in policy evaluations.

Support of performance monitoring and management. DSIP systems offer the possibility of collating real-time policy output data. For example, in Colombia, the SCIENTI Technological Platform has developed STI indicators and metrics that support the monitoring and assessments of government-funded research. DSIP systems can allow more agile short-term policy adjustments. They can improve insights into the policy process for accountability and learning in the medium to long term, so that evaluation becomes an open and continuous process. Policy makers and delivery agencies can consider the circumstances that make it possible and meaningful to use other digitally enabled data resources, such as altmetrics of research outputs and impacts (Priem et al., 2010; Sugimoto and Larivière, 2016). They can also rely on other data collection approaches (e.g. web scraping) to complement and enhance existing approaches to assessing research.

Anticipatory intelligence. Technologies like big data analytics can help detect patterns in data that could be useful for policy, e.g. emerging research areas, technologies, industries and policy issues. Digital technologies can also support short-term forecasting of policy issues and contribute to strategic policy planning (Choi et al., 2011; Zhang et al., 2016; Peng et al., 2017; Yoo and Won, 2018). For instance, DSIP systems could identify labour demand in specific STI fields and address potential mismatches on the supply side of the labour market. In the Russian Federation, for example, the Institute for Statistical Studies and Economics of Knowledge of the National Research University Higher School of Economics, has developed the iFORA system to support foresight studies. Underpinned by advanced computational techniques, iFORA analyses large volumes of administrative data and web data to provide insights on STI breakthroughs, weak signals of change, centres of excellence and emerging technologies.

General information discovery. DSIP systems often include data on a wide range of inputs, outputs and activities. Policy makers and funders can use these data to identify leading experts in a given field (e.g. identify reviewers for project proposals), as well as centres of excellence (Guo et al., 2012; Sateli et al., 2016). This kind of information also helps researchers and entrepreneurs to identify new partners for collaboration and commercialisation. For example, the Ministry of Business, Innovation and Employment of New Zealand has developed the New Zealand Research Information System (NZRIS). It aims to raise the quality of RD&I data and improve information discovery on issues related to research and innovation. The NZRIS provides information on levels of public investments in different research areas, research collaboration networks, and leading researchers and organisations. In doing so, it aims to accelerate research commercialisation and foster close partnerships between academia and industry.

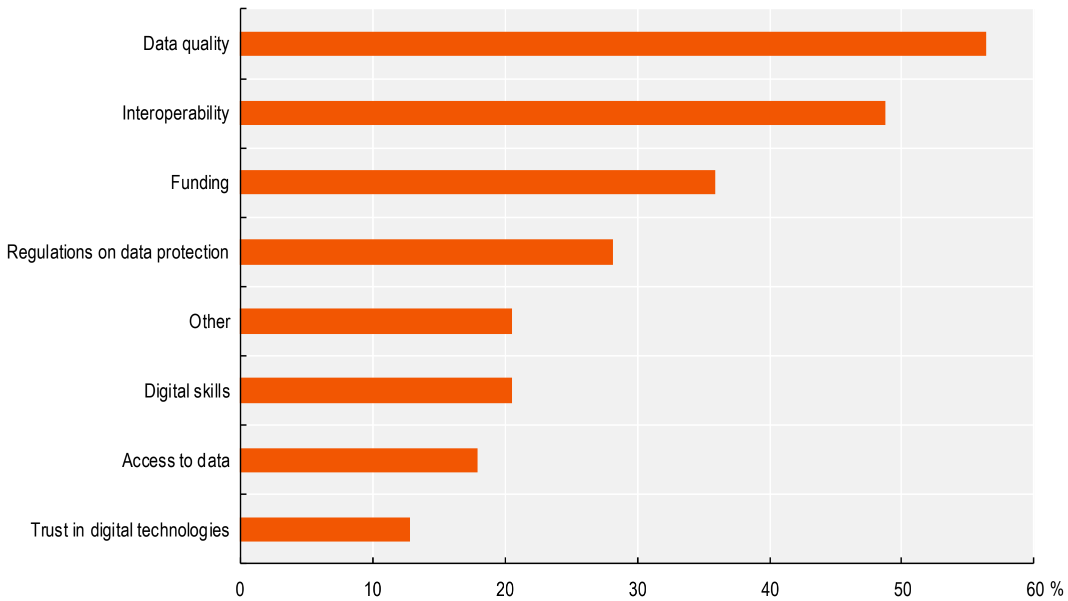
Promotion of inclusiveness in STI policy agenda setting. DSIP systems can contribute to debate with stakeholders on policy options by providing detailed information about a policy problem in an accessible way, e.g. through interactive data visualisation. The increased transparency provided by DSIP systems can empower citizens by providing them with knowledge about the nature and impacts of ongoing research and innovation. Thus, DSIP may be instrumental in building trust and securing long-term sustainable funding for research and innovation. Costa Rica, for example, has launched the Hipatia platform to help citizens better understand national scientific capabilities and the impacts of publicly funded research. Hipatia is an integrated platform created atop a variety of Costa Rican administrative databases. As a "one-stop shop" for research information in Costa Rica, Hipatia aims to improve the transparency and accountability of publicly funded research.

Source: Based on OECD (forthcoming a), *Digital Science and Innovation Policy and Governance*.

Realising the potential of DSIP involves overcoming several possible barriers. In their responses to the OECD questionnaire, DSIP administrators identified data quality, interoperability, sustainable funding and data protection regulations as the biggest challenges facing their initiatives (Figure 7.2). Other challenges cited less often were access to data, the availability of digital skills and trust in digital technologies. Policy makers wishing to promote DSIP face further systemic challenges. These include overseeing fragmented DSIP efforts and multiple (often weakly co-ordinated) initiatives (see Box 7.4, which summarises a case study of Norway's DSIP ecosystem); ensuring responsible use of data generated for other purposes; and balancing the benefits and risks of private sector involvement in providing DSIP data, components and services.

Figure 7.2. Main challenges facing DSIP initiatives

Percentage of surveyed DSIP systems



Notes: DSIP = digital science and innovation policy. Questionnaire respondents could select more than one challenge facing their DSIP initiatives.
Source: OECD survey of administrators of 39 DSIP systems in OECD countries and partner economies.

StatLink https://doi.org/10.1787/888934076077

Box 7.4. OECD case study of Norway's DSIP landscape

The Norwegian Ministry of Education and Research requested the OECD Secretariat to conduct a case study of the Norwegian landscape for DSIP. This study took place in the context of the OECD DSIP project. It involved an extensive literature review of policy issues and technological trends. The authors also analysed policy documents and reports related to Norway's DSIP landscape. In addition, they interviewed key stakeholders during a one-week mission to Norway in April 2018. Interviewees included data providers, regulators, administrators, and developers of digital infrastructures and their users.

The case study describes Norway's DSIP landscape, including its initiatives and main actors, the objectives followed and their outcomes, the level of devoted resources and future development perspectives. It shows that Norway has built substantial capabilities in preservation, access and use of comprehensive administrative datasets that could power analytical solutions used in DSIP systems.

The DSIP landscape in Norway comprises a number of digital infrastructures that collect, preserve and provide access to data on research and innovation activities. These include a digital infrastructure for sharing datasets across Norwegian government agencies; databases of Research Council Norway and Innovation Norway that include data on research inputs and outputs; and Health&Care21 research and innovation monitor, which aims to facilitate decision making on healthcare research.

One of the key elements of the Norwegian DSIP landscape is Cristin, Norway's national CRIS. Cristin is interoperable with several external digital systems managed by Norwegian government agencies and effectively serves as a major data hub on Norwegian research. Cristin provides the evidence base on which the Norwegian government performs its assessments of research performance. Apart from government bodies, all higher education institutes, research hospitals and public research institutions that receive public funding use the system to support research and strategic planning.

A distinguishing feature of Norway is its trust-based social consensus. Individuals and organisations are willing to share data about themselves with the government to improve the quality of policy making and to create more value for citizens. High levels of trust, accountability and transparency in the Norwegian government, combined with a consensus-based culture of decision making, create an excellent environment for developing DSIP initiatives.

Nevertheless, there is considerable fragmentation of efforts around DSIP. For example, several Norwegian ministries and agencies are experimenting with ML algorithms. They wish to extract actionable knowledge from fragmented datasets to support the development of statistical indicators. These, in turn, could help steer science and innovation policy initiatives in a more effective and efficient way. In some cases, these experiments – often in co-operation with external providers – have already helped design early versions of DSIP solutions. However, such efforts could benefit from a more systematic approach, involving greater co-ordination across government.

Source: OECD (forthcoming b), "OECD case study of Norway's digital science and innovation policy and governance landscape".

Interoperability

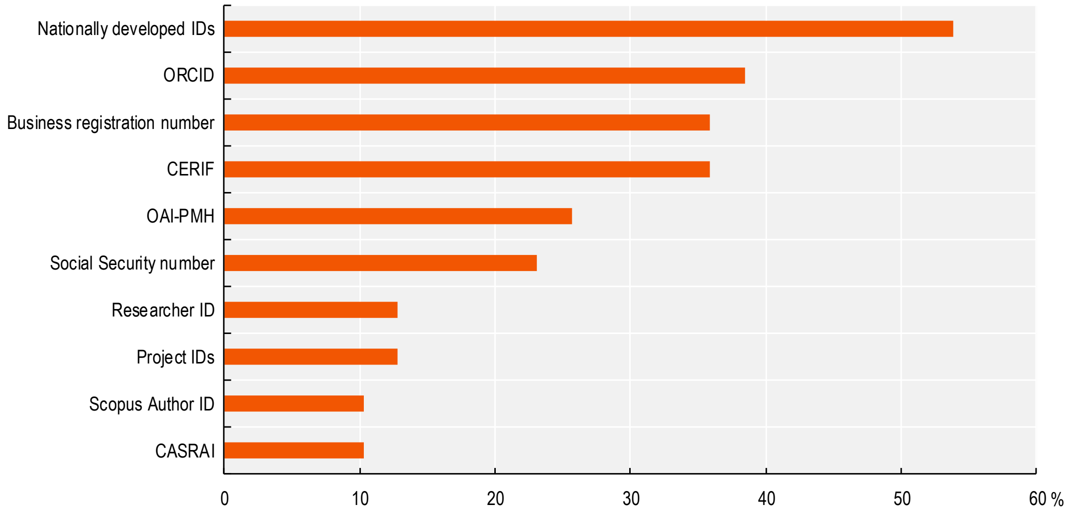
Research and innovation activities, by their nature, have high levels of pervasiveness and are shaped by a large number of stakeholders. As a result, data on the incidence and impacts of research and innovation are dispersed across a variety of public and private databases and the web. Harvesting these datasets from external sources requires the development of common data formats and other interoperability enablers including, but not limited to, application programming interfaces (APIs), ontologies, protocols and UPPIs.

An integrated and interoperable system leads to a considerable reduction in the reporting and compliance burden, freeing up time and money for research and innovation. In addition to the reduced administrative burden, interoperability allows quicker, cheaper and more accurate data matching. This, in turn, makes existing analyses less costly and more robust, and facilitates new analyses. Interoperability can produce more timely and detailed insights, enabling more responsive and tailored policy design. Furthermore, the gradual emergence of internationally recognised identifiers makes it easier to track the impacts of research and innovation activities across borders, and map international partnerships.

Interoperability raises several types of questions. On a technical level, policy makers must ask what kind of digital system can be put in place to make existing and new data interoperable. On a semantic level, they must grapple with metadata and language issues. With respect to governance, they must reflect on how all stakeholders can be aligned to agree upon an interoperability system. A specific issue concerns the role and effectiveness of data standards, particularly in a mixed ecosystem containing both legacy and new systems. In this regard, some DSIP systems use national identifications (IDs) – e.g. business registration and social security numbers – as well as country-specific IDs for researchers (Figure 7.3).

Figure 7.3. Use of interoperability enablers in DSIP systems

Percentage of surveyed DSIP systems



Notes: DSIP = digital science and innovation policy; ORCID = Open Researcher and Contributor ID; CERIF = Common European Research Information Format; CASRAI = Consortia Advancing Standards in Research Administration Information. Questionnaire respondents could select more than one type of interoperability enabler used in their DSIP initiatives.
Source: OECD survey of administrators of 39 DSIP systems in OECD countries and partner economies.

StatLink https://doi.org/10.1787/888934076096

Table 7.1. Examples of interoperability enablers in DSIP and related systems

Type	Examples
UPPIs for STI actors	• Open Researcher and Contributor ID (ORCID) • Digital object identifier (DOI) • Global Research Identifier Database (GRID) • International Standard Name Identifier (ISNI) • Ringgold ID
Author IDs generated by publishers/indexers	• Researcher ID • Scopus Author ID
Management standards for data about STI	• Common European Research Information Format (CERIF) • Consortia Advancing Standards in Research Administration Information (CASRAI) Dictionary • VIVO ontology
Protocols	• Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

Source: OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, https://doi.org/10.1787/sti_in_outlook-2018-en.

In recent years, attempts have been made to establish international standards and vocabularies to improve the international interoperability of DSIP infrastructures (Table 7.1). These include UPPIs, which assign a standardised code unique to each research entity, persistent over time and pervasive across various datasets. Box 7.5 sets out the desirable characteristics for successful UPPIs. Some UPPIs exist as an integral part of, or support for, commercial products such as publication/citation databases, research information systems, supply-chain management services, etc. Others exist solely to provide a system of identifiers for wide adoption and use. One example is ORCID, which aims to resolve name ambiguity in scientific research by developing unique identifiers for individual researchers. These systems provide a simple register of UPPIs and basic associated identity information (e.g. name and organisational affiliation for individuals, name and location for organisations). In addition, they often directly include, or incorporate links to, a wide range of further information. For example, ORCID records allow details of education, employment, funding and research works to be added manually or brought in by linking to other systems including Scopus and ResearcherID.

Box 7.5. Desirable characteristics for UPPIs

McMurry, Winfree and Haendel (6 July 2017) propose various desirable characteristics for identifiers. These have been adapted here for the specific use case under consideration (identifying individuals and organisations):

- **Defined**. The identifiers should follow a formal pattern (regular expression) that will also determine the total set of assignable identifiers; this facilitates validation and use (including by machines).
- **Persistent, stable**. The identifier should stay the same over time, wherever possible, and should never be deleted; this avoids difficulty locating records. In support of this, it is not recommended to include unnecessary detail or information liable to change in the identifier format chosen (e.g. by using a random alphanumeric code of a fixed length and structure).
- **Unambiguous**. The identifier must relate to no more than one entity locally; to avoid confusion between different entities. The identifier format chosen should seek to avoid ambiguity. For example, if an alphanumeric identifier is used, either the number zero or letter “o” should be allowed as these are easily confused by users.
- **Unique**. One entity should ideally be associated with no more than one identifier (and identifiers should never be “recycled” to apply to another entity).
- **Version-documented**. Where important changes occur, these should be clearly logged and, if necessary, new identifiers issued.
- **Web-friendly**. The id should avoid use of characters that perform specific functions html and exchange formats (e.g. XML) such as “:”, “/”, “.” to make the identifiers easier to use, search, etc.
- **Web-resolvable**. The identifier must be resolvable to a web address where the data or information about the entity can be accessed. In practice, this means the identifier should consist of a uniform resource identifier (URI) pattern (e.g. http://orcid.org/) and a local id relating to the specific record (e.g. 0000-0002-2040-1464). When used together, the URI and local id create a resolvable web address (e.g. http://orcid.org/0000-0002-2040-1464). This allows the identifier to be easily checked to ensure it relates to an actual record and that the record relates to the correct entity.
- **Free to assign**. The identifier should ideally be assigned at no costs; this reduces barriers to adoption.
- **Open access (OA) and use**. The identifier appropriate metadata (e.g. the name of the entity to which it relates) should be able to be transparently referenced and actioned (e.g. in a public index or search) anywhere, by anyone, and for any reason; this enables integration on the basis of practical usefulness.
- **Documented**. The identifier scheme, its operation, etc. should be clearly documented; this enables users to understand the system and encourages consistent use. Documented privacy and dispute resolution policies are also important factors.

Source: McMurry, Winfree and Haendel (6 July 2017), “Bad identifiers are the potholes of the information superhighway: Take-home lessons for researchers”, http://blogs.plos.org/biologue/2017/07/06/bad-identifiers-potholes-of-information-superhighway/.

As an UPPI system gains traction there may be a "network effect", whereby each additional registrant increases the value of the system to all users. Eventually the UPPI system may become a generally expected way for entities to unambiguously identify each other. This results in strong incentives to join for those not yet registered.

Besides UPPIs, APIs have become an industry standard for integrating data. They enable machine-to-machine interactions and data exchanges. Within a framework of digital government initiatives, several countries have started to proliferate APIs across the whole landscape of government websites and databases, improving data reuse. Improvements in data access to administrative datasets have positive impacts on the functionality and reliability of the results of analyses delivered by DSIP systems.

Aside from government agencies and other public funders, RD&I-performing organisations store a significant share of research and innovation data. The Common European Research Information Format (CERIF) and metadata formats by Consortia Advancing Standards in Research Administration Information (CASRAI) were originally designed to serve the needs of HEIs in data management. Some DSIP systems use them to harvest curated data from research institutes and directly apply them in analysis (Box 7.6).

Box 7.6. Management standards for data about STI

CERIF

CERIF is a standard maintained by the international not-for-profit organisation EuroCRIS since 2002. It ensures a uniform management and exchange of research information by providing data models (entities, attributes and relationships), exchange models, metadata models and controlled vocabulary terms. CERIF covers related information on publications, projects, organisations, equipment, events, individuals, language, facilities, patents, products and services (Jörg et al., 2012). An important feature of CERIF is the provision of connectivity among different metadata standards by enabling conversion of one standard into another (Jeffery and Asserson, 2016).

CASRAI

CASRAI is an international not-for-profit organisation founded in 2006. It helps key stakeholders in data curation to develop standard agreements for making research information exchange more efficient. Agreements entail standards for managing the full data cycle. Implementation of CASRAI standards can help organisations to improve data quality, interoperability and accessibility. They do this by filtering information (agreements on report format templates) and disambiguating it (agreements on shared glossaries). CASRAI is mainly used in Europe, the United States and Canada; the rest of the world tends to use other standards. Even still, a large number of digital tools in one way or another use CASRAI standards. For example, ORCID uses CASRAI research-output report formats and glossaries, and Snowball Metrics uses CASRAI standard information agreements (CASRAI, 2016).

VIVO

Semantic ontologies can also help address the problem of interoperability in DSIP infrastructures. Launched in 2003 by Cornell University, the VIVO project aims to develop an open-source software and an ontology for research information enabling federated search for research partners. The VIVO ontology includes information on organisations, researchers, activities and their relationships. It builds linkages among various data items to provide a consistent and connected perspective on research and enables more effective data reuse. In a similar vein to VIVO, other initiatives like Semantic Web for Research Communities and Advanced Knowledge Technologies also provide ontologies for scientific research.

Source: OECD (forthcoming a), *Digital Science and Innovation Policy and Governance*.

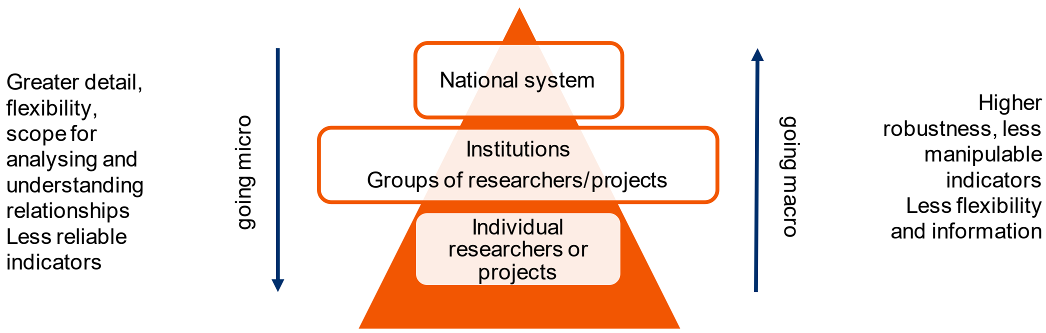
Using DSIP infrastructures in research assessment

In recent years, research funders, research-performing institutes and researchers have faced increasing pressure to demonstrate the value and impact of research. Budgetary discussions implicitly or explicitly compare the value of the marginal dollar placed in science versus other policy areas. All policy areas try to make their best possible case, and data-based assessment has become a core component of evidence-based policy and strategy discussions. As a particular class of evidence-based assessment, data-driven assessments are responding to the complexity of research and innovation systems, and the need for more efficient and faster decisions. They use the opportunity provided by the digital trace of many scientific research activities, as well as growing data processing capacities.

However, there are significant risks that the procedures of data-based assessment fail to meet their intended objectives. A key risk of data-based systems is giving up control over what drives assessment. Decisions, for example, are based on what data are available in a quantitative, apparently compact fashion. Data-based assessment can provide a valid perspective only as long as available data encompass all the relevant parts of the phenomena of interest. Two steps can address this concern. First, policy makers need a broad sense of what science actors of different types do and the extent to which existing data capture these activities and their outcomes. Second, they need to identify to what extent such data can be actually deployed for assessment. This will depend on their accessibility and interoperability with other data sources.

The level of analysis at which impact is examined is critical. One of the great advantages of digitalisation is the technical ability to operate with large, linked databases at very fine levels of granularity so that information is not necessarily lost in the process of aggregation. This micro perspective has, however, somewhat contributed to a loss of perspective in terms of what can be concluded from inferences in such data. One prominent example of a disconnect between data users and producers relates to the confusion between using data to assess features in the performance of individual researchers, their institutions and the country or broader area as a whole, as highlighted in Figure 7.4. For example, while the journal Impact Factor was born out of the need to inform librarians' decisions concerning what titles to acquire and store, over time this became a surrogate measure used to assess the quality of individual researchers and their research outputs. Despite extensive academic discussion of the limitations of journal-based metrics (Moed et al., 2012), these continue to be widely used. Such misuses of data have generated calls for concerted efforts to create an open, sound and consistent system for measuring *all* activities that contribute to academic productivity.

Figure 7.4. Trade-offs between the micro and macro levels of data analysis and indicators



Source: OECD (forthcoming a), *Digital Science and Innovation Policy and Governance*.

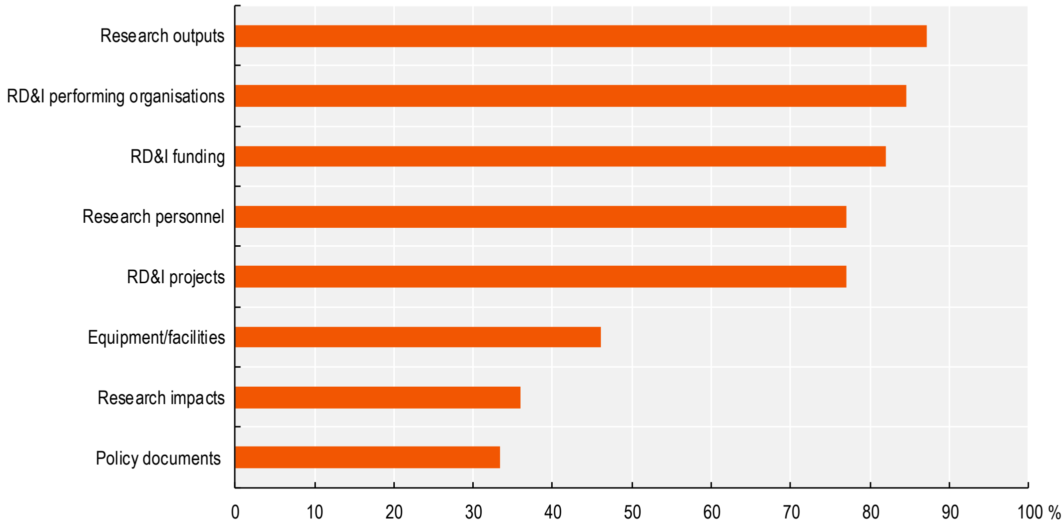
DSIP infrastructures could reinforce existing misuses of data, which could distort the incentives and behaviour of individual researchers and research-performing organisations (Hicks et al., 2015; Edwards and Siddhartha, 2017). But DSIP also brings with it the promise that one day most, if not all, relevant dimensions of research activity and interaction might be represented digitally. This can be described as the "promise of altmetrics". Some argue the emergence of web-based new data sources, especially those generated within online social media platforms, can provide timelier insights into relevant and hitherto unknown dimensions (Priem et al., 2010).

It has been argued that altmetrics could support the assessment of increasingly important, non-traditional scholarly products like datasets and software, which are under-represented in the citation indices frequently used for assessment. Altmetrics could also reward impacts on wider audiences outside the publishing core, such as practitioners or the public in general. The altmetrics movement promotes the use of metrics generated from social media platforms as a source of evidence of research impact broader and timelier than citations. Altmetrics have also been advanced as part of the infrastructure required to facilitate open science, and as an aid to filtering fast-growing amounts of information outside or at the margins of traditional peer-review mechanisms. However, as with more traditional metrics, such as citation counts, questions remain over the extent to which altmetrics qualify as signals of research impact.

More than half of the DSIP systems surveyed play a role in research assessment. Nearly 90% collect information on research outputs and more than one-third gather information on research impacts (Figure 7.5). Some, like the Cristin system in Norway, the Lattes Platform in Brazil and the METIS system in the Netherlands, are the primary sources of data for national research assessments. Few use altmetrics in their research assessments.

Figure 7.5. Types of information harnessed for DSIP systems

Percentage of surveyed DSIP systems



Notes: DSIP = digital science and innovation policy; RD&I = research, development and innovation. Questionnaire respondents could select more than one type of information harnessed by their DSIP initiatives.
Source: OECD survey of administrators of 39 DSIP systems in OECD countries and partner economies.

StatLink https://doi.org/10.1787/888934076115

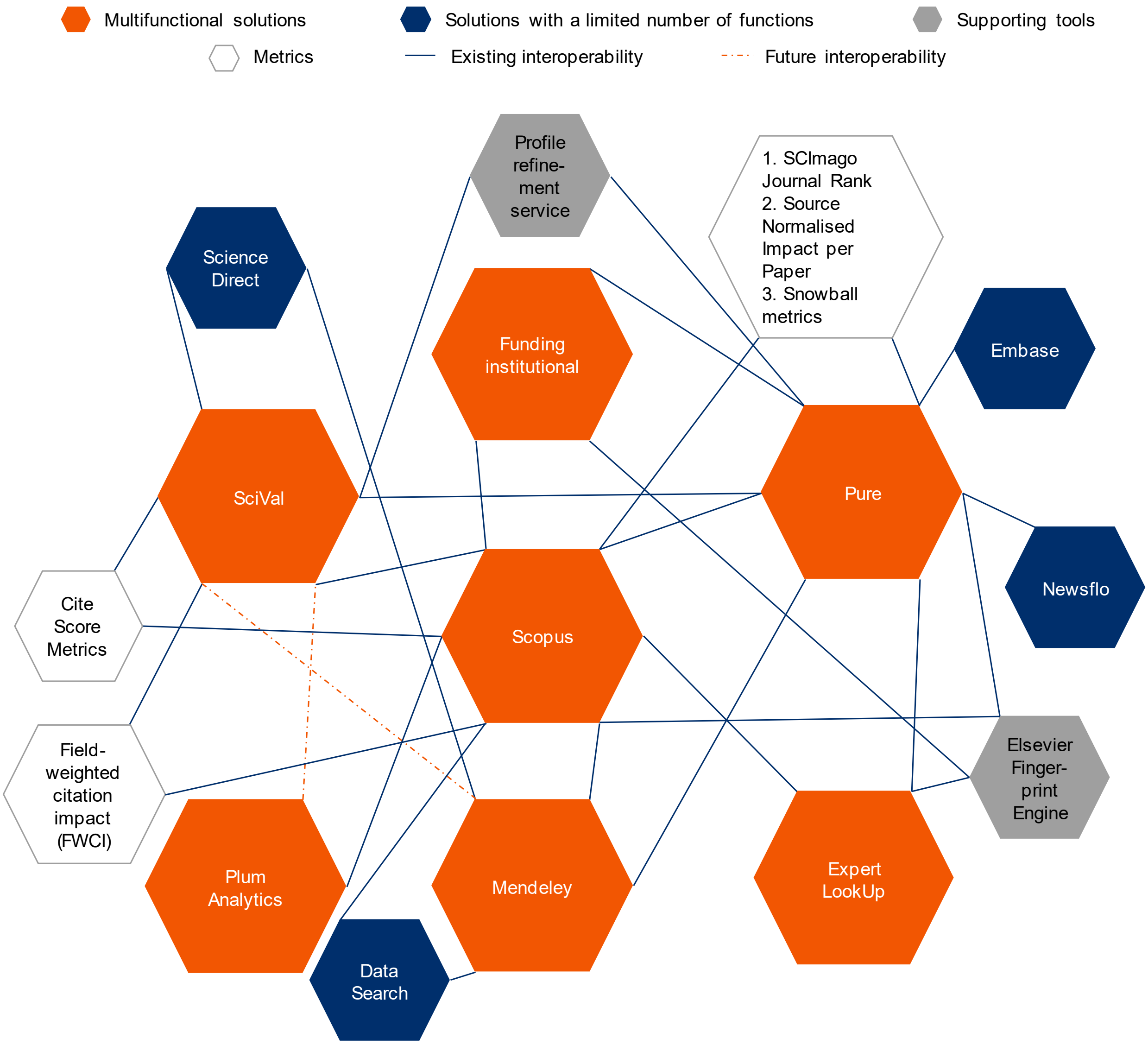
The roles of non-government actors in DSIP

Non-government actors are emerging as one of the main forces for digitalisation of science and innovation policy. Solutions developed by commercial companies and not-for-profit organisations can provide governments with essential capabilities in data management and analytics at an agreed cost and within a required timeframe.

The private sector exerts manifold impacts on the development of DSIP initiatives. DSIP systems can potentially use digital products and services designed by the private sector as building blocks. Essentially, they can extend the functionality and increase the value of DSIP systems for their stakeholders. The private sector designs technological architectures, develops digital tools for data management and provides consulting services related to launching and maintaining digital infrastructures. However, co-operation between the public and private sectors is multidimensional. It is not confined to the purchase of off-the-shelf solutions; there is also considerable co-operation in developing new solutions. For instance, administrators of the Flanders Research Space DSIP system are co-operating with IBM to develop a web-scraping tool that retrieves information on research activities scattered across the web.

The large academic publishers, Elsevier and Holtzbrinck Publishing Group, together with the analytics firm, Clarivate Analytics, are particularly active. They are developing and bundling a mix of products and services into platforms that mimic many features of fully fledged DSIP systems (Figure 7.6 shows the example of Elsevier). Several products developed by these firms, including bibliographic databases, unique identifiers and organisational CRIS (Box 7.2), are often key components in governments' DSIP systems.

Figure 7.6. Interoperability within Elsevier's portfolio of in-house digital products



Source: OECD (forthcoming a), *Digital Science and Innovation Policy and Governance*.

In addition, digital giants like Alphabet and Microsoft, and national technology companies such as Baidu (People's Republic of China) and Naver (Korea) have all designed platforms to search academic outputs. The impact of these companies on the digitalisation of science and innovation policy is limited. However, given their coverage of information on research outputs, these platforms could become key elements in national DSIP systems. For instance, the Academic Knowledge API of Microsoft Academic Graph enables the retrieval of information on publications, citations, contributors, institutions, fields of study, journals and conferences (API, n.d.; Microsoft, n.d.). Developers of DSIP systems can use these data for further analysis, which could spark competition with other established commercial databases of bibliographic information (such as Scopus). Academic search engines (Google Scholar, Microsoft Academic, Baidu Scholar and Naver Academic) could collect information on research publications and citations. This could potentially support research assessments and international benchmarking of research participants, including university rankings (Daraio and Bonaccorsi, 2017; Kousha, Thelwall and Abdoli, 2018).

Another group of firms active in the DSIP area are providers of research administration tools for public funders and research-performing organisations. These tools provide the evidence base for national research assessments and support decisions on allocation of public funding. Some of these companies are involved in consultancy projects to support evidence-informed science and innovation policy making. Science-Metrix, a subsidiary of the Canadian research information management firm 1Science Inc., is a case in point. It was commissioned in 2018 to develop methods and indicators of research and innovation activities for the US National Science Foundation (Côté et al., 2018).

Harnessing these private sector developments for use in public DSIP systems has many potential benefits. Solutions can be implemented quickly and at an agreed cost, sparing the public sector the need to develop the necessary in-house skills beforehand. Private companies can also promote interoperability through their standards and products, which can expand the scope and scale of data within a DSIP system. But there are also risks. For example, outsourcing data management activities to the private sector may result in a loss of control over the future development of DSIP systems. In addition, reliance on proprietary products and services may lead to discriminatory access to data, even if these concern research activities funded by the public sector. Finally, the public sector's adoption of commercial standards for metrics may drive the emergence of private platforms exhibiting network effects that are difficult to contest.

Charities and not-for-profit organisations also contribute to DSIP, as shown above in the discussion of interoperability enablers. These organisations can also directly fund and design DSIP solutions. For instance, the Alfred P. Sloan Foundation has financially supported projects to collect systematic evidence on impacts of publicly funded research (e.g. ETOILE, UMETRICS), to provide free access and sharing of research outputs (e.g. arXiv.org, FORCE11, Impactstory) and to aid data disambiguation (improvement of citations, development of unique identifiers). In another example, an Australian-based not-for-profit social enterprise, Cambia, in co-operation with Queensland University of Technology, has launched the Lens, an open platform for "innovation cartography". The platform aggregates data from databases of several national and international patent offices and scholarly datasets including PubMed, Crossref and Microsoft Academic to provide OA to disambiguated and linked patent information (Lens, n.d.). A number of add-on tools and services provide actionable intelligence that decision makers can use. For example, policy makers can use Lens PatCite to identify, disambiguate and link scientific articles cited in patents (Lens PatCite, n.d.).

Due to their free pricing and high levels of functionality, digital solutions designed by not-for-profit organisations are widely adopted by public organisations, as well as commercial firms. Indeed, in many cases, they serve as important elements of commercial DSIP solutions, enhancing their functionality and contributing to their interoperability. Administrators of several surveyed DSIP systems opted mostly for open software and free digital solutions to better ensure the financial sustainability of their operations and to mitigate the risks of vendor lock-ins.

Conclusion

The digital transformation of STI policy and its evidence base is still in its early stages. This means STI policy makers can take an active stance in shaping DSIP ecosystems to fit their needs. This will require strategic co-operation, through significant interagency co-ordination and sharing of resources (such as standard digital identifiers), and a coherent policy framework for data sharing and reuse in the public sector. Since several government ministries and agencies formulate science and innovation policy, DSIP ecosystems should be founded on the principles of co-design, co-creation and co-governance (OECD, 2018).

This chapter has highlighted some of the challenges facing DSIP. Interoperability remains a major barrier, despite the recent proliferation of identifiers, standards and protocols. There is the potential opportunity for policymakers to influence the development of international UPPI systems in terms of target populations, information captured, compatibility with statistical systems, and especially adoption both by entities and by potential users. In particular, international efforts related to data documentation and the development of standards for metadata could be consolidated to improve data interoperability.

DSIP systems can help broaden the evidence base on which research is assessed by, for example, incorporating altmetrics. They can also empower a broad group of stakeholders to participate more actively in the formulation and delivery of science and innovation policy. However, there is also the danger that these systems reinforce existing data misuses. DSIP systems should uphold and endorse recent initiatives that promote best practices in the responsible use of data. These include the San Francisco Declaration on Research Assessment (ASCB, n.d.) and Leiden Manifesto (Hicks et al., 2015).

Finally, governments can usefully co-operate with the private and not-for-profit sectors in developing and operating DSIP systems. However, they should ensure public data remains outside of "walled gardens" and open for others to readily access and reuse. They should also avoid vendor lock-ins, deploying systems that are open and agile. In a fast-changing environment, this will provide governments with greater flexibility to adopt new technologies and incorporate unexploited data sources in their DSIP systems.

References

API (n.d.), "Academic knowledge", webpage, https://docs.microsoft.com/en-us/azure/cognitive-services/academic-knowledge/home (accessed 16 July 2018).

ASCB (n.d.), "San Francisco Declaration on Research Assessment", webpage, www.ascb.org/dora (accessed 18 August 2016).

Bauer, M.W. and A. Suerdem (2016), "Relating 'science culture' and innovation", presentation at the OECD Blue Sky Forum on Science and Innovation Indicators, Ghent, 19-21 September.

CASRAI (2016), "CASRAI Impacts – 10-year anniversary, Building bridges for research information users", CASRAI.

Choi, S. et al. (2011), "SAO network analysis of patents for technology trends identification: A case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells", *Scientometrics*, Vol. 88/3, Springer International Publishing, Cham, Switzerland, pp. 863-883, https://doi.org/10.1007/s11192-011-0420-z.

Côté, G. et al. (2018), *Bibliometrics and Patent Indicators for the Science and Engineering Indicators 2018 – Technical Documentation*, Science-Matrix, Montreal.

Daraio, C. and A. Bonaccorsi (2017), "Beyond university rankings? Generating new indicators on universities by linking data in open platforms", *Journal of the Association for Information Science and Technology*, Vol. 68/2, Wiley Online Library, pp. 508-529.

Edwards, M. and R. Siddhartha (2017), "Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition", *Environmental Engineering Science*, Vol. 34/1, Mary Ann Liebert Inc., pp. 51-61, http://online.liebertpub.com/doi/pdf/10.1089/ees.2016.0223.

Gibson, E. et al. (2018), "Technology foresight: A bibliometric analysis to identify leading and emerging methods", *Foresight and STI Governance,* Vol. 12/1, National Research University Higher School of Economics, Moscow, pp. 6-24.

Guo, Y. et al. (2012), "Text mining of information resources to inform forecasting innovation pathways", *Technology Analysis & Strategic Management*, Vol. 24/8, Routledge, London, pp. 843-861, https://doi.org/10.1080/09537325.2012.715491.

Hicks, D. et al. (2015), "Bibliometrics: The Leiden manifesto for research metrics", *Nature*, Vol. 520/7548, Nature Research, Springer, pp. 429-431, www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351.

Jeffery, K. and A. Asserson (2016), "Position paper: Why CERIF?", presentation at smart descriptions and smarter vocabularies workshop, Amsterdam, 30 November-1 December, www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_15.

Johnson, R., O. Fernholz and M. Fosci (2016), "Text and data mining in higher education and public research. An analysis of case studies from the United Kingdom and France", Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation, Paris, https://adbu.fr/competplug/uploads/2016/12/TDM-in-Public-Research-Final-Report-11-Dec-16.pdf.

Jörg, B. et al. (2012), "CERIF 1.3 full data model (FDM): Introduction and specification", 28 January, EuroCRIS, www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf.

Kayser, V. and K. Blind (2017), "Extending the knowledge base of foresight: The contribution of text mining", *Technological Forecasting and Social Change*, Vol. 116C, Elsevier, Amsterdam, pp. 208-215.

Kong, D. et al. (2017), "Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country", *Technological Forecasting and Social Change*, Vol. 119, Elsevier, Amsterdam, pp. 80-97.

Kousha, K., M. Thelwall and M. Abdoli (2018), "Can Microsoft Academic assess the early citation impact of in-press articles? A multi-discipline exploratory analysis, *Journal of Informetrics,* Vol. 12/2, arXiv:1802.07677, Cornell University, pp. 287-298.

Lens (n.d.), "About the Lens", webpage, https://about.lens.org/ (accessed 31 August 2018).

Lens PatCite (n.d.), "Lens PatCite" webpage, www.lens.org/lens/patcite (accessed 31 August 2018).

Mateos-Garcia, J. (6 April 2017), "We are building a formidable system for measuring science – but what about innovation?", Nesta blog, www.nesta.org.uk/blog/we-are-building-a-formidable-system-for-measuring-science-but-what-about-innovation/.

McMurry, J., L. Winfree and M. Haendel (6 July 2017), "Bad identifiers are the potholes of the information superhighway: Take-home lessons for researchers", PLOS Biologue Community blog, http://blogs.plos.org/biologue/2017/07/06/bad-identifiers-potholes-of-information-superhighway/.

Microsoft (n.d.), "Microsoft Academic Graph", webpage, www.microsoft.com/ (accessed 16 July 2018).

Moed, H.F. et al. (2012), "Citation-based metrics are appropriate tools in journal assessment provided that they are accurate and used in an informed way", *Scientometrics*, Vol. 92/2, Springer, pp. 367-376.

OECD (forthcoming a), *Digital Science and Innovation Policy and Governance*, OECD Publishing, Paris.

OECD (forthcoming b), "OECD case study of Norway's digital science and innovation policy and governance landscape", *OECD Science, Technology and Innovation Policy Papers*, OECD Publishing, Paris.

OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-en.

Park, H., J. Yoon and K. Kim (2013), “Using function-based patent analysis to identify potential application areas of technology for technology transfer”, *Expert Systems with Applications*, Vol. 40/13, Elsevier, Amsterdam, pp. 5260-5265.

Peng, H. et al. (2017), “Forecasting potential sensor applications of triboelectric nanogenerators through tech mining”, *Nano Energy*, Vol. 35, pp. 358-369, Elsevier, Amsterdam, https://doi.org/10.1016/j.nanoen.2017.04.006.

Priem, J. et al. (2010), “Altmetrics: A Manifesto”, webpage, http://altmetrics.org/manifesto (accessed 5 February 2017).

Sateli, B. et al. (2016), “Semantic user profiles: Learning scholars’ competences by analyzing their publications”, in A. González-Beltrán, F. Osborne and S. Peroni (eds.), *Semantics, Analytics, Visualization. Enhancing Scholarly Data*, *SAVE-SD 2016, Lecture Notes in Computer Science*, Vol. 9792, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-319-53637-8_12.

Shapira, P. and J. Youtie (2006), “Measures for knowledge-based economic development: Introducing data mining techniques to economic developers in the state of Georgia and the US South”, *Technological Forecasting and Social Change*, Vol. 73/8, Elsevier, Amsterdam, pp. 950-965.

Sugimoto, C. and V. Larivière (2016), “Social media indicators as indicators of broader impact”, presentation at the OECD Blue Sky Forum on Science and Innovation Indicators, Ghent, 18 September, www.slideshare.net/innovationoecd/sugimoto-social-media-metrics-as-indicators-of-broader-impact.

Wolfram, D. (2016), “Bibliometrics, information retrieval and natural language processing: Natural synergies to support digital library research”, in *Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries*, ACL Anthology, www.aclweb.org/anthology/.

Yoo, S.H. and D. Won (2018), “Simulation of weak signals of nanotechnology innovation in complex system”, *Sustainability,* Vol. 10/2/, MDPI, Basel, pp. 486, https://doi.org/10.3390/su10020486.

Yoon, J. and K. Kim (2012), “Detecting signals of new technological opportunities using semantic patent analysis and outlier detection”, *Scientometrics*, Vol. 90/2, Springer, pp. 445-461.

Yoon, J., H. Park and K. Kim (2013), “Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis”, *Scientometrics,* Vol. 94/1, Springer, pp. 313-331.

Zhang, Y. et al. (2016), “Technology roadmapping for competitive technical intelligence”, *Technological Forecasting and Social Change*, Vol. 110, Elsevier, Amsterdam, pp. 175-186, https://doi.org/10.1016/j.techfore.2015.11.029.

The Digitalisation of Science, Technology and Innovation

KEY DEVELOPMENTS AND POLICIES

This report examines digitalisation's effects on science, technology and innovation and the associated consequences for policy. In varied and far-reaching ways, digital technologies are changing how scientists work, collaborate and publish. While examining these developments, this book also assesses the effects of digitalisation on longstanding policy themes, from access to publicly funded research data, to the diffusion of technology and its absorption by firms. New and emerging topics are also explored. These include the roles of artificial intelligence and blockchain in science and production, using digital technology to draw on the collective intelligence of the scientific community, advances in the digitalisation of biotechnology, and possible "dark sides" of digitalisation.

Consult this publication on line at ***https://doi.org/10.1787/b9e4a2c0-en***.

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases.
Visit ***www.oecd-ilibrary.org*** for more information.

ISBN 978-92-64-97802-7

OECD*publishing*
www.oecd.org/publishing

9 789264 978027