Data-Driven Innovation

BIG DATA FOR GROWTH AND WELL-BEING

OECD

SYNTHESIS CHAPTER

Data-Driven Innovation

BIG DATA FOR GROWTH AND WELL-BEING

Please cite this publication as:
OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264229358-en

ISBN 978-92-64-22934-1 (print)
ISBN 978-92-64-22935-8 (PDF)

Photo credits: Cover © montage by Baseline Arts

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

Preface

Social and economic activities are increasingly migrating to the Internet. The cost of data collection, storage and processing continues to decline dramatically. Ever larger volumes of data will be generated from the Internet of Things, smart devices, and autonomous machine-to-machine communications. We are now at the cusp of a new era, in which "big data" will play a transformative role.

The "datafication" of the economy and society holds many promises in a wide range of areas, from health to agriculture, from public governance to innovation, and from education to the environment, to name just a few. The "low-hanging fruit" of data-driven innovation (DDI) may be clear, but the full scope of potential benefits is much more difficult to grasp, resulting in opportunities that may be lost.

Seizing these benefits poses a formidable challenge to policymakers. In the years ahead, the pivot to a data-driven world will have important implications for policies ranging from privacy, consumer policy, competition, taxation, innovation and especially jobs and skills.

We will need, for example, to recast how we think about infrastructure in the 21st Century, and expand it to encompass broadband networks, cloud computing and data itself. Ensuring that DDI leads to growth will require focusing on small and medium enterprises and high value-added services, such as design and engineering. The questions of access and ownership of data are also essential. Governments will need to understand and strike the right balance between the social benefits of "openness", and individuals' and organisations' legitimate concerns about such openness.

As well as a catalyst for growth, innovation and productivity gains, DDI will be a disruptive force, with far-reaching effects on the economy and well-being. Policymakers will need to consider the trade-offs, complementarities and possible unintended consequences both of their policy actions - and of inaction. We need to ensure that the benefits of DDI are widely shared, and that far from creating new divides they do not leave anyone behind.

This will be no easy task. This report helps policymakers to be proactive, instead of reactive, by outlining these trade-offs. It uses the breadth of the OECD's expertise to outline the contours of this phenomenon and frames a number of the policy dialogues that need to occur so as to fully benefit from the coming era of ubiquitous data.

Angel Gurría
Secretary-General
OECD

Foreword

Early in 2011 the OECD began a project on *New Sources of Growth: Knowledge-based Capital* (KBC). The project was inspired by findings from the OECD's *Innovation Strategy*, originally published in 2010 and now updated to 2015 (forthcoming). According to these findings, many innovating firms invest, beyond R&D, in a broader range of intangibles assets including i) intellectual property (e.g. patents, trademarks, copyrights, trade secrets, designs); ii) digital data and information (e.g. data and analytics); and iii) economic competencies (e.g. organisational capital and firm-specific skills). These intangible assets are referred to as knowledge-based capital (KBC).

This report focuses on digital data and analytics and their effects on innovation, growth and well-being. It aims to improve the evidence base on the role of data-driven innovation (DDI) in boosting productivity growth and contributing to well-being. It also offers policy guidance for maximizing the benefits of DDI and mitigating the associated economic and societal risks. The insights in the report are intended to help policy makers better understand DDI, incorporate its multidimensionality into policy design and "identify trade-offs, complementarities and unintended consequences of policy choices". This report contributes to the goal of building and maintaining "resilient economies and inclusive societies" while enhancing the productivity and competitiveness of industries, as articulated in the OECD Ministerial Council Statements of 2014 and 2015.

The work on DDI has drawn on expertise from different directorates within the OECD. Supported with financial resources from the Secretary-General's Central Priority Fund and in-kind contributions from the Netherlands, the Directorate for Science, Technology and Innovation led the two-year effort. Other partners have been the Directorate for Employment, Labour and Social Affairs, and the Directorate for Public Governance and Territorial Development. Owing to this co-operative effort, the publication's different chapters were discussed and declassified by various OECD committees, including the Committee on Digital Economy Policy which had oversight responsibility for the project; the Committee on Consumer Policy; the Committee for Scientific and Technological Policy; the Health Committee; and the Public Governance Committee. The comments and inputs received from delegates to these official OECD bodies are gratefully acknowledged.

The material presented in this book will feed ongoing and future OECD projects, most notably the OECD project on the Next Production Revolution (NPR, http://oe.cd/npr). Further information on the work on DDI, including follow-up work, will be available on the OECD website, at http://oe.cd/bigdata.

ACKNOWLEDGEMENTS

The work on Data-Driven Innovation represents an OECD collective effort led and co-ordinated by Christian Reimsbach-Kounatze (Information Economist and Policy Analyst, Division for Digital Economy Policy) under the guidance and oversight of Andrew Wyckoff (Director, Directorate for Science, Technology and Innovation). Jørgen Abild Andersen (Denmark), Chair of the OECD Committee on Digital Economy Policy, and Anne Carblanc (Head of Division, Division for Digital Economy Policy) provided directions and advice throughout the process.

Chapters 1, 2, 3, 4 and 6 were authored by Mr. Reimsbach-Kounatze. In particular, Chapter 1 ("The phenomenon of data-driven innovation") benefited from input from Sabine Gerdon; Chapter 2 ("Mapping the global data ecosystem and its points of control") from Andrea de Panizza and the Netherlands Organisation for Applied Scientific Research (TNO – Jop Esmeijer, Bas Kotterink, Anne F. van Veenstra, Tom Bakker, Merel Ooms, Anna van Nunen, and Silvain de Munck); Chapter 3 ("How data now drive innovation") from Rudolf van der Berg; and Chapter 6 ("Skills and employment in a data-driven economy") from Cristina Serra Vallejo, Sabine Gerdon and the Research Institute for Applied Knowledge Processing, Germany (FAW/n – Estelle L.A. Herlyn, Thomas Kämpke, Franz J. Radermacher, and Dirk Solte). Chapter 5 ("Building trust for data-driven innovation") was authored by Laurent Bernat and Michael Donohue. Chapter 7 ("Promoting data-driven scientific research") was written by Giulia Ajmone Marsan, with guidance from Mario Cervantes. Chapter 8 ("The evolution of health care in a data-rich environment") was authored by Jillian Oderkirk and Elettra Ronchi. Chapter 9 ("Cities as hubs for data-driven innovation") was written by David Gierten, with input from TNO. Chapter 10 ("Governments leading by example with public sector data") was authored by Barbara Ubaldi, with contributions from Graham Vickery. Randall Holden edited the book and Janine Treves, James Arkinstall and Kate Brooks provided support with the overall presentation.

Some chapters benefited from additional expertise within the OECD through extensive rounds of comments. Special thanks therefore go to: John Davies (Competition Division of the Directorate for Financial and Enterprise Affairs); Vincenzo Spiezia (Economic Analysis and Statistic Division of the Directorate for Science, Technology and Innovation); Jesse Eggert, Eric Robert and Liz Chien (Digital Economy Team of the Center for Tax Policy and Administration); and Guillermo Montt (Division for Employment Analysis and Policy of the Directorate for Employment, Labour and Social Affairs).

Analysis and policy conclusions also benefited from advice provided by an international panel of independent experts including Devdatt Dubhashi (Professor, Department of Computer Science and Engineering, Chalmers University of Technology), Brett Frischmann (Director, Cardozo Intellectual Property & Information Law Program and Professor of Law, Benjamin N. Cardozo School of Law), Jakob Haesler (Co-founder, tinyclues), Simon Hania (Corporate Privacy Officer, TomTom), and Sarah Spiekermann (Head of the Institute for Management Information Systems,

Vienna University of Economics and Business). Thanks also go to Brian Kahin (Fellow at the MIT Sloan School Center for Digital Business) for his very informative comments. In addition, the report benefited from the advice of a panel of delegates drawn from the participating OECD committees. Many thanks go to Andre Loth (France), Emilio Garcia Garcia and Ruth Del Campo Becares (Spain), Tony O'Connor (United Kingdom), Hugh Stevenson and Stacy Feuer (United States), and Robin Wilton (Internet Technical Advisory Committee to the OECD).

The work on DDI and this book also benefited from discussions with the authors of some of the most prominent literature on "big data". Thanks therefore go to: Kenneth Cukkier and Viktor Mayer-Schönberger, the authors of "Big Data: A Revolution That Will Transform How We Live, Work and Think"; Robert Kirkpatrick and his team at United Nations Global Pulse for their work on big data for well-being and development; Carl Kalapesi and Joel Nicholson at the World Economic Forum (WEF) for their work on personal data and big data for development; Hasan Bakhshi and Juan Mateos-Garcia at Nesta for their work on the "datavores"; and Paul Hofheinz at the Lisbon Council and Michael Mandel at the Progressive Policy Institute (PPI) for their work on the transatlantic policy issues raised by big data.

Finally, two major events helped scope, develop and test analytic and policy ideas with academics, policymakers and practitioners. One was the *2012 Technology Foresight Forum* (http://oe.cd/tff2012), held at the OECD headquarters in Paris, France, on 22 October 2012; The second event was the *4th Global Forum on the Knowledge Economy* (GFKE, http://oe.cd/gfke2014), held in Tokyo, Japan, on 2-3 October 2014, and co-organised with, and hosted by, the Japanese Ministry for Internal Affairs and Communications and the Japanese Ministry of Economy, Trade and Industry. Special thanks go to the hosts, to Hajime Oiso, Aki Irie and Yuka Miyazaki who helped organise the GFKE, and to all the participants of both events.

Table of contents

Figures

Tables

Abbreviations

AD	Alzheimer's disease
ADRN	Administrative Data Research Centres, United Kingdom
AIC triad	Availability, integrity and/or confidentiality of information
AMI	Acute myocardial infarction
APIs	Application programming interfaces
ATS	Algorithmic trading systems
BiOS Initiative	Biological Innovation for Open Society
BPP	Billion Price Project
CAGR	Compound annual growth rate
CancerLinQ	American Society of Clinical Oncology's Cancer Learning Intelligence Network for Quality
CCD	Ciudad Creativa Digital, Guadalajara, Mexico
CCLA	City Climate Leadership Award
ccTLDs	Country code top-level domains
CDC	Centers for Disease Control and Prevention, United States
CDNs	Content delivery networks
CEPS	Centre for European Policy Studies
CER	Comparative effectiveness research
CLA	Contributor Licence Agreement
CODATA	Committee on Data for Science and Technology
CONIYT	National Commission of Technological Research, Chile
CSV	Comma-separated values
CT	Computed tomography
DBMS	Database management system
DDI	Data-driven innovation
DEC	Department of Environmental Conservation, New York State
DoS	Denial of service
DRM	Digital rights management
DSSs	Decision support systems
DW	Data warehouse
EBI	European Bioinformatics Institute
ECHO	European Collaboration for Healthcare Optimization
EDF	Électricité de France
EDW	Enterprise data warehouse
EHRs	Electronic health records
EITC	Earned income tax credits
EMBL	European Molecular Biology Laboratory
EMIF	European Medical Information Framework
ENoLL	European Network of Living Labs
EP	European Parliament
EPRs	Electronic personal records
ERDF	Électricité Réseau Distribution France

ERP	Enterprise resource planning
Esri	Environmental Systems Research Institute
ESSC	European Statistical System Committee
ETDE	Energy Technology Data Exchange
eTRIKS	Delivering European Translational Information & Knowledge Management Services
EU-ADR	EU Advanced Drug Reporting initiative or should ADR really be "Adverse Drug Reactions"? Please confirm
EUNOIA	Evolutive User-centric Networks for Intraurban Accessibility, European Union
EuroHOPE	European Health Care Outcomes, Performance and Efficiency Project
Fing	Fondation Internet Nouvelle Génération, France
fMRI	Functional magnetic resonance imaging
GfK	Gesellschaft für Konsumforschung, Society for Consumer Research
GIS	Geographic information systems
GP	General practitioner
GPHIN	Global Public Health Information Network
GPS	Global positioning system
HCQI	Health Care Quality Indicators (OECD)
HDDs	Hard disk drives
HES	Hospital Episode Statistics
HGF	Hypothesis generation framework
HMO	Health care maintenance organisation
IaaS	Infrastructure as a service
IAPP	International Association of Privacy Professionals
ICC	Integrated circuit card
ICES	Institute for Clinical and Evaluative Sciences, Canada
ICGC	International Cancer Genome Consortium
ICS-CERT	Industrial Control System Cyber Emergency Response Team, United States
ICSTI	International Council for Scientific and Technical Information
ICSU	International Council for Science
ICTs	Information and communication technologies
IGOs	International governmental organisations
IEA	International Energy Agency
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
INTEGRATE	Integrative Cancer Research through Innovative Biomedical Infrastructures, European Commission
IoT	Internet of Things
IPRs	Intellectual property rights
ISO	International Organization for Standardization
ISPs	Internet service providers
ITU	International Telecommunication Union
JSON	JavaScript Object Notation
Kbit	Kilobit, equals 1 000 bits
M2M	Machine-to-machine (communication)

Mbit	Megabit, equals 1 000 000 bits
MGI	McKinsey Global Institute
MOOCs	Massive open online courses
MR	Magnetic resonance
NCDs	Non-communicable diseases
NDES	National Digital Economy Strategy
NFC	Near field communication
NHGRI	National Human Genome Research Institute
NIT	Negative income tax
NLP	Natural language processing
NPISHs	Non-profit institutions serving households
NSF	National Science Foundation, United States
OCR	Optical character recognition
ODbL	Open Database License
ODI	Open Data Institute, United Kingdom
OLAP	Online analytical processing
OLTP	Online transaction processing
OSS	Open source software
OSTP	Office of Science and Technology Policy, United States
PaaS	Platform as a service
PAW Conference	Predictive Analytics World Conference
PB	Petabytes
PCOR	Patient-centred health outcomes research
PCT	Patent Cooperation Treaty
PCTs	Primary care trusts
PEDW	Patient Episode Database for Wales
PET	Positron emission tomography
PGETIC	Global Strategic Plan for Rationalisation of ICT Costs in Public Administration, Portugal
PII	Personal identifying information
PNAS	Proceedings of the National Academy of Sciences
ProMED	Program for Monitoring Emerging Diseases
PSI	Public sector information
QIN	Quantitative Imaging Network, United States
QoS	Quality of service
RDA	Research Data Alliance
RDF	Resource Description Framework
RFID	Radio frequency identification
RPAS	Remote piloted aircraft systems
SaaS	Software as a service
SALUS	Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies
SCOAP3	Sponsoring Consortium for Open Access Publishing in Particle Physics
Sense-OS	Sense Observation Systems
SGDR	Sui generis database right
SHAs	Strategic health authorities
SIM	Subscriber identity module
SIS	Swedish Standardization Institute
SNA	UN System of National Accounts

SPARC	Scholarly Publishing and Academic Resources Coalition
SPECT	Single photon emission computed tomography
SQL	Structured query language
SSDs	Solid-state drives
STRIDE	Stanford Translational Research Integrated Database Environment
TCGA	The Cancer Genome Atlas
TfL	Transport for London
TNO	Netherlands Organisation for Applied Scientific Research
TRANSFoRm	Translational Research and Patient Safety in Europe
TRIPS	Trade-Related Aspects of Intellectual Property Rights
UAV	Unmanned aerial vehicles
UKDBIS	UK Department for Business Innovation & Skills
URIs	Uniform resource identifiers
USPTO	United States Patent and Trademark Office
VRM	Vendor Relationship Management
WCT	WIPO Copyright Treaty
WIPO	World Intellectual Property Organization
WITSA	World Information Technology and Services Alliance
WPA	Wireless personal area
WT	Wellcome Trust
XML	eXtensible Markup Language

Executive summary

Close to real-time analysis of large volumes of data (big data) – generated from a myriad of transactions, production and communication processes – is accelerating knowledge and value creation across society to unforeseen levels. Data-driven innovation (DDI) refers to significant improvement of existing, or the development of new, products, processes, organisational methods and markets emerging from this phenomenon.

DDI has the potential to enhance resource efficiency and productivity, economic competitiveness, and social well-being as it begins to transform all sectors in the economy, including low-tech industries and manufacturing. The exploitation of DDI has already created significant value-added for many businesses and individuals, and more can be expected to follow. Some estimates put the global market for big data related technology and services at USD 17 billion in 2015, with a growth rate of 40% on average every year since 2010. Available evidence also shows that firms using DDI have raised productivity faster than non-users by around 5-10%.

DDI can also help address social and global challenges, including climate change and natural disasters, health and ageing populations, water, food, energy security, and mass urbanisation. Investments in public administration, research and education, and health care will be particularly fruitful in the short term, as these areas rely heavily on the collection and analysis of information, but still face a relatively low level of computerisation in most countries.

The disruptive nature of DDI requires addressing major economic and societal challenges and calls for a whole-of-government and participatory approach to help maximise the benefits and mitigate associated risks and obstacles.

Two clusters of challenges should be met by policy makers in the transition towards a data-driven economy:

1. Governments should consider addressing the negative effects of "creative destruction" while stimulating investments in:
   - *the infrastructure needed for DDI,* particularly in mobile broadband, cloud computing, the Internet of Things, and data, with a strong focus on small and medium-sized enterprises (SMEs) and high value-added services
   - *the public sector, health care, science and education* to pick the "low-hanging fruit" that can boost efficiency, knowledge sharing and well-being in the short term, and help better address global challenges
   - *organisational change and entrepreneurship in the private and public sector* by encouraging a culture of data-driven experimentation and learning
   - *continuous education training and skills development beyond science, technology, engineering and mathematics (STEM) fields* to take advantage of job creation opportunities and smooth structural change while addressing inequality in earnings in labour markets.

2. Governments should aim to understand and strike the right balance between the social benefits of "openness" and individuals' and organisations' legitimate concerns of such openness by encouraging:

   - *the free flow of data across nations and organisations*. This also includes ensuring that the Internet remains an open platform for innovation; promoting both open access to data and interoperability of data-driven services; and empowering actors to reuse their data across interoperable applications (i.e. data portability).
   - *the responsible usage of personal data and the prevention of harm caused by privacy violations.* This also includes enhancing the participation of individuals; the transparency of data processing; the effectiveness of privacy enforcement; and the adoption of a privacy risk management approach.
   - *a culture of digital risk management across society*, involving all stakeholders of the data ecosystem.
   - *data sharing and the appropriation of returns on investments (ROI)* through a combination of alternative incentive mechanisms such as data citations and intellectual property rights (IPR) licences that enable sharing such as Creative Commons and open source software licences.
   - *coherent assessment of market concentration and competition barriers* through better definitions of the relevant market and the consideration of potential consumer detriments due to privacy violation. This will also require a better dialogue between regulatory authorities (in particular in the area of competition, privacy and consumer protection).
   - *improved measurement* to help better assess the economic value of data assets, prevent base erosion and profit shifting (BEPS), and design better DDI policies.

In addressing these two clusters of challenges, policy makers should acknowledge that DDI may favour concentration and greater information asymmetry and with that, shifts in power: away from individuals to organisations; from traditional businesses to data-driven businesses; and from governments to data-driven businesses (the latter can gain more knowledge about citizens than governments). These shifts could exacerbate existing inequalities and lead to a new digital (data) divide that could undermine social cohesion and economic resilience if not addressed.

Given all of this, governments have an important role to play in promoting DDI and mitigating the associated risks.

Chapter 1

The phenomenon of data-driven innovation

This chapter provides a synthesis of the main findings of Phase II of the OECD project on New Sources of Growth: Knowledge-Based Capital*, in particular its pillar which focuses on data-driven innovation (KBC2: DATA). It first presents available evidence on the increasing role of "big data" and data analytics, highlighting in particular the potential of data-driven innovation (DDI) for economic growth, development, and well-being. It then presents the context and policy issues related to the various aspects of DDI covered in this book, chapter by chapter. The discussion concludes by raising key challenges that most countries will face as DDI takes off and accelerates, and the policy considerations they will need to address.*

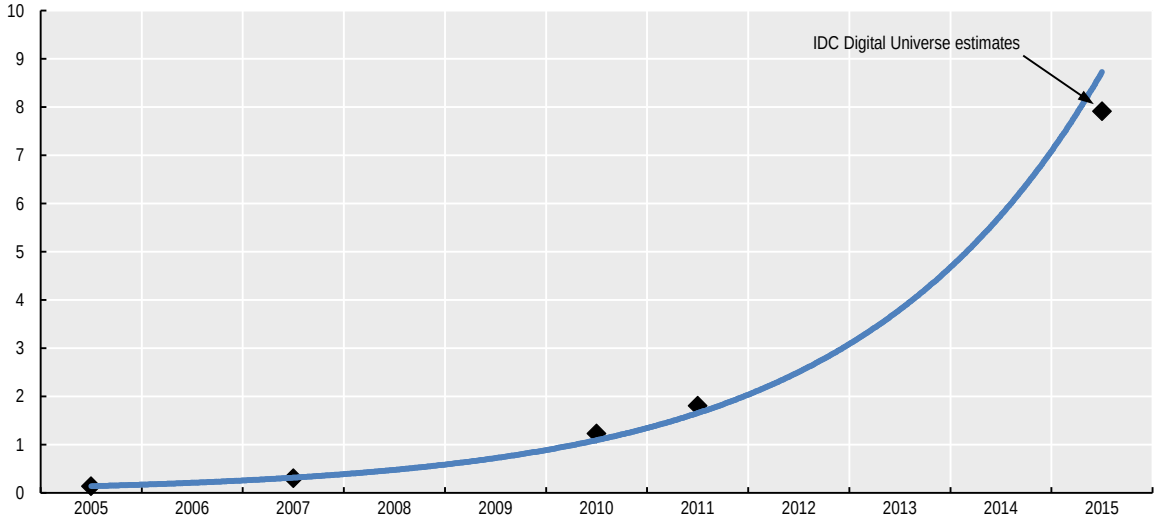
It's difficult to imagine the power that you're going to have when so many different sorts of data are available. (Berners-Lee, 2007)

Software is eating the world (Marc Andreessen, in Anderson, 2012)
… and the world is served in big chunks of data. (Esmeijer, Bakker, and de Munck, 2013)

More and more organisations are starting to leverage large volumes of (digital) data generated from myriad transactions and production and communication processes. These large streams of data, which are now commonly referred to as "big data", are generated through information and communication technologies (ICTs) including the Internet, as well as ubiquitous, wired sensors that are capturing activities in the physical world (see Chapter 3 of this volume). Measurement of the real total data generated, collected and stored is still speculative, but some sources suggest, for instance, that today more than 2.5 exabytes[1] (EB, a billion gigabytes) of data are generated every single day,[2] which is the equivalent of 167 000 times the information contained in all the books in the Library of Congress of the United States. This has led to an estimated cumulative data storage of around 8 zettabytes (ZB, a trillion gigabytes) in 2015 (Figure 1.1) and some estimates suggest that this will multiply by a factor of 40 by the end of this decade.[3] Today, the world's largest retail company, Walmart, already handles more than 1 million customer transactions every hour, which are imported into databases estimated to have contained more than 2.5 petabytes (PB, a million gigabytes) of data in 2010 (*The Economist*, 2010a).

Figure 1.1. **Estimated worldwide data storage**

In zettabytes (ZB, trillions of gigabytes)



Source: Based on the IDC (2012) Digital Universe research project.

The analysis of "big data", increasingly in real time, is driving knowledge and value creation across society; fostering new products, processes and markets; spurring entirely new business models; transforming most if not all sectors in OECD countries and partner economies; and thereby enhancing economic competitiveness and productivity growth. Algorithmic trading systems (ATS), for example, analyse massive amounts of market data on a millisecond basis to autonomously identify what to stock and when, and at what price to trade; this process was unheard of a decade ago (see Chapter 3). Traditional sectors such as manufacturing and agriculture are also being disrupted through the use of data and analytics, and are becoming more and more service-like (see Chapter 2). The German manufacturer of athletic shoes and sports equipment, Adidas, for instance, has redesigned many of its products as *data-driven services*, which are integrated via its online *miCoach* platform. This platform enables services related to physical activities

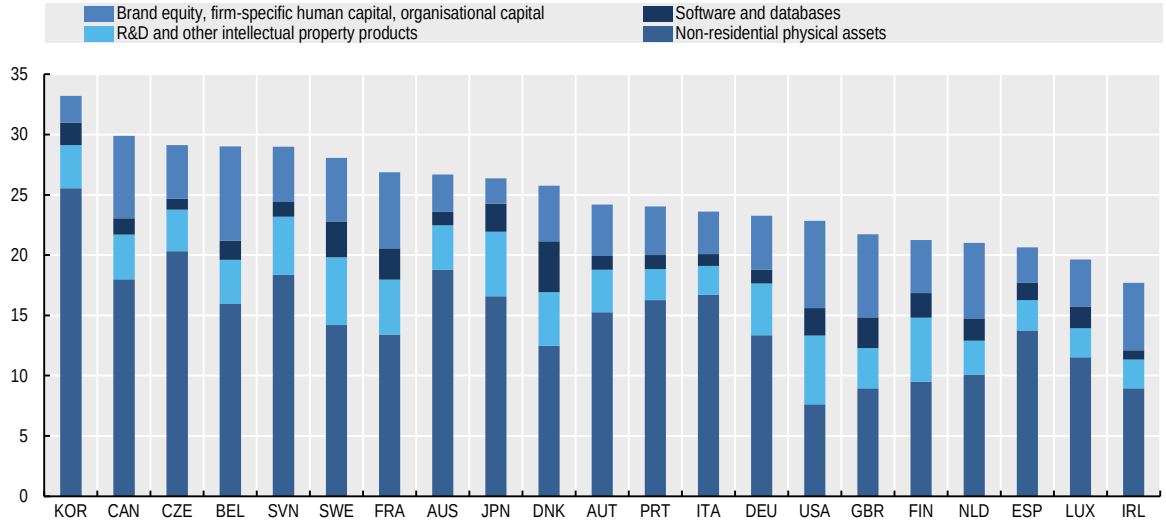
such as performance monitoring and training recommendation. In the public sector, the release of data as "open government data" can increase the transparency and accountability of governments, thus boosting public sector efficiency and public trust in governments (see Chapter 10). Better access to public sector information (PSI, including public sector data) can also empower entrepreneurs to develop new innovative commercial and social goods and services – such as the app "Asthmapolis", which is based on data released by the United States Government, and used to identify highly dangerous spots for asthmatic people. Since the app was created, hospitals in the United States have recorded a 25% decrease in asthmatic incidents.

The use of data and analytics to improve or foster new products, processes, organisational methods and markets – which is referred to hereafter as "data-driven innovation" (DDI) – is a new source of growth. It also represents a key opportunity for governments aiming to rebuild public trust through greater openness, transparency and accountability of the public sector. But governments need to address some major economic and societal challenges and risks in order to unleash the full potential of DDI and assure that its fruits contribute to the well-being of all citizens. These include, most prominently, the risk of i) barriers to the free flow of data (see Chapters 2, 3 and 4), ii) market concentration and competition barriers (Chapter 2), iii) base erosion and profit shifting (BEPS, Chapter 2), iv) privacy violation and discrimination (Chapter 5), v) dislocation effects in labour markets (Chapter 6), and with that vi) an emerging new digital or "data divide" that may hit developing economies particularly hard. Some of these challenges and risks deserve special attention from governments particularly concerned about social cohesion and rising inequality, which could hamper the economic resilience of their countries as stated by Ministers and Representatives[4] in the OECD (2014a) Ministerial Council Statement.[5]

DDI should be seen in a broader social and economic context in which knowledge-based capital (KBC) increasingly forms the foundation of 21st century knowledge economies, with data and software as one key pillar. In 2010, the OECD launched a horizontal project on *New Sources of Growth: Knowledge-Based Capital*, which provides evidence of the impact on growth, and the associated policy implications, of three main types of knowledge-based capital (KBC): i) computerised information (e.g. software and databases); ii) innovative property (e.g. patents, copyrights, designs and trademarks); and iii) economic competencies (e.g. brand equity, firm-specific human capital, networks of people and institutions, and organisational know-how) (OECD, 2013a).[6] The work highlighted that in some countries – such as Sweden, the United Kingdom and the United States – investment in KBC matches or exceeds investment in physical capital such as machinery, equipment and buildings (Figure 1.2). In many countries, such as Denmark, Ireland and Italy, business investment in KBC also rose higher as a share of GDP, or declined less, than investment in physical capital during the 2008-09 financial and economic crisis (OECD, 2013a).

Figure 1.2. **Investment in physical and knowledge-based capital, 2010**

As a percentage of value added of the business sector



Sources: *OECD Science, Technology and Industry Scoreboard 2013*, based on INTAN-Invest Database, www.intan-invest.net, and national estimates by researchers. Estimates of physical investment are based on OECD Annual System of National Accounts (SNA) and the INTAN-Invest Database, May 2013, http://dx.doi.org/10.1787/888932889820.

This synthesis chapter is structured as follows. It first presents available evidence on the increasing roles of data, analytics, and data-driven innovation. It then illustrates the context and policy issues related to the various aspects of DDI that are covered in this book, chapter by chapter. The discussion concludes by raising key challenges that most countries will face as DDI takes off and accelerates, and the policy considerations they will need to address.

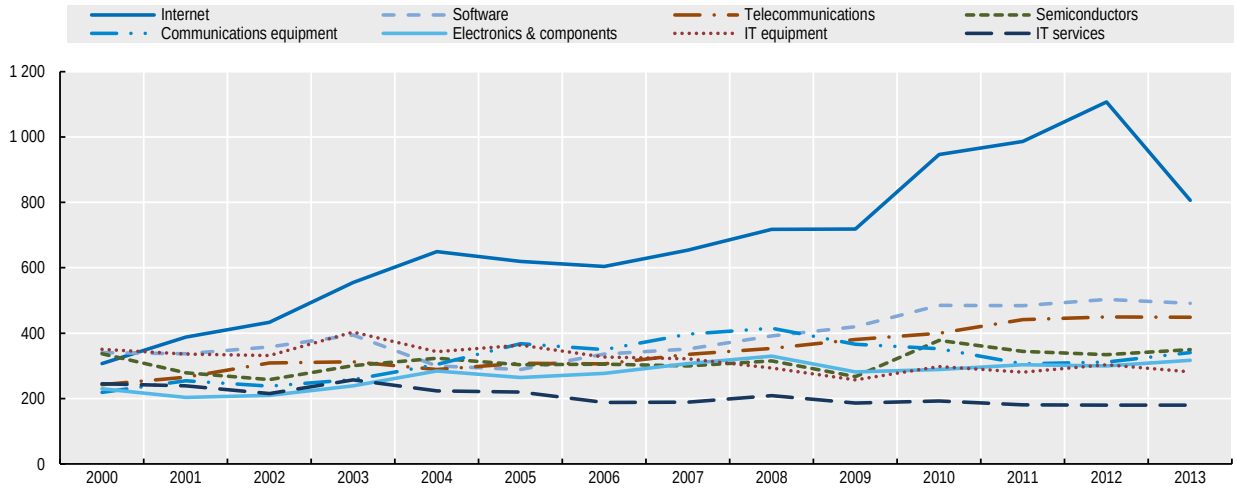
1.1. The rise of "big data" and data-driven innovation

Leading the way: The ICT sector

ICT firms heavily rely on KBC investments, in particular software and data. This is especially apparent in the asset structure of Internet firms, such as Google and Facebook, where physical assets accounted for only around 15% of the firms' worth as of 31 December 2013.[7] Internet firms also enjoy huge productivity gains thanks to their KBC investments in software and data particularly. However, compared with other ICT firms, which also rely heavily on investments in software and data, Internet firms are by far more productive. Among the OECD area's top 250 ICT firms, Internet firms generated on average more than USD 1 million in revenues per employee in 2012 and more than USD 800 000 in 2013, while the other top ICT firms generated around USD 200 000 (IT services firms) to USD 500 000 (software firms) (Figure 1.3).

Figure 1.3. **Average revenue per employee of top 250 ICT firms, 2000-13**

In thousand USD



Note: The presentation is based on averages for those firms reporting in 2000-13.

Sources: Based on OECD Information Technology database; compiled from annual reports, SEC filings and market financials, July 2014.

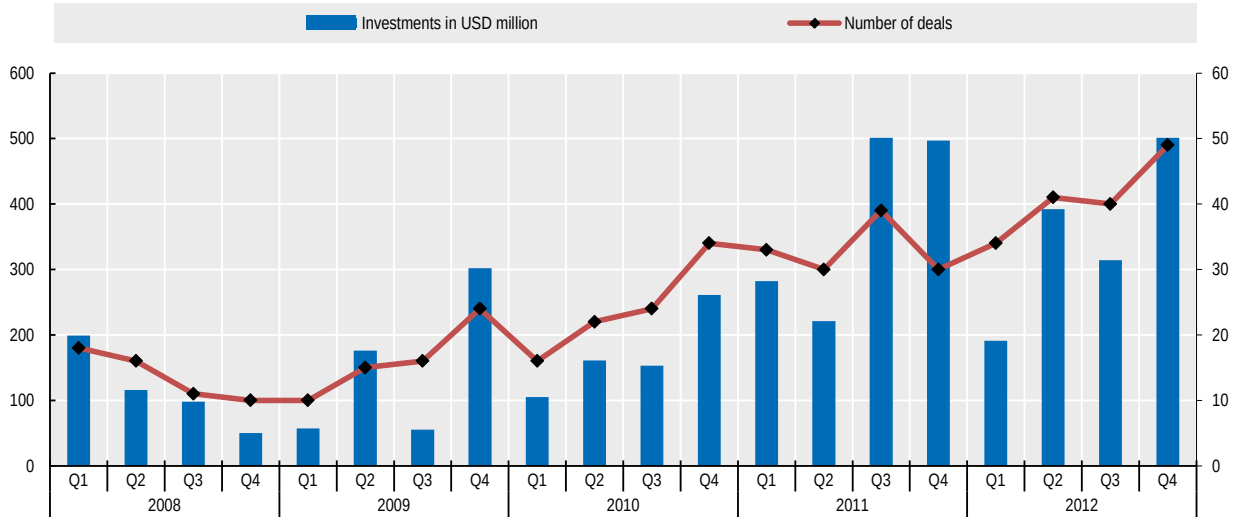
The business models of many Internet firms involve the collection and analysis of large streams of data collected from the Internet (OECD, 2012). By collecting and analysing "big data", a large share of which is provided by Internet users (consumers), Internet companies are able to automate their processes and to experiment with, and foster, new products and business models at much a faster rate than the rest of the industry. In particular, the advanced use of data and analytics enables Internet firms to scale their businesses at much lower costs than other ICT firms, a phenomenon that goes much further than what Brynjolfsson et al. (2008) describe as *scaling without mass*.[8]

The rest of the ICT sector (excluding Internet firms) has begun to recognise big data as a new business opportunity and is making significant investments to catch up and jump on the big data bandwagon. Estimates by IDC (2012) suggest that "big data technology and services" will grow from USD 3 billion in 2010 to USD 17 billion in 2015, which represents a compound annual growth rate (CAGR) of almost 40%. Technologies and services related to storage are expected to be the fastest growing segment, followed by networking and services, which explains the increasing role of IT equipment firms in this relatively new market.[9]

Top ICT companies are also strengthening their position through mergers and acquisitions (M&A) and/or through "co-opetition" (i.e. collaboration with potential and actual competitors). This includes in particular the acquisition of young start-ups specialised in big data technologies and services, and co-opetition via open source projects such as Hadoop (see Chapter 2). Data provided by Orrick (2012) on M&A deals (mainly in the United States) show that M&A activities have increased significantly since 2008 in terms of volume and number of deals: from 55 deals in 2008 to 164 in 2012, with almost USD 5 billion being invested over that period (Figure 1.4). In the first half of 2013 alone, big data companies raised almost USD 1.25 billion across 127 deals. IBM was the most active acquirer of big data companies in 2012, followed by Oracle.

Figure 1.4. **Big data-related financing activities, Q1 2008-Q4 2012**

Volume of investments in USD million (left scale) and number of deals (right scale)

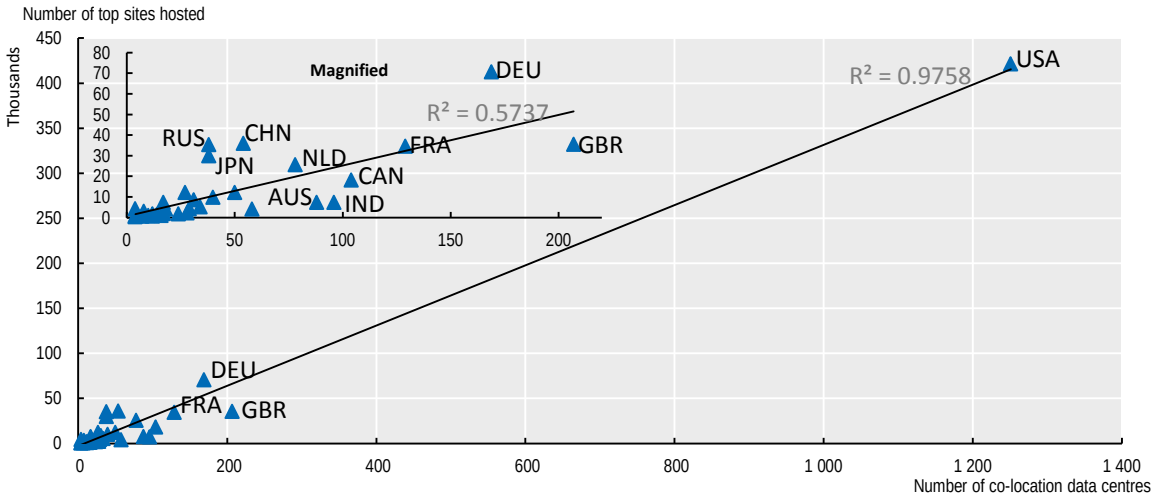


Source: Based on Orrick, 2012.

The combined effect of M&A, co-opetition, and the demand and supply of goods and services related to big data is the emergence of a *global data ecosystem* in which data and analytic services are traded and used across sectors and national borders (see Chapter 2 of this volume). The United States plays a central role, and countries such as Canada, Germany, France, Ireland, the Netherlands, Japan and the United Kingdom, as well as the People's Republic of China (hereafter "China"), India and Russia are catching up. The global data ecosystem involves global value chains (GVCs), in which companies increasingly divide up their data-related processes and locate productive activities in many countries. Figures on the distribution of data-driven services are not known. However, analysis of the world's top Internet sites suggests that data-driven services may be concentrated in the United States, which alone accounted for more than 50% of all top sites hosted in the OECD area, plus Brazil, China, Colombia, Egypt, India, Indonesia, Russia and South Africa in 2013 (Chapter 3). The number of top sites hosted correlates significantly with the number of co-location data centres (Figure 1.5).

Furthermore, top locations for data-driven services tend to be major exporters of ICT services (see Chapter 2). In 2013, the top ten exporters of ICT services were India, Ireland, the United States, Germany, the United Kingdom, China, France, the Netherlands, Belgium, and Sweden, and all – except Ireland, Belgium, and Sweden – are top locations for data-driven services. These countries are more likely to be the largest destination of cross-border data flows. As a consequence, the leading OECD area importers of ICT-related services are also the major sources for trade-related data; they include in particular the United States and Germany. (See Chapter 2 for further discussion on trade in data and ICT-related services.)

Figure 1.5. **Top locations by number of co-location data centres and top sites hosted**



Note: Number of top sites hosted based on analysis of 429 000 country code top-level domains (ccTLD) of the top one million sites collected in 2013. The remaining sites including the generic top-level domains were omitted from the list, as there are no reliable public data as to where the domains are registered.

Sources: Based on Pingdom, 2013; and www.datacentermap.com, accessed 27 May 2014.

Data-driven innovation across society

The economic impact of DDI goes far beyond market prospects for the ICT industry, although evidence strongly suggests that ICT firms are not only supplying products for data collection, processing and analysis, but also still leading in use of advanced data analytics. According to Tambe (2014), for example, only 30% of Hadoop investments come from non-ICT sectors, including in particular finance, transportation, utilities, retail, health care, pharmaceuticals and biotechnology firms. There is, however, a rapidly growing interest from non-ICT businesses across the economy in big data-related technologies and services to exploit data for innovation – that is to say, for developing new, or for improving existing, products, processes and markets (see Box 1.1 for the OECD definition of innovation).

Many organisations across the economy already benefit from significant investment in data in the form of traditional databases[10] for innovation, in particular in Finland, Denmark, Luxembourg, Sweden, the United Kingdom and the United States. As shown in Figure 1.2, investments in software and data (across the economy) accounted for an average share of slightly below 2% of business sector value added in OECD countries, with businesses in countries such as Denmark (4%), Sweden (3%), the United Kingdom and the United States (both 2%) leading in terms of the share of investment. With the exception of Sweden, these latter countries also saw a significant increase in software and data-related investments during the crisis, as did countries such as Luxembourg and Finland. Overall, investments in software have increased to 57% of total ICT investment in 2012, from less than 40% in 2000 (OECD, 2015a).

Box 1.1. **Defining innovation**

The latest (3rd) edition of the *Oslo Manual* defines innovation as the implementation of a new or significantly improved product (good or service), or process, new marketing method, or new organisational method in business practices, workplace organisation or external relations (OECD and Eurostat, 2005). This definition, for measurement purposes, captures the following four types of innovation:

- *Product innovation* – The introduction of a good or service that is new or significantly improved with respect to its characteristics or intended uses. This includes significant improvements in technical specifications, components and materials, incorporated software, user-friendliness and other functional characteristics.
- *Process innovation* – The implementation of a new or significantly improved production or delivery method. This includes significant changes in techniques, equipment and/or software.
- *Marketing innovation* – The implementation of a new marketing method involving significant changes in product design or packaging, product placement, product promotion or pricing.
- *Organisational innovation* – The implementation of a new organisational method in the firm's business practices, workplace organisation or external relations.

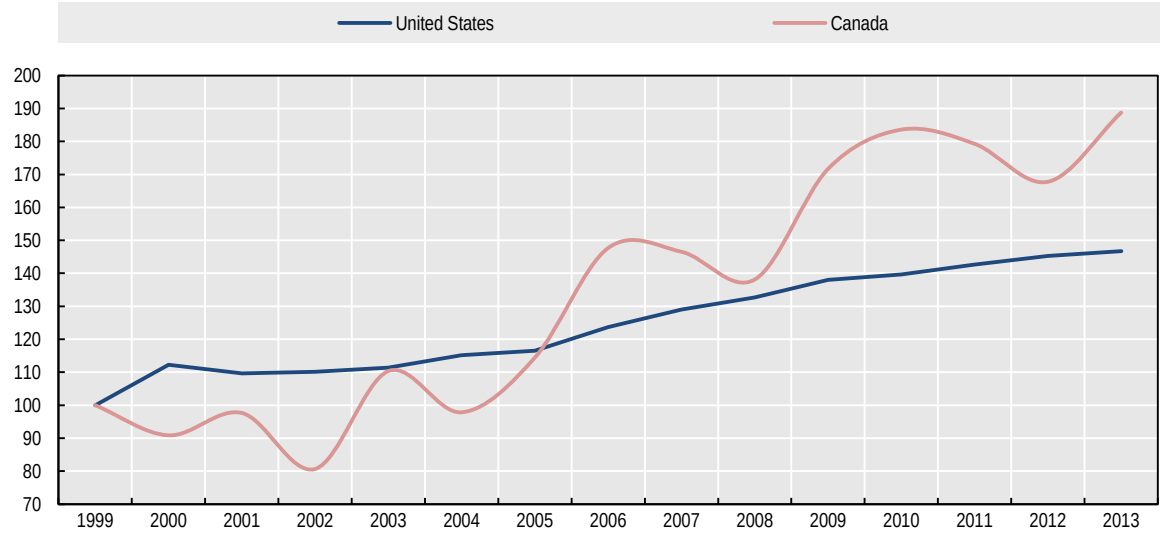
Source: OECD and Eurostat, 2005.

Increasing investments in software and databases go hand in hand with a growing data intensity of the economy as measured, for instance, by the share of data specialists in total employment. Employment figures for Canada and the United States show that the share of data specialists in total employment has increased since 1999 (Figure 1.6). The most data intensive industries employing the highest share of data specialists are still the ICT services industries, and in particular i) the IT and other information services industries, but also ii) insurance and finance, iii) science and research and development, iv) advertising and market research, as well as v) the public sector (see Chapter 6 of this volume). This is in line with findings by Tambe (2014) presented earlier and estimates by MGI (2011), according to which data intensity (measured as the average volume of data stored per organisation) is highest in financial services (including securities and investment services and banking), communication and media, utilities, government, and manufacturing. In these sectors, each organisation stored on average more than one petabyte (one million gigabytes) of data in 2009.

The following three sections describe the potential of DDI to contribute to productivity growth, well-being, inclusiveness and development. The process through which DDI creates value to achieve these policy objectives – the *data value cycle* – is presented in detail afterward.

Figure 1.6. **Trends in data intensity of the Canadian and United States economies, 1999-2013**

Index, share of data specialists[1] in total employment, 1999=100



Note: Data specialists do not correspond here to the 2008 International Standard Classification of Occupations (ISCO-08) definition presented in Box 6.4. in Chapter 6 of this volume. To be consistent across years, the definition has been slightly modified and does not include "Information security analysts" (SOC 2010 code 15-1122), "computer network architects" (15-1143) or "Computer occupations, nec" (15-1199) for the United States, and only include ISCO 08 code 212, "mathematicians, actuaries and statisticians", and (2521), "database designers and administrators" for Canada.

Sources: Occupational Employment Statistics (OES), US Bureau of Labor Statistics, www.bls.gov/oes/home.htm, November 2014; Statistics Canada, labour force survey, February 2015.

DDI can boost productivity growth

DDI is a disruptive new source of growth that could transform all sectors in the economy. Even traditional sectors such as retail, manufacturing and agriculture are being disrupted through DDI, as companies become more and more service-like, a trend that some have described using the term "servicification" (Lodefalk, 2010). Firms like Tesco, the UK supermarket chain, exploit huge data flows generated through their fidelity card programmes. The Tesco programme now counts more than 100 market baskets a second and 6 million transactions a day, and it very effectively transformed Tesco from a local, downmarket "pile 'em high, sell 'em cheap" retailer to a multinational, customer- and service-oriented one with broad appeal across social groups.

The world's largest company, Walmart, is even more progressive in its use of data and analytics. The company develops its own data analytic services via its subsidiary Walmart Labs, which is also actively contributing to the (co-)development of open source analytics. Walmart Labs' (internal) solution *Social Genome*, for example, allows Walmart to reach out to potential customers, including friends of direct customers, who have mentioned specific products online, to provide discounts on these exact products.[11] "This has resulted in a vast, constantly changing, up-to-date knowledge base with hundreds of millions of entities and relationships" (Big Data Startups, 2013).

In manufacturing, companies are increasingly using sensors mounted on production machines and delivered products to collect and process data on the machines' and products' operation, taking advantage of the Internet of Things (IoT) – the interconnection of "real world" objects. This trend, enabled by machine-to-machine

communication (M2M) and analysis of sensor data, has been described by some as "Industry 4.0" (Jasperneite, 2012), the "Industrial Internet" (Bruner, 2013), and "network manufacturing" (Economist Intelligence Unit, 2014). Sensor data are used to monitor and analyse the efficiency of products, to optimise their operations at a system-wide level, and for after-sale services, including preventive maintenance operations. The data are sometimes also used in collaboration with suppliers, and in some cases even commercialised as part of new services for existing and potential suppliers and customers.[12] For example, Germany-based Schmitz Cargobull, the world's largest truck body and trailer manufacturer, uses M2M and sensors to monitor the maintenance, travelling conditions and routes travelled by any of its trailers (Chick, Netessine and Huchzermeier, 2014; see also Vennewald, 2013). The insights generated by analysis of the data are used to help Schmitz Cargobull's customers minimise their usage breakdowns.[13] Quantitative evidence on the overall economic impact of DDI in manufacturing is still limited. Available estimates for Japan, for example, suggest that the use of big data and analytics by some divisions of Japanese manufacturing companies could lead to savings in maintenance costs worth almost JPY 5 trillion (which correspond to more than 15% of sales in 2010) and more than JPY 50 billion in electricity savings (MIC, 2013). For Germany it is estimated that Industry 4.0 can enable companies to boost their productivity by up to 30% (acatech, 2013), and to increase gross value added by a cumulative amount of up to EUR 267 billion by 2025 (BITKOM and Fraunhofer, 2014).

Agriculture is now being further modernised thanks to DDI, a development that is leading to huge productivity improvements and the reduction of environmental impacts. DDI in agriculture builds on geo-coded maps of agricultural fields and the real-time monitoring of every activity, from seeding to watering and fertilising, to harvesting. The data that are thereby generated can now be stored and analysed using cloud computing. As a result, farmers are today sitting on a wealth of agricultural data, which companies such as Monsanto, John Deere and DuPont Pioneer are trying to exploit through new data-driven goods and services (Noyes, 2014). John Deere, for example, is taking advantage of the "Industrial Internet" by integrating sensors into its latest equipment "to help farmers manage their fleet and to decrease downtime of their tractors as well as save on fuel" (Big Data Startups, 2013). The same sensor data are reused and linked with historical and real-time data on (e.g.) weather patterns, soil conditions, fertiliser usage and crop features, to optimise and predict agricultural production.[14] Traditional cultivation methods are thus improved and the wisdom and know-how of skilled farmers formalised. Overall, the use of data and analytics is estimated by some experts to improve yields by five to ten bushels per acre or around USD 100 per acre in increased profit (Noyes, 2014). This productivity increase comes at the right time, as the OECD and the Food and Agriculture Organization of the United Nations (OECD and FAO, 2012) call for a necessary food production increase by 60% for the world to be able to feed the growing population, which is expected to hit 9 billion in 2050.

There is as yet little evidence on the overall economic effects of DDI, but the few studies available suggest that firms using DDI raise labour productivity faster than non-users. A study of 330 companies in the United States by Brynjolfsson, Hitt and Kim (2011) estimates that the output and productivity of firms that adopt data-driven decision making are 5% to 6% higher than would be expected from their other investments in and use of ICTs. These firms also perform better in terms of asset utilisation, return on equity and market value. A similar study based on 500 firms in the United Kingdom by Bakhshi, Bravo-Biosca and Mateos-Garcia (2014) finds that businesses that make greater use of online customer and consumer data are 8% to 13% more productive as a result.[15] A recent

study by Tambe (2014) based on the analysis of 175 million LinkedIn user profiles, out of which employees with skills for big data-specific technologies have been identified, indicates that firms' investment in big data-specific technologies was associated with 3% faster productivity growth.[16] Overall, these studies suggest an approximately 5-10% faster productivity growth of DDI users compared to that of non-users.[17] However, it should be stressed that these estimates cannot be generalised, for a number of reasons. First, the estimated effects of DDI vary by sector and are subject to complementary factors, such as the availability of skills and competences and the availability and quality (i.e. relevance and timeliness) of the data used (see Chapter 4). But more importantly, these studies often suffer from selection bias, which makes it difficult to disentangle the effects of DDI from other factors at the firm level.[18] More comprehensive studies are therefore needed to better assess the impact of DDI on productivity growth.

DDI can contribute to well-being

The full impact of DDI goes beyond its positive effects on productivity growth. DDI can also contribute directly to the well-being of citizens, even if quantification of that contribution remains challenging because many if not most of the benefits related to the use of data are still not captured by market transactions (Mandel, 2012, 2013).[19] Citizens' use of open data as enabled by governments through their open data initiatives, for example, can increase the openness, transparency and accountability of government activities and thus boost public trust in governments. At the same time, it can enable an unlimited range of commercial and social services across society. For instance, "civic entrepreneurs" increasingly use available open data as promoted by the OECD (2008) *Council Recommendation on Enhanced Access and More Effective Use of Public Sector Information* (PSI), in combination with other publicly available data sources, to develop apps that facilitate access to existing public services. Estimates on the economic impact of PSI (EUR 509 billion in 2008 for the reuse of PSI in the OECD area) focus on the commercial reuse of PSI and thus do not cover the full range of (social) benefits.

Science and education, health care services and public administration are the low hanging fruit policy makers can target in the relative short run to leverage DDI for growth and well-being. These sectors may be where adoption of DDI could have the highest impact. They employ the largest share of people who perform work related to the collection, processing and analysis of information and data. However, in these sectors, people are also still performing that work at a relatively low level of computerisation. In the United States, where data on working activities are available via the United States Department of Labor's O*NET system, almost 30% of the total employment in health care and social assistance, for instance, is in occupations largely involving information collection and analysis (e.g. records of patient medical histories, and test data or image analysis to inform diagnosis or treatment), but at the same time also involving a relatively low level of computer interaction.[20] Targeted promotion of the adoption of ICTs, and data and analytics in particular, could thus boost efficiency gains even further in these sectors.

In the area of science, the advent of new instruments and methods of data-intensive exploration could signal the arrival of new "data-intensive scientific discoveries", with new opportunities for knowledge creation. New instruments such as super colliders or telescopes, but also the Internet as a data collection tool, have been instrumental in these new developments in science, as they have changed the scale and granularity of the data being collected (see Chapter 7 of this volume). The Digital Sky Survey, for example, launched in 2000, collected more data through its telescope in its first week than had been amassed in the history of astronomy (The Economist, 2010a), and the new square

kilometre array (SKA) radio telescope could generate up to 1 petabyte (one million gigabyte) of data every 20 seconds (EC, 2010). Furthermore, the increasing power of data analytics has made it possible to extract insights from these very large data sets reasonably quickly. In genetics, for instance, DNA gene sequencing machines based on big data analytics can now read about 26 billion characters of the human genetic code in seconds. This goes hand in hand with the considerable fall in the cost of DNA sequencing over the past five years (see Chapter 3).

These recent developments in science obviously had significant impacts on health research and care, where the demographic evolution toward ageing societies and rising health costs are pressing for greater efficiency and for more responsive, patient-centric services (Chapter 8). At the core of DDI in the health sector are national health data, including but not limited to electronic health records and genetic, neuroimaging and epidemiological data. The efficient reuse of these data sets promises to improve the efficiency and quality of health care. In Finland for example, the content, quality and cost-effectiveness of treatment of a set of selected diseases are analysed by linking patient data across the whole cycle of care from admission to hospital, to care by their community doctor, to the medications prescribed and deaths (OECD, 2013c). The results of the analysis are made publicly available and have empowered patients and led to improvement in the quality of hospitals in Finland. In the particular case of the US health care system, MGI (2011) estimates that the use of data analytics throughout the system (clinical operations, payment and pricing of services, and R&D) could bring savings of more than USD 300 billion, two-thirds of which would come from reducing health care expenditures by 8%.[21]

New sources of data are already being considered by researchers who are seeking to improve research in and the treatment of diseases, as well as by individuals who are taking advantage of DDI to empower themselves for better prevention and care. For example, the social network PatientsLikeMe not only allows people with a medical condition to interact with and derive comfort and learn from other people with the same condition, but also provides an evidence base of personal data for analysis and a platform for linking patients with clinical trials. As another example, the so-called Quantified Self-movement has inspired its followers to use tools, like Fitbit, to track their every move and heartbeat, and to empower individuals to improve their health and overall well-being.

In the case of the public sector (intelligence and security excluded), there is some evidence of insufficient use of data that are generated and collected (see Chapter 10). According to MGI (2011), full use of data analytics in Europe's 23 largest governments may reduce administrative costs by 15% to 20%, creating the equivalent of EUR 150 billion to EUR 300 billion in new value and accelerating annual productivity growth by 0.5 percentage points over the next ten years.[22] The main benefits would be greater operational efficiency (due to greater transparency), increased tax collection (due to customised services, for example), fewer frauds and errors (due to automated data analytics). Similarly, a study of the United Kingdom shows that the public sector could save GBP 2 billion in fraud detection and generate GBP 4 billion through better performance management by using big data analytics (CEBR, 2012). Furthermore, data and analytics can be used to improve policy making by complementing official statistics (Reimsbach-Kounatze, 2015).

DDI can further inclusiveness and development

The potential of DDI to promote growth and contribute to well-being could provide a new opportunity to address the urgent needs of developing economies (Gerdon and Reimsbach-Kounatze, 2015). Increasingly, a wide range of data sources, including mobile phones, social media and the public sector, are being explored by governments, businesses, researchers and citizens groups and used to foster development (UN Global Pulse, 2012; WEF, 2012). International initiatives have formed that investigate the capabilities of data analytics for development. *Paris21*, the Partnership in Statistics for Development in the 21st century, brings together users and producers of statistics in developing and developed countries to strengthen statistical capacities and promote the use of reliable data (Letouzé and Jütting, 2014). Meanwhile, the United Nations (UN) Global Pulse initiative was launched by the Executive Office of the UN Secretary-General in response to the need for more timely data to track and monitor the impacts of global and local socio-economic crises (UN Global Pulse, 2012). The UN, moreover, announced the need for a data revolution for a future development agenda beyond 2015 to succeed the United Nations Millennium Development Goals. The Harvard Humanitarian Initiative, the MIT Media Lab and the Overseas Development Institute jointly formed the Data-Pop Alliance to work on big data for development to improve decisions and empower people in a way that avoids the pitfalls of a new digital divide, de-humanisation and de-democratisation (Letouzé and Jütting, 2014).

Significant progress has been made with the use of data analytics for crisis prevention and disaster management (see Box 1.2).[23] Thailand, for instance, is monitoring natural disaster-prone areas such as the coastline, rivers and forests with satellite and ground sensors in order to better react in emergency situations. The Kenyan-based non-profit software company Ushahidi created a system to collect real-time data from eyewitnesses of violence in the aftermath of Kenya's disputed 2007 presidential election; the system has since been used to gain a better understanding of complex situations such as the 2010 earthquake in Haiti, the Syrian Conflict beginning in 2011, and the Ebola epidemic in 2014. Recently, UN Global Pulse has focused on identifying and quantifying discussion themes in Twitter data in order to investigate how people cope in crisis situations such as food price crises or economic crises (UN Global Pulse, 2014).

Box 1.2. **Big data for disaster management**

Real-time analysis of a wide range of data generated through social media, mobile devices and physical sensors (e.g. the Internet of Things) provides a new opportunity for addressing complex societal challenges, including in particular crisis prevention and disaster management. A series of documentary films, "Disaster Big Data", produced by Japanese public broadcaster NHK has shown how data analytics can help build a better understanding and improve response to tremendous disasters such as the one caused by the 2011 earthquake and tsunami in Japan.[1] Data and analytics, together with other ICTs, play an important role at every stage, from prediction to incident management to reconstruction. M2M communication, for instance, can enable the collection of data from water-level sensors, mudslide sensors and GPS sensors for a real-time monitoring and alert system. Japan is now introducing just such an advanced disaster management system, combining this information with location data.

1. See *www.nhk.or.jp/datajournalism/about/index_en.html*, accessed 15 May 2015.

DDI for development could provide some countries with the capacity to "leapfrog" in critical development areas such as transport, finance and agriculture. In transport, the use of data and analytics could improve transport systems in mega cities (see Chapter 9). The online platform Tsaboin, for instance, crowdsources[24] traffic data based on passenger information around bus stops in Lagos, Nigeria, where no official traffic feed exists, to enable users to check the traffic information in real time and make "smarter" traffic decisions.[25] In finance, Cignifi is used to develop mobile-based credit scores. This start-up mines cellphone data to assign a credit score to unbanked potential clients.[26] In the field of agriculture, data analytics can improve the work of farmers through information, forecasting and evaluation, particularly on the local level. The International Center for Tropical Agriculture (CIAT) developed a climate-smart, site-specific recommendation engine for Colombian rice farmers, based on meteorological data and seasonal forecasts.

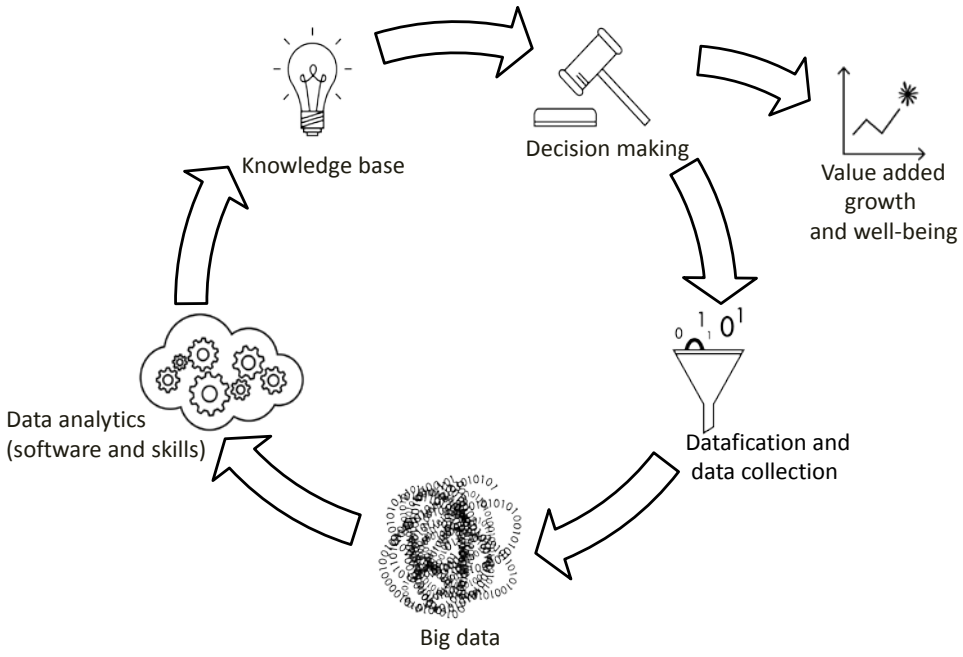
The data value cycle: From datafication to data analytics and decision making

Policy and decision makers aiming at leveraging DDI for growth, well-being and development must understand the process through which data are transformed to finally lead to innovation. In this volume, DDI is described as a sequence of phases from datafication to data analytics and decision making. This process, however, is not a (linear) value chain, but a *value cycle* that involves feedback loops at several phases of the value creation process. The stylised data value cycle illustrated in Figure 1.7 includes the following phases:

- *Datafication and data collection* – These refer to the activity of data generation through the digitisation of content, and monitoring of activities, including real-world (offline) activities and phenomena, through sensors.
- *Big data* – This refers to the result of datafication and data collection that together lead to a large pool of data that can be exploited through data analytics.
- *Data analytics* – Until processed and interpreted via data analytics, big data are typically useless since the first glance reveals no obvious information. Data analytics is increasingly undertaken via cloud computing.
- *The knowledge base* – This refers to the knowledge that is accumulated through learning over time. Where machine learning is involved, the knowledge base reflects the state of the learning system. The knowledge base is the "crown jewels" of data-driven organisations, and therefore enjoys particular protection through legal (e.g. trade secrets – see OECD, 2015b) and technical means (see Chapter 5 on the implications for digital risk management).
- *Data-driven decision making* – The value of data is mainly reaped at two moments: first when data are transformed into knowledge (gaining insights), and then when they are used for decision making (taking action). Decisions taken can in turn lead to more or different data generated and thus trigger a new data value cycle.

Analytics and the value cycle are discussed further below, in the section titled "How data now drive innovation – The focus of Chapter 3".

Figure 1.7. **The data value cycle**



1.2. Objectives and structure of this volume

This volume includes ten chapters discussing the various key aspects of DDI with the aims to: i) improve the evidence base on the role of DDI for promoting growth and well-being, and ii) provide policy guidance on how to maximise the benefits of DDI, and mitigate the associated economic and societal risks. The insights it presents are intended to assist policy makers in better understanding DDI and in incorporating its multidimensionality into policy design. This will, according to the OECD (2014a) Ministerial Council Statement, "help identify trade-offs, complementarities[27] and unintended consequences of policy choices", in line with the common goal of building and maintaining "resilient economies and inclusive societies". These insights can also feed a wide range of future OECD work, including the preparations for the 2016 meeting at Ministerial level organised by the OECD Committee on Digital Economy Policy (CDEP) as well as the current revision of major OECD instruments related to data access, linkage and reuse. These include the OECD (2008) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* of 30 April 2008 and the OECD (2006) *Recommendation of the Council concerning Access to Research Data from Public Funding* of 14 December 2006.

The remainder of this section introduces the context and policy issues related to the various aspects of DDI that the reader will encounter in this volume.

Mapping the global data ecosystem – The focus of Chapter 2

For many of the steps in the value creation process along the data value cycle presented above (Figure 1.7), organisations will have to involve third parties around the world, because they lack the experience, technological resources and/or talent to deal with the multidisciplinary aspects of data and analytics on their own. The resulting global value chain (GVC) is in most cases specifically tailored towards the goal that is being

pursued. What emerges from all this interaction is a global data ecosystem in which, more than ever before, data and analytic services are traded and used across sectors and national borders. The concept of an ecological approach to describe business environments was introduced by Moore (1993) to describe how companies should not be viewed as members of a single industry "[…] but as part of a business ecosystem that crosses a variety of industries." In these ecosystems, collaborative arrangements of firms combine their individual offerings to create coherent, customer-tailored solutions (Adner, 2006).

Key actors in the ecosystem

The global data ecosystem is evolving swiftly due to the increasing number actors, many of which typically have multiple roles, goods and services, technologies, and business models.[28] The data ecosystem is seen in this report as a combination of layers corresponding to key roles of actors, where the underlying layers provide goods and services to the upper layers.

A first layer of actors includes Internet service providers; these form the backbone of the data ecosystem through which data is exchanged. A second layer includes IT (hardware and software) infrastructure providers that offer data management and analysis tools and critical computing resources – including, but not limited to, data storage servers, database management and analytic software, and (most importantly) cloud computing resources. The third layer includes data (service) providers: i) data brokers and data marketplaces that commercialise data across the economy; ii) the public sector with its open data initiatives (see Chapter 10); and iii) consumers, which actively contribute their data to the data ecosystem through new services provided by innovative businesses and through data portability initiatives. A fourth layer includes data analytic service providers – businesses that provide data aggregation and analytic services, mainly to business customers. Finally, there are data-driven entrepreneurs that build their innovative businesses based on data and analytics available in the data ecosystem. DDI from these entrepreneurs can be applied to science and research (see Chapter 7), health care (Chapter 8), and smart cities (Chapter 9), and public service delivery (Chapter 10).

Interactions in the ecosystem

Interaction among the actors that structure the data ecosystem could best be described as "co-opetition", a combination of competition and collaboration. As with many innovation ecosystems, collaboration among individual companies allows them to create value that no single company can deliver on its own. Promising (and often specialist) start-ups emerge, which are eventually acquired by larger companies wishing to improve and augment their propositions with analytics platforms, visualisations and applications (ESG, 2012). In the past five years the focus on mergers and acquisitions, in terms of both deals and (especially) investments, has shifted from big data infrastructure to big data analytics and applications.

Recent years have also seen the emergence of "data markets" – online services that host data from various publishers and offer the (possibly enhanced) data to interested parties (Dumbill, 2012). One important distinguishing factor between data brokers and data market providers is that data brokers are actively engaged in the collection of additional data and their aggregation, while data market providers are intermediaries through which data controllers (including brokers) can offer their data sets.

The data ecosystem's value chains are truly global; companies increasingly divide up their production processes and locate productive activities in many countries. Data may be collected from consumers or devices located in one country through devices and apps developed in another country. They may then be processed in a third country and used to improve marketing to the consumer in the first country and/or to other consumers around the globe. Many global value chain activities are captured in international trade, not only in ICT services provided by actors in the IT infrastructure layer, but also in other data-intensive services such as finance, e-commerce, and research. In fact the leading OECD importers of ICT-related services are also the major sources for trade-related data.

Key challenges in the global data ecosystem

The globally distributed nature of the data ecosystem, its extreme interconnectedness, and the interdependencies of its actors and their technologies and resources raise a number of policy issues. One such challenge is the difficulty of value attribution: this challenges measurement but also taxation policies. A number of governments have raised concerns that characteristics of the global data ecosystem could create opportunities for base erosion and profit shifting (BEPS) through "aggressive tax planning by multinational enterprises making use of gaps in the interaction of different tax systems to artificially reduce taxable income or shift profits to low-tax jurisdictions in which little or no economic activity is performed" (OECD, 2014b). A second concerns the key points of control and competition: some dominant actors in the data ecosystem may have significant control and power over certain activities through which the data ecosystem could be shaped, and eventually disrupted. The third involves the free flow of data, which favours global competition among actors of the data ecosystem. Barriers to the flow can limit the effects of DDI, by limiting for example trade and competition. Finally, there remain barriers to data interoperability – especially in sectors requiring significant investment, with a high threshold for new entrants – and portability, which refers to the capacity of reusing data for new applications.

The following four chapters focus on the key issues that decision and policy makers need to consider in more detail, including:

1. the key factors through which decision and policy makers can leverage DDI for growth and well-being: (i) "How data now drive innovation – The focus of Chapter 3", and (ii) "Drawing value from data as an infrastructure – The focus of Chapter 4"
2. the two major policy challenges decision and policy makers need to address to mitigate the economic and societal risks that come with DDI: (i) "Building trust for data-driven innovation – The focus of Chapter 5", and (ii) "Skills and employment in a data-driven economy – The focus of Chapter 6".

How data now drive innovation – The focus of Chapter 3

While the importance of data, both economically and socially, is not new, a confluence of three major socio-economic and technological trends along the data value cycle (Figure 1.7) is making DDI a new phenomenon today and a new source of growth.

The enablers of data-driven innovation

The first is the exponential growth in data generated and collected, driven by high-speed mobile broadband, and the Internet of Things, including sensors and sensor

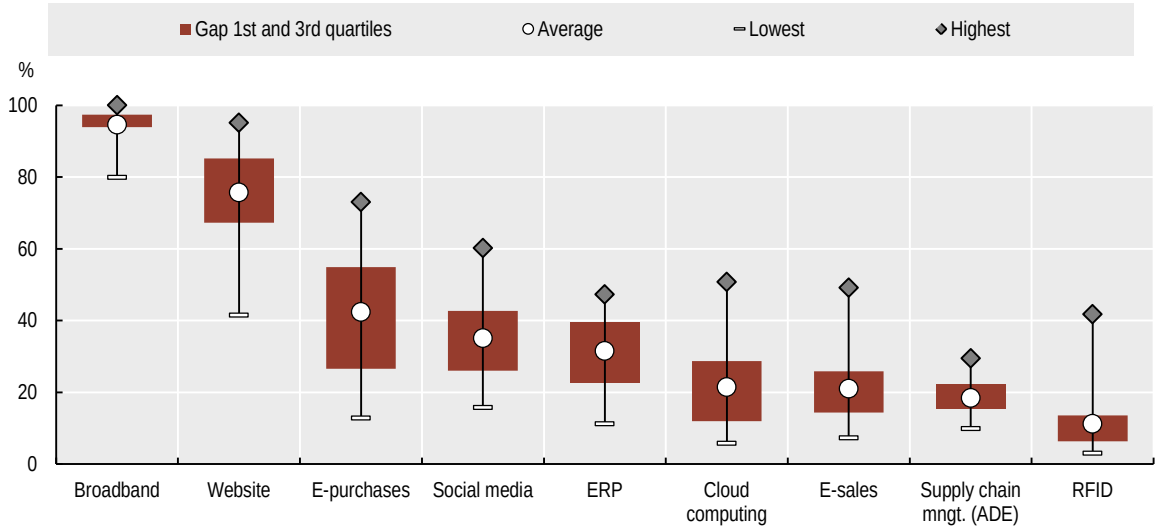
networks enabling the ubiquitous "datafication" of the physical world, and machine-to-machine communication (M2M) empowering data exchange in that world. It is estimated that the average number of Internet-connected devices per household in OECD countries – which today totals ten for an average family of four persons (including two teenagers) – could reach 50 by 2022. One question that arises is whether networks will be able to support all the devices that will be coming on line (an estimated 50 billion devices by 2025).

The second major development favouring DDI is the pervasive power of data analytics, which is now becoming affordable for start-ups and small and medium-sized enterprises (SMEs). The huge volume of data generated by the Internet has no value if no information can be extracted from the data; data analytics refers to a set of techniques and tools that are used to extract information from data, revealing the context in which the data are embedded and their organisation and structure. It reveals the "signal from the noise" – patterns, correlations and interactions among all the pieces of information. The adoption of data analytics has been greatly facilitated by the declining cost of data storage and processing. Cloud computing has been key to this cost reduction. There remain, however, significant issues limiting adoption of cloud computing, which is still used to a much less degree than the high level of broadband connectivity and website adoption would suggest (Figure 1.8). Besides a low capacity to change of many businesses (see Chapter 6), privacy and security (Chapter 5) are among the two most pressing issues limiting cloud computing adoption. Another major challenge is the lack of appropriate standards and the potential for vendor lock-in due to the use of proprietary solutions: applications developed for one platform often cannot be easily migrated to another application provider.

The third factor is the emergence of a paradigm shift in knowledge creation and decision making. Those two moments – when data are transformed into information and knowledge (gaining insights) and then used for decision making (taking action) – are when the social and economic value of data is mainly reaped. Separating these concepts (data, information, and knowledge) is important to better understand data-driven value creation. The distinction can help explain how it is one can have a lot of data but not be able to extract value from them when not equipped with the appropriate analytic capacities; and how one can have a lot of information (extracted from data), but not be able to gain knowledge from it, a phenomenon nowadays better known as "information overload". Data analytics today can help gain insights through i) extracting information from unstructured data (i.e. that lack a predefined data mode); ii) real-time monitoring; and iii) inference – the "discovery" of information even if there was no prior record of such information, through "mining" available data for patterns and correlations – and prediction.

Figure 1.8. **The diffusion of selected ICT tools and activities in enterprises, 2013**

Percentage of enterprises with ten or more persons employed



Note: For countries in the European Statistical System, sector coverage consists of all activities in manufacturing and non-financial market services, and data on e-purchases and e-sales refer to 2013. For Australia, data refer to the fiscal year 2013/14, ending on 30 June and include agriculture, forestry and fishing activities. For Canada and Japan, data refer to 2013 except cloud computing (2012). For Korea, data refer to 2013. For Mexico, data refer to 2012 and to establishments with 10 or more persons employed. For New Zealand, data refer to the fiscal year 2013/14, ending on 31 March. For Switzerland, data refer to 2011.

Source: OECD (2015c).

The implications for the quality of decision-making

Data now have an even bigger role in the decision-making process than in the past. Three major trends account for this: (i) Human decision making is increasingly based on rapid data-driven experiments; (ii) crowdsourcing – "the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community" (Merriam-Webster, 2014) – has been made further affordable; and (iii) decision-making is increasingly being automated thanks to advances in artificial intelligence. In fact, one of the largest impacts of data on (labour) productivity is expected to come from decision automation, thanks to "smart" applications that are "able to learn from previous situations and to communicate the results of these situations to other devices and users" (OECD, 2013d).

As a result of these three major trends, analytics obviates the need for decision makers to understand the phenomenon before they act on it: in other words, first comes the analytical fact; then the action; and last, if at all, the understanding.[29] This can raise serious issues, in particular because the use of data analytics does not come without limitations. There are considerable risks that the underlying data and analytic algorithms could lead to unexpected (false) results, a risk heightened when decision making is automated. Three types of errors could occur: (i) those due to poor-quality data (which will almost always lead to poor results); (ii) those that come with inappropriate use of data and analytics (there will be wrong results if the data used are irrelevant and do not fit the business or scientific questions they are supposed to answer); and (iii) those caused by unexpected changes in the data environment. These last may be intentional; sometimes analytics can be easily "gamed" once the factors affecting the underlying algorithms have

been understood. Or the wrong results may not be intentional and due to constant changes in the data environment; patterns in the data collected are therefore hardly robust over time.

Drawing value from data as an infrastructure – The focus of Chapter 4

Data has become a key infrastructure for 21st century knowledge economies. Data are not the "new oil" as still too often proclaimed. They are rather an infrastructure and capital good that can be used across society for a theoretically unlimited range of productive purposes, without being depleted.

Open data, data commons, and data philanthropy

Data provide economies with significant growth opportunities through spillover effects in the support of the downstream production of goods (including public and social goods).[30] And as with any infrastructure, there can be significant (social) opportunity costs in limiting access. Open (closed) access enables (restricts) user opportunities and degrees of freedom in the downstream production of private, public and social goods (Frischmann, 2012). Especially in environments characterised by high uncertainty, complexity and dynamic changes, open access can be an optimal (private and social) strategy for maximising the benefits of an infrastructure. Data markets may not be able to fully serve social demand for data if there is a demand manifestation problem – as there certainly can be – in the data ecosystem. In addition, the context dependency of data and the dynamic environment in which some data are used (e.g. research) make it almost impossible to fully evaluate ex ante the potential of data, and would exacerbate a demand manifestation problem.

This calls for governing data through non-discriminatory access regimes and commons (see Frischmann, Madison and Strandburg, 2014). In contrast to Hardin's (1968) "tragedy of the commons", where free riding on common (natural) resources leads to degradation and depletion of resources, the "comedy of the commons" (Rose, 1986) – where greater social value is created with greater use of common resources – is possible in the case of non-rivalrous goods such as data. This is the strongest rationale for policy makers to promote access to data, either through "open data" in the public sector, "data commons" such as in science, or through the more restrictive concept of "data portability" to empower consumers. The accumulation of data does come with certain costs (e.g. storage) and risks (e.g. privacy violation and digital security risks). Nevertheless, the advantages for individuals and businesses are clear.

Most definitions for open data point to a number of criteria or "principles". According to the OECD (2006) *Council Recommendation on Principles and Guidelines for Access to Research Data from Public Funding*, for example, openness means i) access that should be granted on equal or non-discriminatory terms, and ii) access costs that should not exceed the marginal cost of dissemination. And it is important to note that the concept of open data is not limited to science (Chapter 7) or the public sector (Chapter 10). For instance, "data philanthropy", whereby the private sector shares data both to enable societal benefits such as by supporting more timely and targeted policy action for development.

Towards a common data governance framework

Among the criteria listed in the many definitions of open data, non-discriminatory access (or "access on equal terms", as stated in the OECD [2006] Recommendation) is

central. Access independent of identity and intent can be crucial for maximising the value of data across society, as it keeps the range of opportunities as wide as possible. Three factors in particular affect the level of non-discriminatory access.

One is the data's technological design – they need to be made available, ideally on line; machine readable, i.e. structured; and linkable. Intellectual property rights (IPRs) are a second factor, for they can limit or prevent the (re-)use and distribution of open data. Some open data initiatives therefore explicitly state that open data should be free of any IPRs, although in other cases innovative IP regimes are used and even promoted through open data regimes, as long as they do not restrict the rights of users to reuse and sometimes redistribute the data.

Pricing, the third factor, will have less of an impact on the degree of openness than technological design or IPRs, but it can still be one of the most challenging factors, because optimal pricing can be hard to determine. Many governments wish to engage in cost recovery, partly for budgetary reasons and partly based on the principle that those who benefit should pay. But the calculation of the overall benefits can be problematic due to significant spillover effects through the creation of public and social goods based on open data. Furthermore, as Stiglitz et al. (2000) have argued, if government provision of a data-related service is a valid role, generating revenue from that service is not. Many open data initiatives therefore encourage the provision of data "at the lowest possible cost, preferably at no more than the marginal cost" as stated in the OECD (2005) Recommendation.

Pricing is challenging mainly due to the fact that data have no intrinsic value, as the value depends on the context of their use. A number of factors can affect that value, in particular the accuracy and the timeliness of data. The more relevant and accurate data are for the particular context in which they are used, the more useful and thus valuable data will be. This of course implies that the value of data can perish over time, depreciating as they become less relevant for their intended use. There is thus a temporal premium that is motivated by the "real-time" supply of data, for example in the financial sector.

Better data governance regimes are needed to overcome barriers to data access, sharing and interoperability. These regimes can have an impact on the incentives to share and the possibility of data being used in interoperable ways. The elements to consider for effective data governance include data access and reuse; portability and interoperability; linkage and integration; quality and curation; "ownership" and control; and value and pricing.

Ownership is singled out, because it a questionable appellation when it comes to data and personal data in particular. In contrast to other intangibles, data typically involve complex assignments of different rights across different data stakeholders. Those different stakeholders will typically have different power over the data, depending on their role. In cases where the data are considered "personal data", the concept of data ownership by the party that collects personal data is even less practical since privacy regimes grant certain explicit control rights to the data subject, as for example specified by the Individual Participation Principle of the OECD (2013e) *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*.

Building trust for data-driven innovation – The focus of Chapter 5

Critical to reaping the substantial economic benefits of DDI – as well as to realising the full social and cultural potential of that innovation – is trust. Trust is a complex issue,

and yet there is consensus that it plays a central if not vital role in social and economic interactions and institutions. Trust is seen as central for efficiency gains realised thanks to the reduction of transaction costs in social and economic interactions. In reducing transaction costs and frictions, trust generates efficiency gains. Trust is therefore considered by some to be a "social capital" and a determinant of economic growth, development, and well-being. The OECD (2011) provides quantitative evidence that high country trust is strongly associated with high household income levels. While trust can be built, it can also erode over time if overexploited as discussion on the recent financial crisis (Allen, 2013; OECD, 2013f) and the revelations about intelligence gathering (Croft, 2014; Naughton, 2015) have suggested.[31] The main components of *trust in the digital economy* are security, privacy and consumer protection.

From traditional security to digital security risk management

DDI relies on an intricate, hyper-connected ICT environment in which security threats have changed in both scale and kind. They include organised crime groups, "hacktivists", foreign governments, terrorists, individual "hackers" – and sometimes, business competitors. There are in addition the non-intentional digital threats, such as hardware failure and natural disasters.

Many stakeholders continue to adopt a *traditional security approach* that not only falls short of appropriately protecting assets in the current digital environment, but also is likely to stifle innovation and growth. That traditional approach aims to create a digital environment secure from threats that can undermine the "AIC triad": data's availability (accessibility and usability upon demand by an authorised entity); integrity (quality in terms of accuracy and completeness); and/or confidentiality (prevention of data disclosure to unauthorised individuals, entities or processes). To preserve each of these dimensions, security experts put in place "controls", "mechanisms" or "safeguards", generally based on technologies, that form a perimeter around the protected assets to secure them.

The problem here is that data-intensive economic and social activities introduce a level of complexity to the point where the traditional security approach cannot scale up. First, these data-intensive activities rely on information systems and networks to become more open and interconnected, enabling data flows to be exchanged easily, flexibly and cheaply, with a potentially unlimited number of partners outside the perimeter. Second, DDI relies on the capacity to exploit the dynamic nature of the digital environment – rapidly connecting, matching and analysing what was previously not related in order to create new assets. Third, traditional security can deal with increased volumes and diversity if the data are located within that defined perimeter and their processing is not subject to continuously unpredictable uses and flows. However, the uncertainty already introduced by the open and dynamic nature of DDI grows, sometimes exponentially, with these increases.

As a result, the traditional security approach, which can only operate at the cost of reducing complexity and increasing stability, will inevitably slow innovative usage and, ultimately, undermine the economic and social benefits of interoperable ICTs.

With the risk-based management approach, the value of data-intensive activities is not limited to the digital storage and processing of a large quantity of data ("big data"), but rather to the capacity to manage a data value cycle (Figure 1.7). The objective of digital security risk management is therefore to increase the likelihood of economic and social benefits from the data value cycle by minimising potential adverse effects of uncertainty

related to the availability, integrity and confidentiality of the cycle (the AIC triad). Unlike the traditional security approach, digital security risk management does not aim to create a secure digital environment to eliminate risk. Instead, it creates a framework to select proportionate and efficient AIC security measures in light of the benefits expected from the cycle.

That raises the key question of responsibility. Traditional security focuses on securing the digital environment. Therefore, in most cases, the party responsible for the provision of the environment (generally the IT department) takes responsibility for its security, and users of the environment do not have to be concerned with it. In contrast, from a digital security risk management perspective, responsibility cannot be delegated to a separate party. Managing risk means accepting a certain level of risk – or deciding not to accept it, and therefore not to realise the benefits. The primary responsibility for managing risk should therefore mirror the responsibility for achieving the objectives and realising the benefits (leadership).

Privacy protection for data-driven innovation

Each step of the data value cycle (Figure 1.7) on which data-driven innovation relies can raise privacy concerns. *Step 1* is the initial data collection, which is becoming increasingly comprehensive, diminishing an individual's private space. Some of the data collected is *volunteered* and thus knowingly and willingly provided by the individual as it is often essential to the completion of an online transaction. An increasing share of data in contrast is *observed*, based on the online tracking of individuals and the collection and analysis of related personal information.

Step 2 is the massive storage of data, which increases the potential of data theft or misuse by malicious actors and other consequences of a data security breach, the risks of which may not be easy to ascertain. Where personal data are collected, stored or processed, security incidents can heavily affect individuals' privacy as high-profile *data breaches*[32] have demonstrated. Cyber-attacks still remain the most frequent cause for data breaches in terms of records stolen but not in number of incidents. These incidents come along with significant costs to individuals but also to the firms suffering the data breaches.

Steps 3 and 4 involve inferences of information and knowledge enabled by data analytics, which often go well beyond the data knowingly provided by a data subject, diminishing an individual's control and creating information asymmetry. Advances in data analytics, make it increasingly easy to generate *interferences* from data collected in different contexts, even if individuals never directly shared this information with anyone. Once *linked* with sufficient other information, data analysts can predict, with varying degrees of certainty, the likelihood that an individual will possess certain characteristics, building a profile. This increased capacity of data analytics is illustrated by Duhigg (2012) and Hill (2012), who describe how the United States based retailing company Target "figured out a teen girl was pregnant before her father did" based on specific signals in historical buying data.[33]

Finally, data-driven decision making (*Step 5*) can lead to a real-world (discriminatory) impact on individuals and other harms. Concerns have been raised that the information inferred through data analytics could be used to exploit the vulnerabilities and receptiveness of individuals in a way that not only induces them to undertake certain actions (e.g. purchase products), but that alters their preferences for these actions. In addition, private actors increasingly rely on the predictive capabilities of data analytics in

their search for competitive advantage. While these predictive analyses may result in greater efficiencies, they may also perpetuate existing stereotypes, limiting an individual's ability to escape the impact of pre-existing socio-economic indicators. A well-known example in this regard is "price discrimination" where firms are selling the same good to different customers for different prices, even though the cost of producing for the two customers is the same. Certain uses of data analytics may also have more serious implications for individuals, for example, by affecting their ability to secure employment, insurance or credit, and this is the more severe when decision-making processes are fully automated.

Data analytics may thus impact core societal values such as individuals' liberty, when for instance creating a "chilling effect" in which an individual curtails communications and activities in fear of uncertain but possibly adverse consequences, or a "filter bubble" (Pariser, 2012) which narrows the range of views exposed to an individual as a result of efforts to personalise content and other products and services. Where the issue of information asymmetry is further exacerbated by the limited transparency of data analytics, individuals will remain unaware that data analytics is affecting their decision making and even preferences, and they will have considerable difficulty ascertaining how exactly analytics is being used to influence them.

There have been several policy responses to improve the effectiveness of privacy protections in the context of DDI. One set of initiatives is grouped under a heading of improving transparency, access and empowerment for individuals. An element in a number of these initiatives is data portability, which allows users to more easily change data controllers by reducing switching costs, and enables them to analyse their own data for their own benefit by receiving it in a usable format (see also Chapter 4). Another emerging element includes the means through which the transparency of the processes and algorithms underlying data analytics (i.e. *algorithmic transparency*)[34] can be increased (see Box 1.3, sell also Annex on the highlights of the *2014 OECD Global Forum on the Knowledge Economy*). A second area of focus is the promotion of responsible usage of personal data by organisations. The promise of technologies used in the service of privacy protection has been long noted as another area. Finally, application of risk management to privacy protection is highlighted as providing another possible avenue. Perhaps the most difficult policy prescription is a need for greater effort to articulate substantive boundaries within which responsible uses of data and analytics would be limited, including the boundaries within which fully automated decision-making would be appropriate. Determining where these boundaries lie – and who should make this determination – will become an increasingly necessary task.

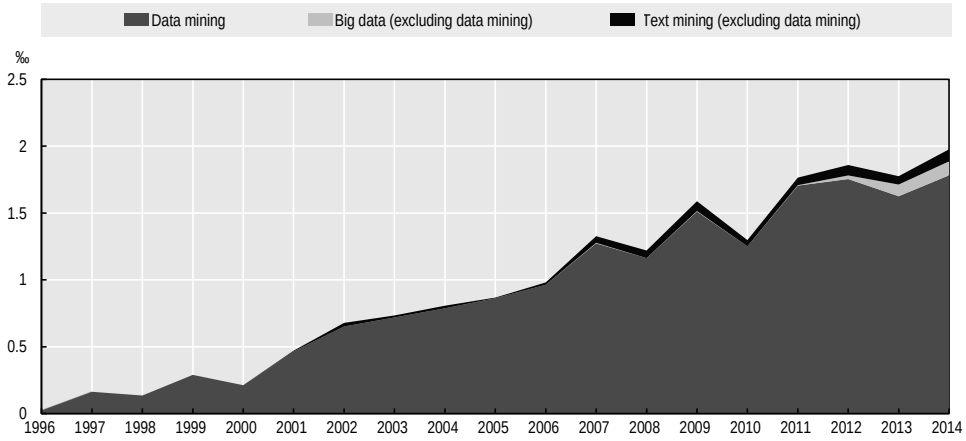
Box 1.3. **The role of an open scientific community for algorithmic transparency**

Increasing transparency related to the functioning of data analytics can be challenging as it may in some cases put at risk proprietary intellectual property rights (IPRs) including trade secrets, which some businesses would consider the "secret sauce" of their business operations (OECD, 2015b). Open scientific communities can play a key role for enhancing algorithmic transparency, while preserving the IPRs of data controllers and increasing awareness about the potentials and risks of data analytics (see Chapter 7 of this volume for further discussions on open science). Many innovative uses of data analytics are disclosed and discussed in conferences and/or scientific papers. Within the last 10 years between 2004 and 2014 scientific articles on data analytics (and related terms) have grown by 9% a year on average (Figure 1.9).

Box 1.3. **The role of an open scientific community for algorithmic transparency (cont.)**

Figure 1.9. **Data analytics related articles in the Science Direct repository, 1995-2014**

Per thousand articles available



Source: OECD, 2014c, *Measuring the Digital Economy: A New Perspective*, based on ScienceDirect repository, www.sciencedirect.com, July 2014.

Two cases are highlighted here as illustrative for the potential of open scientific communities:

- Target's use of data and analytics to predict pregnancy (mentioned above) – In this case, the work was presented in 2010 by a Target statistician at the Predictive Analytics World (PAW) Conference under the title "How Target Gets the Most out of Its Guest Data to Improve Marketing ROI" (Pole, 2010). The author's presentation remained unnoticed by the public for two years, until Duhigg (2012) and later Hill (2012) discussed Target's practice.

- Facebook's experiment on massive-scale contagion conducted by Facebook in 2012 – This experiment was disclosed in Kramer et al. (2014) and based on the manipulation of content presented to more than 689 000 Facebook users. The result showed that "emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks" (Kramer et al., 2014). In the case of Facebook's experiment, it was the paper (Kramer et al., 2014) that revealed the details of the experiment and that was immediately discussed by the public after being published and made openly available through the *Proceedings of the National Academy of Sciences* (PNAS).

Skills and employment in a data-driven economy – The focus of Chapter 6

DDI is disruptive and may induce the "creative destruction" of established businesses and markets. Evidence suggests that creative destruction is an essential engine of long-term economic growth in market economies. In particular in the current context of weak global recovery, business and policy leaders need "to take advantage of the process of 'creative destruction' to accelerate structural shifts towards a stronger and more sustainable economic future" (Guellec and Wunsch-Vincent, 2009). However, managing structural change is challenging, as the pursuit of (short-term) profit may result in

reluctance to change – see what Christensen (1997) refers to as the "innovator's dilemma". In addition, too many businesses (and their employees) may have a weak capacity to change and realise the potential of DDI as suggested by the still low level of ICT adoption across countries (see Figure 1.8).

Structural change in labour markets

DDI may further increase pressure for structural change in labour markets, since it enables the automation of an increasing number of cognitive and manual tasks.[35] This includes the use of data analytics for a wider range of intellectually demanding tasks, such as diagnosis of diseases based on analysis of complex information. It also includes use of a new generation of autonomous machines and robots that are no longer restricted to very precisely defined environments, and that can be deployed and redeployed at much faster rates compared to current generation robots.

Many observers see a high risk that "smart" applications will further broaden employment polarisation, at least in the short run. The effects could be that more middle income jobs may be negatively affected – jobs largely held by the segment of the population that "glues" our societies together. Furthermore, DDI will also affect manufacturing, and in fact could reduce the number of blue collar jobs needed.

Certain categories of jobs, however, are less likely to be susceptible to computerisation and (data-driven) automation. These include jobs that involve solving unstructured problems, including problems that lack rules-based solutions; working with new information, including making sense of new data and information for the purpose of problem solving or decision making, or to influence the decisions of others; and non-routine manual tasks, carrying out physical tasks that cannot be well described via rules because they require optical recognition and fine muscle control that continue to prove difficult for robots to perform. While solving unstructured problems and working with new information will be particularly important for high-end jobs, carrying out non-routine manual tasks will become more and more important for low-paying jobs.

The growing importance of data specialist skills and employment

At the same time, the increasing use of data and analytics across the economy has driven demand for new types of skills and jobs, most prominently involving data specialisation. All these data-specialist professions have one common denominator: working with data constitutes a main part of the job. However, there is currently a relatively low availability of critical data specialist skills and competence required for DDI, and this may prove not only a barrier to adoption of DDI, but also a missed opportunity for job creation: some have suggested that the demand for data specialist skills exceeds the supply on the labour market. An Economist Intelligence Unit (2012) survey, for instance, shows that "shortage of skilled people to analyse the data properly" is indicated as the second biggest impediment to make use of data analytics. There is especially a need for people that possess the skills needed for extracting insights from data.

Employment opportunities will remain for people with the right mix of skills and competencies. The most data-intensive industries employing the highest share of data specialists are still the ICT service industries, and in particular i) IT and other information service industries, ii) insurance and finance, iii) science and research and development, and iv) advertising and market research. But v) the public, and (vi) health care sector are identified as promising areas for new data specialist jobs.

Policy considerations for smoothing structural change

DDI may further increase unemployment and inequality through skill-biased technological change, if not addressed by policy measures – including via social and tax policies. This suggests the need for a smart "double strategy" that promotes continuous education, training and skills development, while addressing the risks of worsening inequality in earnings in labour markets. That need is especially acute given the current weak global recovery and lingering high unemployment in major advanced economies. In the context of DDI, inequality could become a major issue if access to urgently needed high-quality education to take advantage of the job creation opportunities ahead is limited to a few.

Education systems should support a broader interdisciplinary understanding of multiple complex subjects but also deeper insights into some domain-specific issues. Soft skills such as creativity, problem solving and communication skills are key for ensuring employment in a data-driven economy; skills involving fine muscle control will also become a key competitive advantage of humans over machines. These skills, if cultivated with the support of education systems and accompanied by political attention and good co-operative global governance, may lessen concerns related to technological unemployment. This will be more the case, if individuals can enhance and complement their talents to use technology to "dance" with the machines instead of "racing" against them.

The policy considerations discussed in the first six chapters presented above apply across all sectors and application domains. There are in addition domain-specific issues that policy makers need to address, in particular when promoting DDI in specific sectors. As highlighted above, the "low hanging fruit" from the adoption of DDI are in research and education, health care, as well as the public sector. The remaining four chapters presented below focus on these key areas, and highlight specific opportunities and challenges that decision and policy makers need to consider in more detail:

1. Promoting data-driven scientific research – The focus of Chapter 7
2. The evolution of health care in a data-rich environment – The focus of Chapter 8
3. Cities as hubs for data-driven innovation – The focus of Chapter 9
4. Governments leading by example with public sector data – The focus of Chapter 10.

Promoting data-driven scientific research – The focus of Chapter 7

Data analytics now makes it possible to collect, generate, access, use and reuse research and scientific material (articles and data sets but also images and digital lab records) at no or extremely low marginal cost. As a result, the speed at which knowledge can be transferred among researchers and across scientific fields can be increased, opening up new ways of collaborating and new research domains. Different scientific domains are becoming increasingly interconnected: data generated in one field of research may nowadays be treated with models and techniques traditionally belonging to other fields of research. The term often used to describe this transformation of science into a more open and data-driven enterprise is *open science*.

Impacts of open access to scientific data

Data generated at the global level that relate to issues of global concern, such as the environment and climate change or the ageing population and health, may be more

powerfully exploited if properly interconnected and used by large networks of scientists and researchers world-wide. This can avoid duplication of effort and enhance co-ordination in science and research. Furthermore, open scientific data have the potential to strengthen relations between the scientific community and society. Scientists now have a broader range of mechanisms (e.g. through social networks, personal scientific blogs, videos, interviews and discussion forums) to communicate with citizens. And these new scientific communication mechanisms can help build public trust in science.

Collaborative efforts in science and research can thus reach beyond the research community to increasingly involve citizens and "amateur researchers" at different stages of scientific processes, from data collection to solving more complex scientific problems. The involvement of non-professional scientific communities in science and research efforts is often referred to as *citizen science* and it comes with several benefits; for example, it allows the development of a more democratic environment in science by engaging amateurs as well as professionals in research and scientific efforts.

There are other examples of crowdsourcing for technical skills that can solve scientific problems, such as online platforms where solutions to scientific problems are requested from the public. Private companies and research teams publish unsolved problems related to specific data sets (also published on the platform), and data scientists from all over the world compete to find the best solutions and highest-performing algorithms for prize money. The approach relies on the fact that there are countless strategies to solve the problem, each with a different computational efficiency.

The involvement of citizens in scientific projects tends to have an educational value, both implicit and explicit. While in the majority of projects the informal learning aspect of adult citizens is addressed, schools are increasingly considered an important target for the introduction and promotion of citizen science. Teachers play a significant role in facilitating the deployment of experiments and transmitting the socio-scientific values of their contributions to the young audience. In fact, a number of countries are investing in the educational skills building necessary for data analytics, as these skills are currently lacking.

Greater access to scientific inputs and outputs can improve the effectiveness and productivity of the scientific and research system, by reducing duplication costs in collecting, creating, transferring and reusing data and scientific material; by allowing more research from the same data; by multiplying opportunities for domestic and global participation in the research process; and by ensuring more possibilities for testing and validating scientific results. With unrestricted access to publications and data, firms and individuals may use and reuse scientific outputs to produce new products and services. Developing countries in particular may benefit from open access to scientific material.

Challenges to open data and data sharing in science and research

However, several barriers to data sharing still remain. Some are of a technical nature, such as issues related to storage, the technical infrastructure to allow data sharing, interoperability and standards. Other types of barriers are related to the lack of an open data culture or the disincentives that researchers and scientists face with respect to the disclosure and sharing of data sets, especially relative to research at the pre-publication stage. This raises the question of the "optimal" level of openness to boost research and innovation without discouraging data collection from individual researchers.

Additional challenges relate to the definition of ownership of the data itself. Barriers to legal, cultural, language and proprietary rights of access hinder cross-national collaboration and international data exploitation, especially in the social sciences. There are issues with regard to propriety databases that could impede open research data efforts in academia as well. There is thus tension between open research data and IPRs, and a balance must be struck between efforts to promote open data in science and efforts to promote commercialisation of public research, especially in the case of public-private partnerships involving companies. The tension can however be lessened by policies that clarify IP ownership and promote non-exclusive licensing possibilities, as well as by greater IP awareness among researchers.

Another legal issue that comes into play in the context of open scientific data is privacy and personal data protection. Data gathered in the course of research often contain personal information (e.g. medical records), and so opening such data has to respect the rights of data subjects (Lane et al., 2014). This does not mean that the data cannot be opened, but it does calls for implementing effective protective procedures.

Data collection, curation and sharing vary by scientific discipline; some fields have been traditionally more data-intensive than others. Researchers belonging to scientific disciplines not involving large-scale experiments managed by teams of hundreds of researchers, notably in the social sciences and humanities, traditionally collect and built their own data sets, in some cases manually or by developing surveys and questionnaires. That makes this kind of the data set more tied to the individual researcher, and therefore less easily ready to be shared without proper curation, cleaning and metadata compilation. Scientists and researchers do not have necessarily the incentives or the skills to perform those tasks, since proper curation and dissemination of data sets are costly and time-consuming. Also, they traditionally compete to be first to publish scientific results, and may not see the benefits of disclosing information on the data they want to use to produce as yet unpublished research outcomes.

A possible solution to the above-mentioned disincentives is data citation. Researchers wishing to be acknowledged for their work could release data sets through mechanisms similar to the one already in place for citations of academic articles. Data citation is not, however, necessarily a standardised or widely accepted concept in the academic community.

The evolution of health care in a data-rich environment – The focus of Chapter 8

The health sector is a knowledge-intensive industry: it depends on data and analytics to improve therapies and practices. There has been tremendous growth in the range of information being collected, including clinical, genetic, behavioural, environmental, financial and operational data. Every day, health care professionals, biomedical researchers and patients produce huge amounts of data of great value from an array of devices. At the same time, the potential to process and analyse these emerging multiple streams of big data and to link and integrate them is growing.

Drivers of growth of digitised health data

Five principal factors drive the increased collection and use of large-scale data in the health sector. One is demographic change and a 20-year shift in the burden of disease, from infectious conditions to long-term non-communicable diseases (NCDs) brought on by lifestyle choices and environments. A second factor is that fiscal pressures have led to

a need for greater efficiencies. Continuing pressure to find ways to make systems more productive has moved the focus from cost containment to performance-based governance. To evaluate health sector performance, managers and governments will need timely and accurate information about the prices and volumes of services provided and the health outcomes produced, at levels sufficiently detailed to take corrective policy action.

The third has to do with the role of the patients themselves in the care process, which has taken on much greater importance in recent years. Patients' taking command of and managing their health will especially aid in the management of chronic diseases. The fourth driver is the need for co-operation to tackle global public health challenges such as infectious diseases, and improve early detection and warning of emerging health threats and events. Complementing the traditional case-based and syndromic surveillance systems, monitoring of unstructured events – through news and Internet media, web searches, etc. – has been a significant component of public health early warning and response over the past decade.

But the fifth driver of health data use is possibly most important: the sheer volume, velocity and variety of health data available. Many health care systems are rapidly digitising immense amounts of data and using them for a wide range of activities, including preventive care, e.g. early detection; field data to support emergency and urgent care; coaching, rehabilitation and maintenance; context-sensitive intervention, e.g. reminders; epidemiological assessments; post-market surveillance and analysis; health care quality and performance monitoring.

The increasing use of electronic medical records promotes patients' participation – in their own care, in self-management of health conditions, and in informed decision making. Patients' interest in their diagnostic test results and medical records, in their options for care, in the quality of providers, and in scheduling visits on line will keep growing. Over the past decade multiple studies have documented the value of electronic personal records (EPRs) in supporting greater patient-centred services. Patients and practitioners are also increasingly interested in devices, tools and computer applications that assist in monitoring and improving health and well-being. They recognise that these can help patients live longer in their own homes rather than in considerably more expensive hospital or nursing home facilities; and encourage personal responsibility for healthier lifestyles.

Towards smarter models of care

Any systematic effort to address today's health and wellness challenges will also require data to support new and "smarter" models of care. That will require enhanced capacity for the sharing, processing and analysis of health and behavioural data to support patient-centric care, and a more efficient clinical research enterprise for improved prevention and better disease management. Today's care is reactive, episodic and focused on disease. The new health care will need to be proactive, preventive, and focused on quality of life and well-being.

The ubiquitous care model, for example, is based on the utilisation of smart sensing and biometric devices for real-time monitoring, analysis and transmission of health data. The information can then be accessed by health care providers for informed diagnosis, clinical decisions regarding treatments, and evaluation of outcomes. It can also be viewed and acted upon by patients for both education and prevention.

Mobile health (mHealth) offers a wide range of smart modalities by which patients can interact with health professionals, or with systems that can provide helpful real-time feedback along the care continuum, from prevention to diagnosis, treatment and monitoring. mHealth is of particular value in managing health conditions where continuous interaction is important, such as diabetes and cardiac disease. The devices utilised include mobile phones, tablets, global positioning system (GPS) devices, mobile tele-care devices, and mobile patient monitoring devices.

Crowdsourcing is emerging as a means of allowing science to be conducted at scales of magnitude greater than before. It involves capitalising on the Internet and large groups of people, particularly via online Web 2.0 communities, to harvest "collective intelligence" and accomplish tasks that might have traditionally been given to small research groups. Crowdsourcing can help process data quickly, on unprecedented scales, and with better quality control than any individual or small research group can attain. Crowdsourcing therefore has cost and speed benefits, although careful attention must be paid to policy regarding in particular privacy, security, and data stewardship.

Critical success factors and policy priorities

In fact, a number of challenges must be overcome before the benefits from DDI in the health sector can be reaped. One of these is that electronic health records – EHRs – are being collected in health care systems that are often fragmented, with points of care functioning as silos. Questions of privacy also have to be addressed, and skill building will be needed to analyse the voluminous health data sets. Standards and interoperability are other central issues that must be addressed: while health care organisations have access to an ever-increasing number of information technology products, many of these systems cannot "talk" to each other, and if the systems cannot communicate, big data will not meet its potential in the health care system. Attention is also needed to ensure that individuals who wish to restrict or withdraw their data from their contribution to research and statistics can reasonably do so.

Additional critical success factors for governments to realise value from investments in health data are strategic planning; ensuring legislative and regulatory requirements that support planning; engaging all stakeholders in planning and governance; promoting global co-operation; setting standards for data governance; and providing financial stimulus toward data development and use.

Cities as hubs for data-driven innovation – The focus of Chapter 9

A large share of the 65 million sensors estimated to be deployed (e.g. in security, health care, the environment, transport and utilities) are today embedded in urban infrastructures, facilities and environments (MGI, 2011). With around three-quarters of the OECD area population expected to be living in urban areas by 2022, cities will host at least 10 billion out of the 14 billion devices estimated to be in use in member countries by then (OECD, 2010; OECD, 2012). This makes cities a potential hot spot for DDI.

The urban data ecosystem

The data produced and collected in cities can be divided into three categories. There are data on flows; sensors embedded in urban infrastructures increasingly allow the digitisation and datafication of flows of resources, products, people and information across cities. Data on states of urban spaces and environments, subject to constant natural and manmade changes – the density of people or things (e.g. vehicles), air temperature

and quality, light and sound levels, etc. – are monitored by in situ sensors. Finally, data can relate to activities – transaction, consumption and communication patterns that include people's personal and professional activities, communication and interactions; interaction between people and their environment; and interactions among components of their environments, such as communicating or autonomous machines and devices.

Many actors are involved in data collection and use in cities. Key among them are citizens and consumers; innovators and entrepreneurs; governments and utilities; data brokers and platforms; and infrastructure and system operators. Each of them is in principle connected to all the others, through a digital layer and in multiple possible combinations. The extent to which data can be exchanged among these actors and across systems in cities, as well as the extent to which they can easily be reused for different purposes, determines their potential for DDI.

Opportunities for data-driven innovation in cities

Much of the data on flows and states in cities, and some of the data on activities, can be used to increase the efficiency of urban systems and promote their integration. The availability of historical and real-time data on flows in transport, energy, water and waste systems enables analysis at unprecedented depth and granularity, as well as targeted interventions in and precise management of urban systems. So far, the most promising effects of information and communication technologies (ICTs) and data use in cities can be found in transport and electricity, two systems that share an important lever for data-driven improvements: the direct match of demand and supply, based on fuller and often real-time information.

Synergies can be reaped through integration of these systems. Understanding urban infrastructures and sectors as systems, a city can be considered a "system of systems", within which ICTs and the digitisation of urban flows are creating the potential for deep integration (CEPS, 2014). The Internet of Things will continue to multiply the systems, machines, devices and services connected via electricity grids and information systems – such as solar cells on roofs, detailed weather forecasts, home heating systems and air conditioning, supermarket stocks, etc.

Over the past years, innovative start-ups have penetrated established urban sectors with data-driven mobile apps and online platforms. Known under the label "sharing economy", new business models are using real-time and geo-locational data on online platforms and mobile apps that allow commercial "sharing" (renting) of cars, rides and bikes as well as vacant homes, offices and shops in cities. On the supply side, car owners can rent their car if they are not using it, sell seats on trips they are taking anyway, or work as private drivers when time permits; real estate owners can rent out vacant living, office or commercial space for short periods. On the demand side, urbanites get more and cheaper mobility options, and travellers a larger and cheaper choice of accommodations; freelancers and the self-employed gain flexible access to office and commercial space.

City administrations are also increasingly using urban (crowdsourced) data to gain fine-grained real-time information on aspects such as public service delivery, system performance and infrastructure conditions. Mobile apps now allow citizens to report on stray garbage, potholes, broken lamps and the like via their smartphone, directly to city hall. Online, crowdsourced, real-time and geo-locational data can also play an important role in disaster management in cities. The UN Global Pulse project, for instance, uses real-time analytics to turn unstructured online information into actionable information for decision makers to improve resilience.

Greater data availability and more powerful computing are bringing urban modelling back into the spotlight of urban planning, and have the potential to significantly improve the forecasting of societal demand. Geo-referenced data collected via (e.g.) crowdsourcing, remote sensing and social networking – combined with new computational power, including cloud computing – offer fresh possibilities, notably as applied to integrated land use and transport planning (Nordregio, 2014). Data analysis and modelling of societal demand for urban infrastructures and services have the potential to significantly improve resource allocation and investment decisions in urban areas.

Challenges and policy priorities

There are opportunities to be seized in spurring DDI in cities, but there are also challenges. Cities need to build the requisite capacity and skills for collecting, storing and analysing data in a depth and at a scale that are unprecedented, in addition to acquiring the infrastructure and computing power needed to store and process all the data. Sensitive questions need to be addressed when it comes to the type of data cities should collect in the first place and what they should publish thereafter. An important condition for advancing integration of urban systems and system-to-system communication is interoperability across different systems and components at different levels.

Cross-sector data sharing is likely to pose challenges. Data collected in different sectors tend to be stored in different formats, and few incentives exist for harmonising them. Without open standards, data sharing may be limited by and locked into proprietary formats. Linked issues are privacy protection, overcoming silo structures in administrations, and improving co-operation among jurisdictions and levels of government.

While increasing system integration in cities can yield benefits, it also creates new risks. The more ICT, energy, transport and other critical urban infrastructures and systems are interconnected, the more a city as a system-of-systems will become vulnerable to both internal and external threats, ranging from technical failures to cyber attacks and natural disasters. That vulnerability calls for a digital security risk management framework to reduce risk to an acceptable level in light of the expected benefits, through security and preparedness measures that fully support the economic and social objectives at stake.

Governments leading by example with public sector data – The focus of Chapter 10

The public sector is one of the economy's most data-intensive sectors. Its importance as an actor in the data ecosystem is twofold: as a key user of data and analytics, and as a key producer of data that can be reused for new or enhanced products and processes across the economy. The idea behind open access to public sector data is that value can be derived through the reuse of that data by any user from within or outside the public sector. Governments can therefore promote DDI by leading by example in their use and supply of public sector data.

The potential of public sector data

For government – The use of open government data (OGD) by government agencies can lead to efficiency improvements in the public sector. It can, for example, help bring down silos and foster collaboration across and within public agencies and departments. Furthermore, the increasing amount of data made available in formats that enable reuse and linkage is supporting the expansion of data analytics in the public sector; here too,

there is great potential for value creation. Predictive data analytics can, for example, facilitate identification of emerging governmental and societal needs. Use of this data by the public sector can also make for better decisions, inform policies, support the development of data-driven processes and services, and deliver more innovative services. There are also, of course, considerable risks in governments' use of data analytics, in particular with regard to the privacy of citizens.

For citizens – Open government advocates believe that OGD can be a powerful force for public accountability, by making existing information easier to process, combine and analyse. OGD can then promote greater transparency, and allow a new level of public scrutiny that can increase public accountability. E-participation also aims at enhancing citizens' engagement in public life, e.g. in lawmaking, policy making and service design and delivery. Citizens become not just passive consumers of public sector content and services but also active contributors and designers in their own right, empowered to make more informed decisions that can enhance the quality of their lives.

For the private sector – First of all, granting the private sector better access to public sector information (including public sector data) can increase efficiency, effectiveness and innovation in public service delivery. The strategy is to provide innovators from outside governments with the opportunity to develop modular services that are more agile and targeted to citizens' needs than those developed in-house by governments. Secondly, as the importance of data in the development of new services, products and markets has increased dramatically, open access to public sector data can stimulate innovation in the course of that development as promoted by the OECD (2008) *Council Recommendation on Enhanced Access and More Effective Use of Public Sector Information* (PSI).

The OECD market for PSI was estimated to be around USD 97 billion in 2008, and could have grown to around USD 111 billion by 2010. Aggregate OECD economic impacts of PSI-related applications and use were estimated to be around USD 500 billion, and there could be close to USD 200 billion of additional gains if barriers to use are removed, skills enhanced and the data infrastructure improved. There is cross-country evidence that significant firm-level benefits are to be had from free or marginal cost pricing, with small and medium-sized enterprises (SMEs) benefiting most from less expensive data.

Key challenges in implementing open data and PSI strategies

The lack of procedures and standards on how to deal with open data in governments can compromise the quality of the data and eventually the output of OGD and PSI initiatives. Public sector data often are not harmonised, making it difficult from the user perspective to know which data are valid or should be trusted. Critical to access is knowing the source of what one is searching for, and in many instances where to start searching is a challenge. Accessibility can also be limited if data cannot be reused, and data transparency may be hindered if data are not simple to access or reuse due to their format. Interoperability is equally a priority concern for policy makers tasked with implementing OGD or PSI strategies.

The economic climate undoubtedly plays a role. At a time of budget pressures and cuts in government expenditures, it is important to articulate clearly the advantages of opening up public sector data for wider use and, where necessary, to compensate the providers of public sector data for any initial extra funding necessary to open up and digitise the data. Consequently, great emphasis is now being placed on devising more solid methodologies to assess the impact of open access.

Finally, there are organisational, cultural and legal challenges. Having a consistent legal framework in place is critical; fragmented and diverse legislation concerning privacy, reuse of data and related fees can create confusion for end users. And legislation, IT platforms and codes need to be matched by a culture within the public service that supports a presumption to publish, release and share data. Raising awareness of civil servants, citizens, civil society organisations and the private sector with regard to their rights is important for society as a whole to fully capture the benefits of public sector data.

1.3. Common key challenges and policy considerations

DDI is disruptive and comes with major economic and societal challenges and risks to be addressed across all application areas and sectors. Some of the challenges are the result of serious tensions between opposing private and social (collective) interests. Addressing these tensions is complex and cannot be undertaken in silos; these require governments to invite the democratic participation of all citizens – in addition to stakeholders including civil society, the technical Internet community and business groups – in order to be resolved. This also calls for a whole-of-government strategy to promote DDI, as countries such as Australia,[36] Japan,[37] and the United Kingdom[38] as well as the European Commission (EC)[39] have developed or are envisioning.[40]

Two sets of challenges (tensions) need to be addressed by policy makers in order to maximise the benefits of DDI, and mitigate the associated economic and societal risks. A set of key policy issues discussed across the chapters of this volume are related to the need to i) promote "openness" in the global data ecosystem and thus the free flow of data across nations, sectors, and organisations, and at the same time ii) address legitimate considerations of individuals' and organisations' opposing interests (including in particular their interests in the protection of their privacy and their intellectual property rights).[41] Another set of policy issues aims at iii) activating the enablers of DDI, and at the same time ii) addressing the effects of the "creative destruction" induced by DDI, in particular with a focus on small and medium enterprises (SMEs) and on labour markets.

These two sets of tensions may at first appear unrelated. However, a closer look at the policy issues discussed across the chapters suggests that a move towards more "openness" may further the disruptive effects of DDI and thus lead to more "creative destruction". That said, there is no one-size-fits-all optimal level of "openness"; instead the optimal level strongly depends on the domain and the cultural environment in question. Furthermore, in addressing these tensions, policy makers should be aware of the "path-dependency" of current actions (and inaction) that could limit future choice. History of the diffusion of new standards provides examples of path-dependency where early adoption may prevent a more efficient standard at a later stage.[42] In the case of DDI, path-dependency is not only related to the adoption of standards, which play a key role for "openness" in the data ecosystem. The interaction of individuals with data analytics can shape their preferences (including for privacy), and set society on a path that could become impossible to change in the future. This calls for a careful assessment of current policy actions (and inaction) to maximise the long-term benefits of DDI.

Finally, policy makers should acknowledge that DDI may favour concentration and greater information asymmetry and thus shifts in power. This may lead to a new digital (data) divide that could undermine social cohesion and economic resilience. As discussed above, the economic value of DDI is reaped when better insights (knowledge) can be extracted from data. With this knowledge come better insights and more capacity to

influence and control. Where the agglomeration of data leads to greater information asymmetry, power could shift away: (i) from individuals to organisations (incl. consumer to business, and citizen to governments); from traditional businesses to data-driven businesses given potential risks of market concentration and dominance; (iii) from governments to data-driven businesses, where businesses can gain much more knowledge about citizens (and politicians) than governments can; and (iv) from lagging economies to data-driven economies.

Overall, countries will be able to maximise the benefits of DDI, if they can connect to the global data ecosystem (Chapter 2), leverage the enabling factors of DDI (Chapters 3) and promote investments in data as infrastructure (Chapter 4), and address the various key policy challenges (Chapters 5-6), including the domain-specific ones (Chapters 7-10). Given all of this, governments have an important role to play in promoting DDI and mitigating the associated risks.

Annex – Highlights of the 2014 Global Forum on the Knowledge Economy

On the occasion of the 2014 OECD Ministerial Council Meeting, under the Chairmanship of Japan on the 50th anniversary of its accession to the OECD, Ministers affirmed the importance of knowledge-based capital to provide new sources of growth in the face of long-term challenges, such as ageing and environmental degradation, and that the OECD's work on the digital economy is important.

The 4th Global Forum on the Knowledge Economy (GFKE) held in Tokyo, Japan, on 2-3 October 2014, focused on data, one example of knowledge-based capital. Policy makers, business, civil society and other stakeholders from OECD Member and Partner (i.e. non-member) economies participated in active discussions on data-driven innovation for a resilient society.

Throughout the entire forum, participants acknowledged the high value of big data in spurring economic growth or solving various social challenges, and discussed policy options to promote the use of big data that will inform the discussion at future OECD meetings. Highlights of the discussions include:

1. *Illustrating the economic benefits* – Participants discussed the positive economic impacts of big data across industries, and in particular manufacturing, and emphasised that data-driven innovation is likely to promote economic growth in both OECD member and non-member economies, directly or through spillover effects. Participants mentioned the value of optimising existing services and of analytics for decision making. Participants discussed the global dimensions of data-driven innovation, including the importance of cross-border data flows for trade, as well as the need to understand and address the implications of data-driven innovation for jobs.
2. *Addressing complex societal challenges* – Participants recognised the potential of big data analysis for disaster response (for example, based on the ex post analysis of the Great East Japan Earthquake), but also more generally for improving quality of life. They underlined the need for government leadership, awareness and collaboration among all actors in the adoption and implementation of disaster risk management approaches to enhance human security. As an example, it was shown how big data can be used to relieve traffic congestion and improve construction standards.
3. *Leveraging data-driven innovation in ageing societies* – Participants recognised the opportunities that data-driven innovation presents for ageing societies, but agreed that most of the potential for value creation is still unclaimed. They discussed the need to overcome data silos and create the appropriate conditions for broader data access, linkage and integration. It was recognised that local data on vulnerable elderly populations are necessary for central government actions and disaster planning. Defining minimum standards for data was considered essential, as well as interoperability. An important idea put forward was the need to create conditions for a risk-based approach to protect data. Finally, participants concluded that there is a need to strengthen the capacity to analyse data, build expertise, and increase business opportunities.

4. *Promoting skills for the data-driven economy* – Participants were aware of the gap between the demand and supply of data scientists, and the need for skills development and education. Potential displacement effects were highlighted in particular with regard to certain middle income, white collar jobs as well as the need to address the resulting inequality implications. Problem solving and entrepreneurial competences building on human creativity and intuition, in combination with data analysis and software engineering skills, were highlighted as critical as well as basic ICT literacy. Participants recognised the importance of lifelong learning as a means to fill the potential employment gap.

5. *Building trust in the data-driven economy* – Participants recognised that the trust of individuals is crucial and that big data users should respect fundamental values. They underlined the importance of risk-based approaches to the collection and use of personal data. Algorithmic transparency raises complex issues, but providing information on key elements informing decisions is important to avoid discrimination. Other key issues discussed included security, ethics, privacy-enhancing technologies and better metrics. The impact of data concentration on privacy, but also on competition, transparency and accountability, was considered worthy of further examination.

6. *Encouraging open data across society* – Participants underlined the necessity of promoting open data so as to make it possible to use public data to create new services and effective administrative procedures. Public value depends on data use. In response, governments' role is evolving from the direct provision of data and regulation. It also now encompasses the creation of enabling conditions supporting communities of providers and users, building trust, enforcing principles of non-discrimination for public entities, civil society and the private sector to improve open data sharing and use.

7. *Policy conclusions* – Governments and stakeholders need to develop a coherent policy approach to harness the economic benefits of data-driven innovation. They need to assess the context for data collection, analysis and use to ensure that data-driven innovation serves societal values in an ethical and equitable manner.

Notes

1 As a point of reference, one exabyte corresponds to one billion gigabytes and is equivalent, for example, to around 50 000 years of DVD-quality video (see http://exabyte.bris.ac.uk/, accessed 15 May 2015).

2 Estimates are provided by IBM (see www-01.ibm.com/software/data/bigdata/what-is-big-data.html, accessed 15 May 2015).

3 It is estimated that 90% of all this data were created in the past few years (ScienceDaily, 2013; Wall, 2014).

4 These include Ministers from and Representatives of Australia, Austria, Belgium, Canada, Chile, Colombia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, the United States and the European Union.

5 As stated in the OECD (2014a) Ministerial Council Statement, "Rising inequality endangers social cohesion and weakens social resilience, thereby hampering economic resilience. A key challenge is to achieve inclusive growth by providing social protection and empowerment to people, which can strengthen human security. Appropriate flexibility and security in labour markets and relevant education and skill programmes can facilitate greater inclusion and participation of under-represented groups. We welcome OECD initiatives targeting these groups, including on gender equality, youth employment, ageing society and the integration of migrants. We also recognise that regional and urban policies can play a key role in empowering people and building resilience at all levels of our economies and societies".

6 The outcomes of the first phase of the OECD horizontal project on *New Sources of Growth: Knowledge-Based Capital* (KBC1, see OECD, 2013a) were discussed at the conference on "Growth, Innovation and Competitiveness: Maximising The Benefits Of Knowledge-Based Capital" on 13-14 February 2013, and the final conclusions were presented to ministers at the 2013 OECD Ministerial Council Meeting (MCM) (see http://oe.cd/kbcconference).

7 Calculated based on annual balance sheet data as follows: $(p - d) / a$, where p = the total gross value for property, plant, and equipment; d = total accumulated depreciation; and a = total assets.

8 Brynjolfsson et al. (2008) highlighted how information technology (IT) had "enabled firms to more rapidly replicate improved business processes throughout an organization, thereby not only increasing productivity but also market share and market value". Internet firms, however, are not only replicating business processes throughout their organisations, but also increasingly relying on automated business processes that are empowered by software and in particular data analytics.

9 The growth of storage-related technologies and services can be explained by the fact that many top ICT companies are trying to strengthen their market position through the development of new "big data" branded products, many of which are based on open source storage management solutions initially developed by Internet firms such as Hadoop, a major big data technology (see Chapter 2 of this volume).

10 The market value of (traditional) relational database management systems alone was worth more than USD 21 billion in 2011, having grown on average by 8% a year since 2002 according to some estimates (OECD, 2013b).

11 *Social Genome* builds on public data from the web (including social media data) as well as Walmart's proprietary data, such as its customer purchasing and contact data.

12 For example to analyse and predict potentially vulnerable components; the resulting analysis is further used to optimise product design and production control.

13 Similar services are observed in the energy production equipment sector, where M2M and sensor data are used to optimise contingencies in complex project planning activities for instance (Chick, Netessine and Huchzermeier, 2014).

14 Some of the data and analysis results are presented to farmers via the MyJohnDeere.com platform (and its related apps), to empower farmers to optimise the selection of crops and the time and place for planting and ploughing them (Big Data Startups, 2013).

15 The study is based on a survey by Bakhshi and Mateos-Garcia (2012), but extended by "matching survey responses about data activities with historical performance measures taken from respondents' company accounts, and by conducting an econometric analysis of the link between business performance and data activity while controlling for other characteristics of the business". The analysis shows that, other things being equal, a one-standard deviation greater use of online data is associated with an 8% higher-level of total factor productivity (TFP). Firms in the top quartile of online data use are 13% more productive than those in the bottom quartile. The study furthermore shows that "use of data analysis" and "reporting of data-driven insights" have the strongest link with productivity growth, "whereas amassing data has little or no effect on its own" (Bakhshi, Bravo-Biosca and Mateos-Garcia, 2014). Another study by Barua, Mani and Mukherjee (2013) suggests that improving the quality of and access to data by 10%, by presenting data more concisely and consistently across platforms and allowing it to be more easily manipulated, would increase labour productivity by 14% on average, but with significant cross-industry variations.

16 The estimated output elasticity of 3% resulted after controlling for firms' adoption of data-driven decision making. The OLS (ordinary least squares) estimate on the Hadoop measure indicated an output of 10%, which Tambe (2014) attributed to other omitted variable bias, including firms' adoption of data-driven decision making.

17 See also a survey by the *Economist Intelligence Unit* (2012) of business executives, according to which expectations are that the use of "big data" could improve organisational performance by 25% and by more than 40% over the next three years. The use of data analytics by businesses depends primarily on the type of data sets used. Business activity data and point-of-sale data are more frequently subject to data analytics, whereas online data including social media data and clickstream data are less frequently used among firms across the economy. According to the survey by the

Economist Intelligence Unit (2012), of more than 600 business executives around the world, two-thirds "say that the collection and analysis of data underpins their firm's business strategy and day-to-day decision-making". The respondents considered in particular "business activity data" as the most valuable data sets and in the case of the consumer goods and retail sector, "point-of-sale data" as well.

18 For instance, it is unclear whether those firms adopting DDI became more productive due to DDI-related investments or whether they were more productive in the first place. Furthermore, these studies rarely control for the possibility that some firms may have eventually seen a reduction in their productivity due to DDI, and as a result may have discontinued their investments in DDI.

19 As Mandel (2012) highlights: "[…] economic and regulatory policymakers around the world are not getting the data they need to understand the importance of data for the economy. Consider this: The Bureau of Economic Analysis […] will tell you how much Americans increased their consumption of jewelry and watches in 2011, but offers no information about the growing use of mobile apps or online tax preparation programs. Eurostat […] reports how much European businesses invested in buildings and equipment in 2010, but not how much those same businesses spent on consumer or business databases. And the World Trade Organization publishes figures on the flow of clothing from Asia to the United States, but no official agency tracks the very valuable flow of data back and forth across the Pacific".

20 Occupations were identified from the O*NET database, which provides ratings for hundreds of occupations in relation to many different features including working activities and the level and importance of those activities. Working activities considered for identifying potential occupations included: i) "getting information", ii) "processing information", and iii) "analysing data or information", with the level and importance of all three activities above the 75th percentile, and iv) "interacting with computers" at a level and importance below the 75th percentile. In the health care sector, potential occupations included, for instance, registered nurses, physicians and surgeons, and radiologists.

21 These estimated numbers should be taken with a great deal of caution, given that their underlying methodologies and data are not available.

22 Caution should be exercised when interpreting these results, as the methodologies used for these estimates are not available.

23 These initiatives are based on research results providing evidence of a link between real-world events and spikes in the volume of Twitter conversations related to food prices in Indonesia, illustrating the potential value of employing regular social media analysis for early warning and impact monitoring.

24 Crowdsourcing is "the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community" (Merriam-Webster, 2014).

25 Tsaboin also creates opportunities for others to innovate around the traffic data collected by making the data accessible to everyone.

26 The pilot of Cignifi's credit scoring platform in Brazil was conducted with data from pre-paid mobile customers located in the northeast of the country, one of the poorest regions, and provided evidence that the risk score calculated can be a significant

discriminator of default risk. There is great demand for such services in regions with very low levels of financial penetration.

27 Some major policy issues related to DDI have not been addressed in depth in this volume, in particular in respect to the complementary effects between data/analytics and the other types of KBC. This includes the role of intellectual property rights (IPR), which are highly relevant and often used as complementary assets to data and analytics, and sometimes therefore as strategic points for controlling DDI-related activities (Chapter 2). It is therefore recommended that this report be read in conjunction with OECD (2015a), which focuses on the second KBC pillar on IPR. The complementary effects in regard to the third KBC pillar, economic competencies, are highlighted in this report in terms of the skills and organisational change needed to realise the potential of DDI (see Chapter 6; see also Squicciarini and Le Mouel, 2012).

28 The data ecosystem relies on other key elements including technologies and actors that are rarely represented in a simplified model of the system. In particular, Figure 2.2 in the next chapter abstracts from the "cytoplasm" that lies between the layers of the data ecosystem and that enables the smooth interoperability of the different actors, their technologies, and services. These include (open) standards, some of which are related to application programming interfaces (APIs). Representations such as Figure 2.2 tend to be strongly biased toward the ICT sector, and do not sufficiently take into account other roles that are key to the functioning of the data ecosystem (e.g. legal consultants to address privacy risks). Ignoring these other actors will lead to systematic underestimation of the full size and impact of the data ecosystem.

29 Anderson (2008) has even gone so far as to challenge the usefulness of models in an age of massive data sets, arguing that with big data, machines can detect complex patterns and relationships that are invisible to researchers. The "data deluge", he concludes, makes the scientific method obsolete, because correlation is enough (Anderson, 2008; Bollier, 2010).

30 This property is at the source of significant spillovers which provide the major theoretical link to total factor productivity growth according to a number of scholars including Corrado et al. (2009).

31 See also OECD, 2015a according to which "[c]oncerns about government access requests – particularly to data entrusted to providers of cloud computing services – predate the revelations by Edward Snowden in 2013 and are not limited to intelligence gathering. But those revelations have brought into sharper focus the need for transparency. Today, Internet and communications businesses are under increasing pressure to be open about the manner in which they address government access requests."

32 A data breach is "a loss, unauthorised access to or disclosure of personal data as a result of a failure of the organisation to effectively safeguard the data" (OECD, 2012). Where the security breach of intellectual property does not involve personal data, the term "unauthorised access" will be used instead.

33 Duhigg (2012) describes the analysis process as follow: "[…] Lots of people buy lotion, but one of Pole's colleagues noticed that women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women

loaded up on supplements like calcium, magnesium and zinc. Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitisers and washcloths, it signals they could be getting close to their delivery date". As data analytics is not perfect, false positives are to be accounted for (see Harford, 2014). Target therefore mixes up its offers with coupons that are not specific to pregnancy (Piatetsky, 2014).

34 At the fourth meeting of the OECD Global Forum on the Knowledge Economy (GFKE) on "Data-driven Innovation for a Resilient Society" held in 2-3 October 2014 in Tokyo, Japan (www.gfke2014.jp/), Electronic Privacy Information Center (EPIC) President, Marc Rotenberg, highlighted the need for "algorithmic transparency", which would make public data processes that impact individuals (see Annex of Chapter 1 of this volume on the highlights of the *2014* GFKE*).*

35 Several authors including Ford (2009), Cowen (2013), Mayer-Schönberger and Cukier (2013), Frey and Osborne (2013), Levy and Murnane (2013), Brynjolfsson and McAfee (2014), Rifkin (2014) and Elliott (2014), have highlighted the potential negative implications of data-driven automation on wage and income inequalities.

36 Australia's National Digital Economy Strategy (NDES) foresees under its action item 12 the release of its national "Big Data strategy".

37 The national strategy presented in Japan's *Declaration to be the World's Most Advanced IT Nation* highlights the promotion of open and big data.

38 In 2013, the government of the United Kingdom published its "government strategy for how the United Kingdom can be at the forefront of extracting knowledge and value from data" (see www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf, accessed 12 May 2015).

39 In July 2014, the EC outlined its strategy Towards a Thriving Data-driven Economy in its Communication COM(2014) 442 final. The strategy aims at "supporting and accelerating the transition towards a data-driven economy in Europe".

40 Other countries have established, or are about to establish, sector- or domain-specific big data strategies. In 2012 for instance, the United States released its Big Data Research & Development Initiatives, which foresees investments worth USD 200 million in new R&D (see www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf, accessed 14 May 2015) as well as its strategic paper on "Big Data: Seizing Opportunities, Preserving Values" (EOP, 2014; see also PCAST, 2014). In 2012, Korea established its Big Data Master Plan, which promotes step-by-step big data use. This includes in particular the establishment of an infrastructure for big data sharing, and the provision of technical support and expert training.

41 In addressing privacy considerations, for instance, policymakers should seek to preserve the openness of the data ecosystem and the Internet.

42 The difficulties in transitioning from IPv4 standards towards the more efficient IPv6 are well known to policy makers (see OECD, 2014d).

References

Acatech (2013), "Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0", Final report of the Industrie 4.0 Working Group, April, www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Material_fuer_Sonderseiten/Industrie_4.0/Final_report__Industrie_4.0_accessible.pdf.

Adner, R. (2006), "Match your innovation strategy to your innovation ecosystem", *Harvard Business Review*, April, http://pds12.egloos.com/pds/200811/07/31/R0604Fp2.pdf, accessed 15 June 2015.

Allen, K. (2013), "Global financial crisis hit happiness and trust in governments – OECD", *The Guardian*, 5 November, www.theguardian.com/business/2013/nov/05/global-financial-crisis-happiness-trust-governments-oecd.

Anderson, C. (2012), "The man who makes the future: *Wired* icon Marc Andreessen", *Wired*, 24 April, www.wired.com/2012/04/ff_andreessen/5/, accessed 5 May 2015.

Anderson, C. (2008), "The end of theory: The data deluge makes the scientific method obsolete", Wired Magazine, 23 June, www.wired.com/science/discoveries/magazine/16-07/pb_theory/.

Bakhshi, H., A. Bravo-Biosca, and J. Mateos-Garcia, (2014), "Inside the datavores: Estimating the effect of data and online analytics on firm performance", Nesta, March, www.nesta.org.uk/sites/default/files/inside_the_datavores_technical_report.pdf, accessed 13 May 2015.

Bakhshi, H. and J. Mateos-Garcia (2012), "Rise of the datavores: How UK businesses analyse and use online data", Nesta, November, www.nesta.org.uk/sites/default/files/rise_of_the_datavores.pdf, accessed 13 May 2015.

Barua, A., D. Mani, R. Mukherjee (2013), „Impacts of effective data on business innovation and growth", Chapter Two of a three-part study, University of Texas at Austin, www.businesswire.com/news/home/20100927005388/en/Sybase-University-Texas-Study-Reveals-Incremental-Improvement, accessed 20 May 2015.

Berners-Lee, T. (2007), "Q&A with Tim Berners-Lee", Bloomberg Business, 9 April, www.bloomberg.com/bw/stories/2007-04-09/q-and-a-with-tim-berners-leebusinessweek-business-news-stock-market-and-financial-advice, accessed 4 May 2015.

BLS-OES (2014), Occupational Employment Statistics, US Bureau of Labor Statistics, November.

Big Data Startups (2013), "Walmart is making big data part of its DNA", www.bigdata-startups.com/BigData-startup/walmart-making-big-data-part-dna/, last accessed 22 August 2014.

BITKOM and Fraunhofer (2014), "Industrie 4.0 – Volkswirtschaftliches Potenzial für Deutschland", www.bitkom.org/files/documents/Studie_Industrie_4.0.pdf, accessed 15 June 2015.

Bollier, D. (2010), *The Promise and Peril of Big Data*, The Aspen Institute, Washington, DC.

Bruner, J. (2013), "Defining the industrial Internet" O'Reilly Radar, 11 January, http://radar.oreilly.com/2013/01/defining-the-industrial-internet.html.

Brynjolfsson, E. and A. McAfee (2014), *The Second Machine Age – Work, Progress and Prosperity in a Time of Brilliant Technologies*, W.W. Norton, New York.

Brynjolfsson, E., L.M. Hitt and H.H. Kim (2011), "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?", Social Science Research Network (SSRN), 22 April, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486.

Brynjolfsson, E. et al. (2008), "Scale without mass: Business process replication and industry dynamics", Harvard Business School Technology & Operations Mgt. Unit Research Paper No. 07-016, 20 September, http://dx.doi.org/10.2139/ssrn.980568.

CEBR (2012), "Data equity: Unlocking the value of big data", Centre for Economics and Business Research Ltd, London, www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf, accessed 15 April 2015.

CEPS (2014), "Shaping the integrated infrastructures of cities", presentation for the IEC CEPS workshop on "Orchestrating smart city efficiency", Centre for European Policy Studies, www.ceps.eu/system/files/article/2014/04/CEPS_IEC_Smart_City_3-Integrated_Infrastructures.pdf, accessed 15 May 2015.

Chick, S., S. Netessine and A. Huchzermeier (2014), "When big data meets manufacturing", *Knowledge*, INSEAD, 16 April, http://knowledge.insead.edu/operations-management/when-big-data-meets-manufacturing-3297, accessed 15 May 2015.

Christensen, C. M. (1997), *The Innovator's Dilemma*, Harvard Business School Press, Boston.

Corrado, C., C. Hulten and D. Sichel (2009), "Intangible capital and U.S. economic growth", Review of Income and Wealth, Series 55, No.3, September, www.conference-board.org/pdf_free/IntangibleCapital_USEconomy.pdf, accessed 15 May 2015.

Cowen, T. (2013), *Average is Over: Powering America Beyond the Age of the Great Stagnation*, Dutton Adult.

Croft, A. (2014), "Obama says U.S. needs to win back trust after NSA spying", *Reuters*, 25 March, www.reuters.com/article/2014/03/25/us-usa-security-obama-spying-idUSBREA2O18T20140325, accessed 15 May 2015.

Duhigg, C. (2012), "How companies learn your secrets", *New York Times*, 16 February, www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Dumbill, E. (2012), "Microsoft's Plan for Big Data", in *O'Reilly Planning for Big Data*, http://oreilly.com/data/radarreports/planning-for-big-data.csp, accessed 15 May 2015.

EC (2010), "Riding the wave: How Europe can gain from the rising tide of scientific data", Final report by the High-level Expert Group on Scientific Data, European Commission, October, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf, accessed 15 May 2015.

Economist Intelligence Unit (2014), "Networked manufacturing: The digital future", *Economist Intelligence Unit* sponsored by Siemens, 7 July, www.economistinsights.com/technology-innovation/analysis/networked-manufacturing.

Economist Intelligence Unit (2012), "The deciding factor: big data & decision making", *Economist Intelligence Unit* commissioned by Capgemini, 4 June, www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/.

Elliott, S. (2014), "Anticipating a Luddite revival", *Issues in Science and Technology*, Spring, pp. 27-37, http://issues.org/30-3/stuart/, accessed 25 May 2015.

EOP (2014), "Big Data: Seizing opportunities, preserving values", Executive Office of the President, United States, www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf, accessed 15 May 2015.

ESG (2012), "Boiling the ocean of control points in the Hadoop big data market", Enterprise Strategy Group, www.esg-global.com/blogs/boiling-the-ocean-of-control-points-in-the-hadoop-big-data-market/, accessed 24 May 2015.

Esmeijer, J., T. Bakker, and S. de Munck (2013), "Thriving and surviving in a data-driven society", TNO, 24 September, http://publications.tno.nl/publication/34610048/xcv74S/TNO-2013-R11427.pdf.

Ford, M. (2009), *The lights in the tunnel: Automation, accelerating technology and the economy of the future*, Acculant Publishing.

Frey, C.B. and M. Osborne (2013), *The Future of Employment: How Susceptible Are Jobs to Computerisation?*, Oxford Martin School, University of Oxford.

Frischmann, B.M. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press.

Frischmann, B.M., M.J. Madison, and K.J. Strandburg (2014), *Governing Knowledge Commons*, Oxford University Press.

Gerdon, S. and C. Reimsbach-Kounatze (2015), "Data-driven innovation for development: Unleashing 'big data' for inclusive growth", *OECD Digital Economy Papers*, OECD Publishing, Paris, forthcoming.

Guellec, D. and S. Wunsch-Vincent (2009), "Policy responses to the economic crisis: Investing in innovation for long-term growth", *OECD Digital Economy Papers*, No. 159, OECD Publishing, Paris, http://dx.doi.org/10.1787/222138024482.

Hardin, G. (1968), "The tragedy of the commons", Science (AAAS), Vol. 162, No. 3859, pp. 1243-1248, http://dx.doi.org/10.1126/science.162.3859.1243. PMID 5699198.

Harford, T. (2014), "Big data: Are we making a big mistake?", *Financial Times*, 28 March, www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html.

Hill, L. (2012), "How Target figured out a teen girl was pregnant before her father did", Forbes, 16 February, www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

IDC (2012), "Worldwide big data technology and services 2012-2015 forecast", IDC, March.

Jasperneite, J. (2012), "Was hinter Begriffen wie Industrie 4.0 steckt", computer-automation.de, 19 December, www.computer-automation.de/steuerungsebene/steuern-regeln/artikel/93559/0/.

Kramer, A.D.I., J.E. Guillory, and J.T. Hancock (2014), "Experimental evidence of massive-scale emotional contagion through social networks", Proceedings of the National Academy of Science of the United States of America (PNAS), Vol. 111, pp. 8788-8790, www.pnas.org/content/111/24/8788.full, accessed 2 June 2015.

Lane, J. et al. (eds.) (2014), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press.

Letouzé, E. and J. Jütting (2014), "Official statistics, big data and human development: Towards a new conceptual and operational approach", *Data-Pop Alliance White Papers Series*, 17 November, http://static1.squarespace.com/static/531a2b4be4b009ca7e474c05/t/546984d1e4b054b6f2656ac5/1416201425149/WhitePaperBigDataOffStatsNov17Draft.pdf.

Levy, F. and R.J. Murnane (2013), "Dancing with robots: Human skills for computerized work", Third Way, 1 June, http://dusp.mit.edu/uis/publication/dancing-robots-human-skills-computerized-work, accessed 2 June 2015.

Lodefalk, M. (2010), "Servicification of manufacturing - Evidence from Swedish firm and enterprise group level data", Working Papers No. 2010:3, Örebro University, School of Business, http://ideas.repec.org/p/hhs/oruesi/2010_003.html.

Mandel, M. (2013), "The data economy is much, much bigger than you (and the government) think", *The Atlantic*, 25 July, www.theatlantic.com/business/archive/2013/07/the-data-economy-is-much-much-bigger-than-you-and-the-government-think/278113/.

Mandel, M. (2012), "Beyond goods and services: The (unmeasured) rise of the data-driven economy", Progressive Policy Institute, 10 April, www.progressivepolicy.org/2012/10/beyond-goods-and-services-the-unmeasured-rise-of-the-data-driven-economy/.

Mayer-Schönberger, V. and K. Cukier (2013), *Big Data: A Revolution That Will Transform How We Live, Work and Think*, John Murray, London.

Merriam-Webster (2014), "Crowdsourcing", *Merriam-Webster.com*, www.merriam-webster.com/dictionary/crowdsourcing, accessed 24 September 2014.

McKinsey Global Institute [MGI] (2011), "Big data: The next frontier for innovation, competition and productivity", McKinsey & Company, June, www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx, accessed 15 June 2013.

MIC (2013), "Information and Communications in Japan", White Paper, Ministry of Internal Affairs and Communications, Japan.

Moore, F. (1993), "Predators and prey: A new ecology of competition", *Harvard Business Review*, May-June, http://blogs.law.harvard.edu/jim/files/2010/04/Predators-and-Prey.pdf, accessed 15 June 2013.

Naughton, J. (2015), "Don't trust your phone, don't trust your laptop – this is the reality that Snowden has shown us", *The Guardian*, Opinion, 8 March, www.theguardian.com/commentisfree/2015/mar/08/edward-snowden-trust-phone-laptop-sim-cards.

Nordregio (2014), "Urban planning and big data – Taking LUTi models to the next level?", www.nordregio.se/en/Metameny/Nordregio-News/2014/Planning-Tools-for-Urban-Sustainability/Reflection/, accessed 19 September 2014.

Noyes, K. (2014), "Cropping up on every farm: Big data technology", Fortune, 30 May, http://fortune.com/2014/05/30/cropping-up-on-every-farm-big-data-technology/.

OECD (2015a), *Digital Economy Outlook 2015*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264232440-en.

OECD (2015b), *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris, forthcoming.

OECD (2015c), *OECD Science, Technology and Industry Scoreboard 2015*, OECD Publishing, Paris, forthcoming.

OECD (2014a), 2014 Ministerial Council Statement, *Resilient Economies and Inclusive Societies – Empowering People for Jobs and Growth*, OECD Publishing, Paris, 07 May, www.oecd.org/mcm/2014-ministerial-council-statement.htm.

OECD (2014b), *Addressing the Tax Challenges of the Digital Economy*, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264218789-en.

OECD (2014c), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264221796-en.

OECD (2014d), "The economics of transition to Internet Protocol version 6 (IPv6)", *OECD Digital Economy Papers*, No. 244, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxt46d07bhc-en.

OECD (2013a), *Supporting Investment in Knowledge Capital, Investment and Innovation*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264193307-7-en.

OECD (2013b), "Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by 'big data'", *OECD Digital Economy Papers*, No. 222, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k47zw3fcp43-en.

OECD (2013c), "Strengthening health information infrastructure for health care quality governance: Good practices, new opportunities and data privacy protection challenges", OECD Health Policy Studies, OECD Publishing, Paris.

OECD (2013d), "Building blocks for smart networks", OECD Digital Economy Papers, No. 215, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k4dkhvnzv35-en.

OECD (2013e), Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, OECD Publishing, Paris, 11 July, www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

OECD (2013f), *How's Life? 2013*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264201392-en.

OECD (2012), *OECD Internet Economy Outlook 2012*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264086463-en.

OECD (2011), *Society at a Glance 2011*, OECD Publishing, Paris, http://dx.doi.org/10.1787/soc_glance-2011-en.

OECD (2008), Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information, OECD Publishing, Paris, 30 April, C(2008)36, www.oecd.org/internet/ieconomy/40826024.pdf.

OECD (2006), Recommendation of the Council concerning Principles and Guidelines for Access to Research Data from Public Funding, OECD Publishing, Paris, 14 December, www.oecd.org/sti/sci-tech/38500813.pdf.

OECD and Eurostat (2005), *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*, OECD Publishing, Paris.

OECD and FAO (2012), *OECD-FAO Agricultural Outlook 2012-2021*, OECD Publishing, Paris.

Orrick (2012), "The big data report", Orrick, www.cbinsights.com/big-data-report-orrick, accessed 19 May 2015.

Pariser, E. (2012), *The Filter Bubble: How the New Personalised Web is Changing What We Read and How We Think*, Penguin Books, April.

PCAST (2014), "Big data and privacy: A technological perspective", President's Council of Advisors on Science and Technology, United States, May, www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf, accessed 19 May 2015.

Piatetsky, G. (2014), "Did Target really predict a teen's pregnancy? The inside story", KDnuggets, 7 May, www.kdnuggets.com/2014/05/target-predict-teen-pregnancy-inside-story.html.

Pingdom (2013), "The top 100 web hosting countries", 14 March, available at http://royal.pingdom.com/2013/03/14/web-hosting-countries-2013/, accessed 19 May 2015.

Pole, A. (2010), "How Target gets the most out of its guest data to improve marketing ROI", Presentation at the 2010 PAW - the Predictive Analytics World Conference, 19-20 October, www.rmportal.performedia.com/node/1373, accessed 19 May 2015.

Reimsbach-Kounatze, C. (2015), "The proliferation of data and implications for official statistics and statistical agencies: A preliminary analysis", *OECD Digital Economy Papers*, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js7t9wqzvg8-en.

Rifkin, J. (2014), *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*, Palgrave Macmillan.

Rose, C. (1986). "The comedy of the commons: Custom, commerce, and inherently public property", The University of Chicago Law Review, pp. 711-81, http://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=2827&context=fss_papers, accessed 19 May 2015.

ScienceDaily (2013), "Big Data, for better or worse: 90% of world's data generated over last two years", 22 May, www.sciencedaily.com/releases/2013/05/130522085217.htm, accessed 11 April 2015.

Squicciarini, M. and M. Le Mouel (2012), "Defining and measuring investment in organisational capital: Using US microdata to develop a task-based approach", OECD Science, Technology and Industry Working Papers, No. 2012/5, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k92n2t3045b-en.

Stiglitz, J., P. Orszag and J. Orszag (2000), "Role of government in a digital age", Computer and Communications Industry Association, October, www.ccianet.org/CCIA/files/ccLibraryFiles/Filename/000000000086/govtcomp_report.pdf, accessed 10 October 2013.

Tambe, P. (2014), "Big data investment, skills, and firm value", Management Science, forthcoming, http://ssrn.com/abstract=2294077, accessed 10 June 2015.

The Economist (2012), "High-frequency trading: The fast and the furious", The Economist, 25 February, www.economist.com/node/21547988.

The Economist (2010a), "Data, data everywhere", 25 February, www.economist.com/node/15557443, accessed 10 June 2015.

UN Global Pulse (2014), "Mining Indonesian tweets to understand food price crises", United Nations Global Pulse, February, www.unglobalpulse.org/sites/default/files/Global-Pulse-Mining-Indonesian-Tweets-Food-Price-Crises%20copy.pdf, accessed 19 May 2015.

UN Global Pulse (2012), "Big data for development: Opportunities & challenges", United Nations Global Pulse, May, www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf, accessed 10 June 2015.

Vennewald, L. (2013), "A Quantum Leap Forward for Logistics Players", T-Systems Best Practices, October, www.t-systems.pt/umn/short-messages-about-customer-projects-innovations-and-solutions-for-cloud-computing-and-big-data-t-systems/1162170_1/blobBinary/Best-Practice_03-2013_News_EN.pdf, accessed 19 May 2015.

Wall, M. (2014), "Big data: Are you ready for blast-off?", BBC, 4 March, www.bbc.com/news/business-26383058, accessed 19 May 2015.

WEF (2012), "Big data, big impact: New possibilities for international development", World Economic Forum, www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf, accessed 6 June 2013.

Glossary

Knowledge is understood as *information and experience internalised or assimilated* through a process, commonly referred to as "learning". It provides the "learner" with the capacity to make effective decisions autonomously. Knowledge can be explicit, in which case it can be cost-effectively externalised to be communicated and embedded in tangible products, including books, standard procedures and intangible products such as patents, design and software. But it can also be tacit, based on an "amalgam of information and experience", which is too costly to codify and thus to externalise.

Information is often seen as the *meaning* resulting from the interpretation of facts as conveyed through *data* or other sources such as words. This meaning is reflected in the structure or organisation of the underlying source, including its hidden relationships and patterns of correlations, which can be revealed through *data analytics*. Information is therefore always context-dependent: it depends on the capacity to extract meaning from the information source; this capacity depending on available data analytic techniques and technologies as well as the skills and (pre-)knowledge of the data analyst.

Data are understood as the *representation of facts* stored or transmitted as qualified or quantified symbols. Data have no inherent meaning; however, they can be domain-specific. In contrast to knowledge and information, data are assumed to have an "objective existence", and they can be measured, namely in bits and bytes (see Table below). Data are typically gained from information when that information is *encoded* so it can be stored or communicated. Data can also be the result of *datafication*, a portmanteau for "data" and "quantification", where a phenomenon or object is transformed into quantified symbols. Datafication should not be confused with *digitisation*, which refers to the process of encoding information into *binary digits* (i.e. bits) so it can be processed by computers. Data that have not been digitised cannot be processed by computers.

Big data initially referred to data for which the i) *volume* became an issue in terms of data management and processing. Further definitions highlighted other important characteristics of "big data", such as ii) *velocity*, or the speed at which data are generated, accessed, processed and analysed (referring to real-time data), and iii) *variety* (referring to *unstructured* data and the capacity to link diverse data sets). These three properties – volume, velocity and variety – are therefore often considered to be the three main characteristics, and are commonly referred to as the three Vs, of big data. There is a major limitation with definitions based on the 3Vs, however: they are in continuous flux, as they describe technical properties that depend on the evolving state of the art in data storage and processing. Furthermore, these definitions misleadingly suggest that data are the main source of value. While it is true in the case of volume, what is behind variety and velocity is primarily *data analytics* – that is, the capacity to analyse unstructured diverse data in (close to) real time. Furthermore the term "big data" does not suggest how the data are used what type of innovation they can enable, or a how they relate to other concepts such as (e.g.) *open data*, *linked data*, and *data mashups*.

Units for measuring the volume of data

Unit	Size	What it means
Bit (B)	1 or 0	Short for "binary digit", after the binary code (1 or 0) computers uses to store and process data.
Byte (B)	8 bits	Enough information to create a number or an English letter in computer code. It is the basic unit of computing.
Kilobyte (KB)	1 000 B	From "thousand" in Greek. One page of typed text is 2 KB.
Megabyte (MB)	1 000 KB	From "large" in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4 MB.
Gigabyte (GB)	1 000 MB	From "giant" in Greek. A two-hour film can be compressed into 1-2GB.
Terabyte (TB)	1 000 GB	From "monster" in Greek. All the catalogued books in the US Library of Congress total around 15 TB.
Petabyte (PB)	1 000 TB	All letters delivered by America's postal service in 2011 will amount to around 5 PB; Google processes around 1 PB every hour.
Exabyte (EB)	1 000 PB	Equivalent to 10 billion copies of *The Economist.*
Zettabyte (ZB)	1 000 EB	The total amount of information in existence in 2011 was around 1.2 ZB.
Yottabyte (YB)	1 000 ZB	Currently too big to imagine.

Note: The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures. Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.

Source: Adopted from *The Economist (2010), "Data, data everywhere", The Economist, 25 February,* www.economist.com/node/15557443.

Structured data are data based on a predefined *data model* (i.e. an abstract representation of "real world" objects and phenomenon). Such models can be explicit, as in the case of a structured query language (SQL) database, where the data model is reflected in the structure of the database's tables. The data model can also be implicit, as in the case of *semi-structured data* (e.g. structured web content), where the underlying model can be made explicit at relatively low cost. In contrast, **unstructured data** are data that have no predefined data model and where such a model cannot be cost-effectively extracted. Typical examples include text-heavy data sets such as text documents and e-mails, as well as multimedia content such as videos, images and audio streams. The difference between structured, semi-structured, and unstructured data is becoming less important since with rising computing capacities, *data analytics* are increasingly able to automatically extract some structures embedded in unstructured data, including multimedia content.

Linked data typically refers to structured data that are published so that they can be interlinked. Data linkage is a means to contextualise data and thus enable the extraction of further information, which is greater than the sum of the information from the isolated **data silos**. The concept of linked data is closely related to the concept of open data, for which the full benefits can only be achieved if the isolated open data sets can be interlinked. Open standards play an important role in an interlinked data ecosystem.

Metadata are data about entities, including (**primary) data**. Metadata provide the necessary context without which the primary data cannot be accessed, linked, or fully understood. Metadata can be i) descriptive (based on attributes used to search and find an entity), ii) structural (describing the structure and organisation of an entity such as databases), and iii) administrative (providing information to help manage a resource). The concept of metadata is closely related to the concept of linked data, since metadata and primary data are by definition linked.

Personal data are defined by the OECD *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* as "any information relating to an identified or identifiable individual (data subject)". Any data that are not related to an identified or identifiable individual are therefore "non-personal" data. However, *data analytics* has made it easier to relate seemingly non-personal data to an identified or identifiable individual, thus blurring the boundaries between non-personal and personal data (see Chapter 5). It should be noted that the definition of personal data applied here does not distinguish between data (as inherently meaningless representation of facts) and information (as the *meaning* resulting from the interpretation of data). In other words, personal data and personal information are used as synonyms in this report.

Data can be **volunteered** when they are explicitly shared (by a data subject). Examples include creating a social network profile and entering credit card information for online purchases. They can be **observed** when it is captured by recording activities. In contrast to volunteered data where the data subject is actively and purposefully sharing its data, the role of the observed data subject is passive. Examples of observed data include location data of cellular mobile phones, and web usage behaviour. And finally, information can be **inferred** as the result of *data analytics*. Examples include credit scores calculated based on an individual's financial history. It is interesting to note that personal information can be "inferred" from several pieces of seemingly "anonymous" or "non-personal" data.

Public sector (government) data, in respect to the OECD *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* (PSI), are data generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for the government or public institutions (see Chapter 10). They are: i) dynamic and continually generated, ii) often directly produced by the public sector, or iii) associated with the functioning of the public sector (e.g. meteorological data, geo-spatial data, business statistics), and iv) often readily useable in commercial applications with relatively little transformation, as well as being the basis of extensive elaboration. Public sector data are a subset of PSI, which includes not only data but also *digital content*, such as text documents and multimedia files. The terms "public sector data" and "government data" are used as synonyms. The often used term "open government data" refers to public sector data made available as *open data*.

Open data does not describe a specific type of data. The key characteristic is the attribute "open", which specifies how access to data is *managed*, namely on *non-discriminatory terms* or "access on equal terms" as stated in the OECD *Recommendation of the Council on Principles and Guidelines for Access to Research Data from Public Funding*. In other words, data become "open" when access is not limited based on users' identity or intended use of the data (see Chapter 4). "Openness" should not be understood as a binary attribute but rather as a *continuum*, ranging from i) *closed* (with access only by e.g. the data controller or data subject), to ii) *commons* with possible restriction to a community (e.g. of researchers), to iii) *(unlimited) access granted to the public* as the highest degree of openness. Three key factors affect the degree of openness:

- technological design (including e.g. availability, machine readability and interoperability)
- intellectual property rights (IPRs) (including copyright as well as other IPRs applicable to databases and trade secrets)

- pricing, with marginal cost pricing being recommended by the OECD (2006) Council Recommendation on Access to Research Data from Public Funding and the OECD *Recommendation of the Council on Enhanced Access and More Effective Use of PSI*.

Data analytics refers to the set of techniques and tools used to extract information from data by revealing the context in which the data are embedded, their organisation and their structure. In the case of visual analytics the emphasis lies on data visualisation including (interactive) data exploration. Data analytics reveals the signal from the noise and with that the data's manifold hidden relations (patterns) including correlations, and interactions between facts, entities, and concepts. A number of terms are used (in this volume as synonyms) to refer to data analytics, some of which may include aspects that go beyond data analysis:

- **Data (text) mining** and **knowledge discovery** typically refer to data analysis but include aspects such as data pre-processing (cleaning), as well as model and inference considerations.
- **Profiling** is often used to describe the construction of profiles and the classification of entities in specific profiles.
- **Business intelligence**, a term that refers to the analysis of business-related data as often stored in databases (data warehouses) and mainly used for business reporting and monitoring purposes.
- **Machine** or **statistical learning** is a subfield in computer science, and more specifically artificial intelligence (AI), concerned with the design, development and use of data analytic algorithms that allow computers to "learn" – that is, to improve performance with every data set analysed.

Data-Driven Innovation

BIG DATA FOR GROWTH AND WELL-BEING

Contents

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-22934-1
93 2015 02 1 P