

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Introduction: Opening Up Data in Scientific Research

Vera J. Lipton

“Imagine a world in which every single human being can freely share in the sum of all knowledge.”

Wikimedia Foundation

1.1 Why open access to scientific outcomes matters

The need to effectively disseminate and share science outcomes is pressing. As nearly every region feels the effects of climate change, as food insecurity is rising, and as the demand for natural resources is increasing, the world looks to science for solutions. In this interconnected globe where over 50% of its population can access the Internet,¹ science offers hope. It offers hope for those living in prosperous societies and for the remaining half of the world—over 3 billion people—who live on less than US\$2.50 a day.²

In this age of communication, if anything is to secure the future of our planet and the well-being of its civilisation, then it is likely to be science, as Australian science commentator Julian Cribb noted. Yet it will not be science alone, rather the knowledge that it imparts and the learning that it yields when it is shared broadly and applied wisely. If science is to deliver its full value to society, it must be easily and freely accessible [5].

But the majority of science is not accessible easily, and only a fraction of it is accessible freely. This is despite the fact that scientific knowledge is plentiful and is growing rapidly—doubling, on average, every 15 years.³ Much of the knowledge and research data underpinning science remains guarded by elites; much of it stays locked in institutional repositories or costly scientific journals.⁴

The low accessibility and subsequent uptake of scientific knowledge are not ideal for researchers who produce and use science and who are unable to access in a timely manner the scientific outputs produced by their peers.⁵ The situation is also not ideal for universities and public research organisations that train and employ researchers or

¹ Statistics sourced from Internet Source Stats [3].

² Statistics sourced from Shah [4].

³ See Larsen and von Ins [6]. The rate of doubling of the body of scientific knowledge was calculated as an average number of scientific records included in the following databases: Web of Science (owned by Thomson Reuters), Scopus (owned by LexisNexis), and Google Scholar. Duplicate entries were removed.

⁴ In a consultation carried out in 2012, the European Commission determined that there were huge barriers to accessing research data. Of the 1140 subjects questioned, 87% contradicted the statement that there was no access problem to research data in Europe. See European Commission [7]. See also McCain [8].

⁵ In a study conducted in 2011 by Tenopir et al. [9], 67% of 1300 researchers pointed to a lack of access to data generated by other researchers or institutions. See also point 4 above.

for governments that fund the majority of basic research.⁶ Equally, the low availability of scientific knowledge is not ideal for taxpayers and citizens whose hope for improved living standards increasingly depends on the development and application of science and technology. And it is not an ideal scenario for technology companies that require timely access to knowledge as they increasingly innovate by combining research outputs from external and internal sources [11, 12]. Nobody benefits from science that stays limited to those who initially create it. Such science is lost—lost to follow-on innovation and lost to society at large.

Fortunately, there are ways to increase access to scientific knowledge and opportunities to accelerate scientific discovery.

The open access movement has developed over the past two decades. It advocates the sharing of scientific knowledge over the Internet by challenging the application of exclusive property rights over scientific outputs.⁷ This movement further promotes ‘digital openness’ in the conduct of science and of scientific communication, facilitating online access to scientific publications and the underlying research data. The open scientific content that results from this movement and emerging online communities are shaping the fundamental processes of science creation and dissemination. These processes are taking place alongside—and are intimately connected with—the evolution of digital technologies and interactive communications. Open science is developing and building upon the body of digital knowledge, data and infrastructure that it inherently generates.

More recently, an increasing number of governments, research funders and publishers have mandated the release of research data over the Internet with a view to making scientific results more easily and more broadly available for research, innovation, education, technology development and other applications. In 2010, the National Science Foundation in the United States announced that grant proposals would require a data management plan and that the plan would be subject to peer review [16]. The policy was a tipping point leading to similar mandates emerging in other nations.

In 2011, the Research Councils in the United Kingdom (RCUK)⁸ released *Common Principles on Data Policy* [18] and, subsequently, many RCUK funders have mandated the requirements for data management plans with new grant applications. The principles encourage research data to be made openly available with as few restrictions as possible in a timely and responsible manner.⁹ Similarly, the *Recommendation of the European Commission on Access to and Preservation of Scientific Information* [19] encouraged the European Union member states to develop policies for open access to scientific results, including research data and information [20]. This was followed with the Open Research Data Pilot in 2014, aimed at exploring the digital sharing of research data resulting from the Horizon 2020 research grants [21].

The Australian Government was among the first to invest in the development of research data infrastructures. The Australian National Data Service (ANDS) was

⁶ According to the OECD, industry in 2014 funded just over 1% of gross domestic expenditure on basic research in Australia, Austria, Belgium, Denmark, France, Germany, Sweden, and the United States. The figure was less than 1% in Canada, Chile, Czech Republic, Greece, Hungary, Italy, New Zealand, and the United Kingdom. Source: OECD Main Science and Technology Indicators [10].

⁷ The term ‘digital science’ is often referred to as ‘open science’ or ‘Science 2.0’. Its roots go back to the emergence of the Internet and communication technologies. In general terms, open science refers to changing scientific practice based on cooperative work facilitated by diffusing knowledge by using digital technologies. See, for example, European Commission [13]; Fecher and Friesike [14]. The concept of digital science is discussed in Chapter 2, Section 2.2 of this study. For a good overview of the evolution of open access movement, see Suber [15].

⁸ Integrated in April 2018 into a new body [17].

⁹ *Ibid.*, point 2.

established in 2008 to develop an Australian Research Data Commons platform [22]—an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions. Funding was also allocated for the development of metadata tools through the ‘Seeding the Commons’ initiative¹⁰. Open research data is a priority area for the Data to Decisions Cooperative Research Centre established in July 2014.¹¹ The two principal research funders in Australia—the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC)—both ‘strongly encourage’ the recipients of their grants to share data and metadata arising from their research.¹²

Many private research funders also require the public release of data resulting from the research they fund.¹³ The open data mandates introduced by governments and research funders follow similar policies promoting open access to publications¹⁴ and public sector information.¹⁵ The mandates were first introduced in the United States, Europe and Australia and quickly spread to other parts of the world. The sharing of scientific data in electronic formats has a long tradition in medical research, biotechnology and geospatial sciences. Within a span of 6 years, the policies mandating open access to scientific data have expanded to all fields of research, including humanities and social sciences.

At a multilateral level, the United Nations Educational, Scientific and Cultural Organization (UNESCO) adopted the *Revised Draft Strategy on UNESCO’s Contribution to the Promotion of Open Access to Scientific Information and Research* in 2011 [29]. The strategy was established to promote open access to scientific information and research¹⁶, and it called for examining the feasibility of developing a UNESCO convention on open access to scientific information and research.¹⁷ In recent years, the UNESCO has viewed open access to research results as a prerequisite for reaching its Sustainable Development Goals.¹⁸ As such, the UNESCO believes that open science has a fundamental role in supporting poverty reduction. The organisation is committed to making open access to research one of its central supporting agendas.¹⁹

The open data policies vary in their scope and among research funders, yet they share some common objectives—to advance science by making research data available to others more quickly and more broadly; to enable reproducibility of scientific outcomes; and to increase the uptake, use and quality of scientific knowledge, including in developing countries. Such arguments reflect the desire to tackle some of the big challenges facing humanity and the planet today. Building upon and reusing open scientific knowledge can expedite these global efforts [31].

¹⁰ Seeding the Commons was a programme funded by the Australian National Data Service to improve discoverability and use of university research data. Full description of associated university projects is available at: <https://projects.ands.org.au/getAllProjects.php?start=sc>.

¹¹ The centre brings together researchers and industry to contribute to the development of Australia’s big data capability. See Data to Decisions CRC [23].

¹² The National Health and Medical Research Council (NHMRC)’s [24]. *Open Access Policy* (previously also referred to as the *NHMRC Policy on the Dissemination of Research*) took effect from 15 January 2018, 7.

¹³ One of the earliest advocates of open research data was the Wellcome Trust. See Wellcome Trust [25, 26].

¹⁴ For a good overview of open access to publications, see Swam [27].

¹⁵ For a good overview of OA policies for public sector information, see Fitzgerald [28].

¹⁶ *Ibid*, Par. 1.1.

¹⁷ *Ibid*, Annex 3.

¹⁸ The UNESCO states at least 10 out of the 17 Sustainable Development Goals comprising the 2030 Agenda for Sustainable Development require constant scientific input. See UNESCO [30].

¹⁹ *Ibid*.

1.2 The early challenges facing open scientific data

The arguments put forward for open scientific data certainly are plausible. At the same time, understanding the requirements for responsible data sharing and ensuring compliance with these requirements pose fresh challenges to research organisations. Indeed, making research data available, legible and useful to unknown audiences, and for unanticipated purposes, may not be an easy task. Maintaining the privacy of subjects involved in data collection, particularly in clinical trials, is an additional concern for medical research institutes. Furthermore, digital curation of research data requires substantial investments in data infrastructures, human resources and new business models. Many research organisations point out that open scientific data cannot be covered from research budgets. Further still, the very nature of scientific research is changing profoundly as open scientific data is increasingly being curated and shared. In the face of these challenges, some stakeholders feel that the future of open scientific data is uncertain.

This book argues, however, that these challenges help us understand how best to achieve open access to scientific data. Open scientific data is not only desirable, it is possible. Yet it requires careful balancing of the needs of all stakeholders and especially the need to balance ‘collective benefits’ with ‘individual responsibilities’. The ‘collective benefits’ are likely to accrue from the provision of open data to society, while ‘individual responsibilities’ for curating open data and developing supporting infrastructures are vested in researchers and their organisations. Those tensions between ‘individual responsibilities’ and ‘collective benefits’ form an overarching theme of this book. I investigate the ways in which the two concepts are merged and confused and what might be done to clarify them.

Ultimately, it is argued that the responsibility for open data cannot be placed on researchers if their efforts for data curation are not recognised and rewarded and if open data cannot be reused successfully. I suggest that ‘collective societal benefit’ and ‘individual responsibility’ would be best balanced within a staged model for releasing scientific data as open data.

The proposed model is researcher-centric and puts a major emphasis on data quality, which is the key prerequisite for data reuse. The model rests on the observation that even organisations fully committed to openness, such as the European Organization for Nuclear Research (CERN), are unable to share all of their research data as open data at this point.

However, CERN has developed a useful classification model for research data based on the levels of data granularity and processing. I adopt and slightly adjust the CERN classifications to propose the stages of open data release for all research organisations and across all scientific disciplines. I specifically acknowledge that the definitions of ‘research data’ and the timing of data release will vary not only among research disciplines but even among the defined data stages and among individual research projects. This reasoning is consistent with a major finding of this book that the opening up of research data requires an open mindset while acknowledging that ‘one size does not fit all’. The open data mindset finds the practice of open scientific data to be a diverse, ongoing and ever-evolving process that is as important a driver of research practice as are the scientific results and the data that underpin them.

Central to the proposed model is an understanding of the social context in which scientific knowledge is created and used. Historically, science has had a role in creating data and assessing validity of the data. As science develops and changes over time, the meaning and relevance of data also change.

This study finds that the nature, dissemination and use of scientific knowledge are profoundly changing in the context of the digital revolution. However, the core theories of knowledge production and dissemination in the digital age—namely, the theories of a knowledge society²⁰—envisage that merely releasing scientific data into the public domain is sufficient for the economic and social benefits of open data to accrue.

The model proposed in this book rebuts this argument, positing that simply providing *access* to data in the public domain is useless to society and that only data *reuse* can realise the envisaged benefits.

At present, the scientific community is the sole community sector that appears capable of the competent reuse of scientific data. Consequently, it is argued the rationale for open scientific data should be narrowed to focus on data-enabled science, rather than data-enabled society. This argument rests on the fact that science is a form of social organisation in its own right [34, 35]. Therefore, the first task is to facilitate improved data sharing and reuse of open scientific data among relevant researchers. Once researchers embrace open scientific data and learn how to better describe and embed the data into their daily research practice, only then the benefits can spill over to a wider society.

At the same time, open scientific data is profoundly changing the economic context in which power and control over science are distributed in society. In today's world, where knowledge means power²¹ and where science regulates cutting-edge knowledge,²² the control of science is also becoming more important. Open access is a response to a trend towards the commodification of knowledge.

Looking at different scientific disciplines, we see that commodification is pervasive in engineering, in the biological and medical sciences, and—on a somewhat smaller scale—in the physical sciences [37]. Although these trends have roots in policy changes in intellectual property and the economics of information, critical data mass has led to new markets. Some of the data can be exchanged only within academia, but some have commercial value and can lead to new partnerships with industry, just as commercial data can lead to academic research. However, such data exchanges lead to new tensions [38–40].

Open scientific data empowers researchers, not markets, to control scientific knowledge into the future. Open scientific data thus leads towards a more transparent and accountable governance of science that, in turn, advances a more open, collaborative and democratic society.

To summarise how things stand today—while attempts at sharing research data in electronic formats go back to the late 1950s—rapid and pervasive technological changes have enabled the storing, processing and transmitting of large volumes of data and have stimulated collaboration among scientists. The recent and successful practice of enabling open access to scientific publications and government data

²⁰ There is a range of definitions of the term 'knowledge society' but, broadly speaking, a knowledge society is one that generates, processes, shares, and makes knowledge that may be used to improve the human condition available to all its members [32]. Also in 1994, Gibbons et al. [33] examined changes in forms of scientific knowledge production. This is one of the key theories examined in this study.

²¹ Castells argues that the rise of networks that link people, institutions, and countries characterise contemporary society. The purpose of these networks is for information to flow in what Castells defines as an 'informationalised society'—one in which 'information generation, processing, and transmission become the fundamental sources of power and productivity.' See Castells [36].

²² Nowotny and others ([33], pp. 12–13) observed that specialised knowledge plays a crucial role in many dynamic markets. Specialised knowledge is an important source of created comparative advantage for both its producers and users of all kinds, and not only in industry. As a result, the demand for specialist knowledge is increasing.

(public sector information) brought a new impetus to digital research data and has inspired new methods for scientific research in online spaces. Such data-led science holds promise for the development of new scientific knowledge and is transforming the conduct of science and the communication of scientific outcomes.

1.3 Exploring a way forward

This book investigates how the open data policies recently introduced by research funders are being implemented in practice. Drawing on early experiences with open data at CERN and experiences with data resulting from clinical trials—two scientific fields in which the sharing of research data in digital formats is already well-established—this study aims to determine whether open data policies can achieve open access to scientific data. More specifically, the principal goal of this study is to investigate optimal ways to tackle the challenges associated with the practical implementation of open scientific data.

In the context of this study, ‘open scientific data’ refers to the evidence that underpins scientific knowledge produced by publicly funded organisations²³ and that meets the FAIR standards for openness—data that are findable, accessible, interoperable and reusable.²⁴

This book contributes to the ongoing discussion and considers four principal research questions.

Firstly, I examine the objectives and benefits stated for open scientific data.

Secondly, I analyse the open data policies and seek to establish whether open scientific data is an achievable objective.

Thirdly, I ask how selected data-centric public research organisations are implementing open data in practice. Specifically, in this context, I enquire about the legal and other challenges emerging in the process of open data implementation and investigate how data-centric research organisations are dealing with these challenges.

Finally, I seek to establish what can be done to promote open access to data and whether the open data mandates need to be revised.

For the purposes of clarity, the research questions are summarised as follows:

1. *Vision*. What are the expected benefits associated with the curation and release of scientific data?
2. *Policy*. What is the scope of the open data policies?
3. *Practice*. How are selected data-centric public research organisations implementing open data? What are the legal and other challenges emerging in the process of implementation? Is open scientific data an achievable objective?
4. *A way forward*. What can be done to promote open access to scientific data across different research disciplines? Is there a need to revise the open data mandates?

The research questions are answered in the specific chapters of the book as depicted in **Figure 1**.

In answering the stated research questions, I provide a theoretical framework based on the social theories of innovation—especially the models of science and

²³ The reasons for this focus on publicly-funded organisations are discussed at 1.4: ‘Matters beyond the scope of this study’.

²⁴ See Wilkinson et al. [31] at point 27. The meaning of ‘open scientific data’ and the parameters of ‘openness’ are discussed in detail in Chapter 4 and described in **Table 1**.

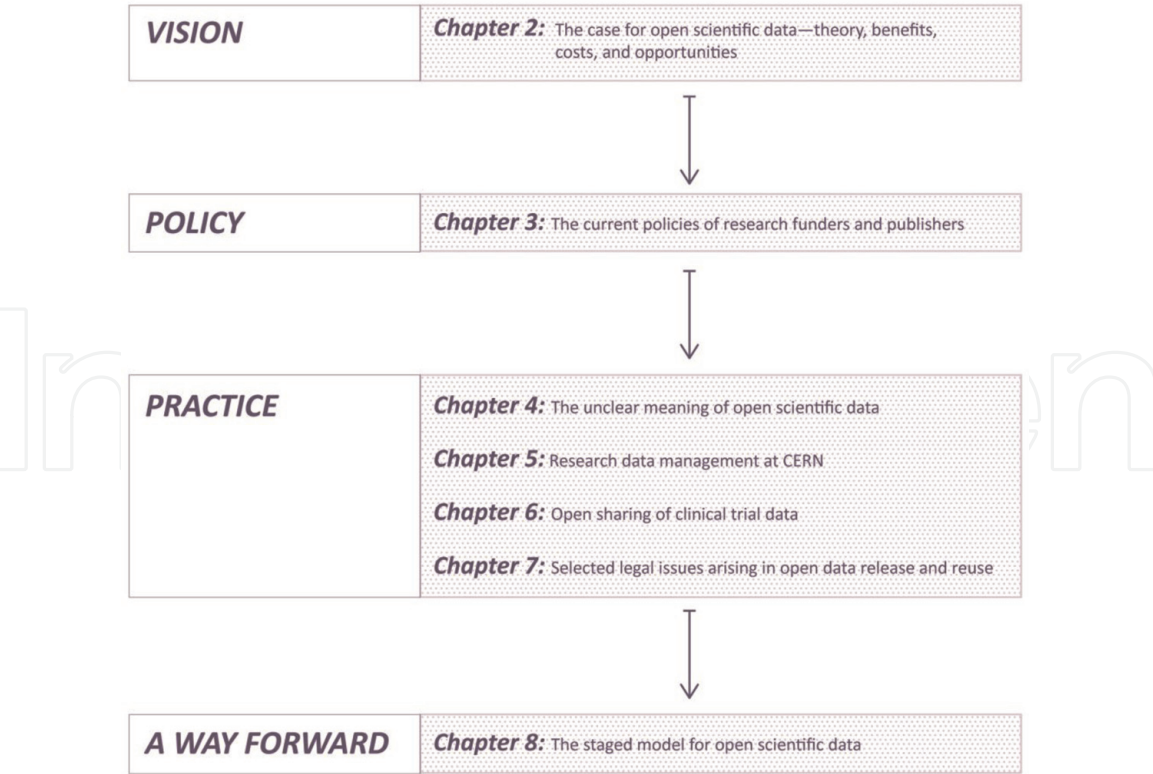


Figure 1.
Structure of this book.

knowledge production and how these are changing in the context of the digital revolution. I place a special emphasis on the reusability of open research data and the reproducibility of research findings, which are the primary stated objectives of open scientific data. I then analyse the open data mandates and the requirements they place on researchers and research organisations. I further discuss and conceptualise the meaning of ‘open scientific data’ and examine how stakeholders understand the term.

This inquiry helps to identify the many facets of open scientific data and the lack of clarity among stakeholders regarding the meaning of ‘research data’ and associated terms such as ‘data use’ and ‘data reuse’. The meaning of ‘data’ is also discussed in the context of copyright law and the parameters that ‘data’, ‘datasets’ and ‘databases’ need to meet to be protected by copyright. The discussion informs the proposed staged model for open scientific data, but that model is also largely informed by the experiences with research data management at CERN and with clinical trial data.

The purpose of the present study is to bring together the existing research into open scientific data, its practice and the challenges that have occurred in its implementation over the past decade. In this regard, it is a survey of theoretical work, policy and legal documents and practical experiences with open data in two scientific fields.

Specifically, this study documents the implementation of the open data mandates recently introduced by research funders and publishers. Further, it evaluates the potential of the mandates for driving open scientific data into the future.

1.4 What this book contributes

Open scientific data is a recent and vast subject—spanning many scientific and scholarly disciplines and various types of research data. To date, all efforts to study open scientific data have been piecemeal—focusing on specific initiatives, successful case studies, desired objectives and the envisaged benefits of open science, the

evolving parameters for data openness and factors that may motivate researchers to release their research data to other users. Such issues have been assessed through the lenses of specific research disciplines or data initiatives or through individual technical issues that were deemed necessary to successfully implement open data. Scholarly research that is more systematic and that brings together the practice of open scientific data across various scientific disciplines is in short supply.

At the same time, research funders, government officials and librarians tend to approach the treatment of open scientific data with the same methods that proved successful in the implementation of open access to scientific publications. Such established approaches have been embedded in open data mandates. However, it is not clear, at this point, whether the current open data mandates can achieve the desired objectives.

There are significant differences evident in the implementation of open access processes. While the implementation of open publication mandates has been achieved within a few years, implementing open access to research data does not appear to be as straightforward. The rates of data deposit in online repositories remain low, and the implementation of the mandates in practice is lagging, as documented in Chapter 6 of this book. Some of the impediments to data sharing are known but have not been systematically studied to date. Other challenges have only become apparent as research organisations have begun the process of implementing the mandates. These last challenges are less known and are not well-documented. Apart from its newness as a concept, open data is proving to hold far more complexity in its execution than does the practice of open scientific publications.

Examining the challenges faced by open scientific data practice is a much-needed contribution to the ongoing debate. This book looks in detail at challenges in policy, data management and legal administration. It draws upon research and experiences with open data at CERN and in clinical trial data—two data-intensive fields at the frontline of the debate regarding open data. It is hoped that lessons learnt in these fields may help other research disciplines to adopt open data policies.

This investigation is significant because many key stakeholders are not aware of the challenges in implementing open data. And even when they are interested in applying the principles of it, many smaller research organisations appear to struggle with the concepts and definitions of open research data and its management, curation and funding. Some organisations yet to embrace open data are questioning not only the costs but whether the broad mandates for open data are fit for purpose. At the same time, while many research funders have introduced open data mandates, others appear to be backtracking from earlier commitments to do the same.

With regard to stakeholders, this book focuses mainly on those matters predominantly raised by researchers and others who execute open data mandates in research practice. Their voices are important and need to be heard by those who believe, mistakenly, that open scientific data requires nothing but the publishing of research data in public repositories.

In this uncertainty, it is hoped that a coherent exploration of the early experiences and challenges with open data in one of the largest data-centric research organisations in the world can bring fresh insights and unearth areas of best practice that, together, may help refine approaches to open data mandates.

1.5 The research methodology

The research in this book is interdisciplinary due to the diverse nature of open scientific data ([41], p. 2). The concept incorporates information, perspectives, concepts, practices and theories from several fields—science, public policy, science

policy, law, data management, information technology, scientific communications and library scholarship, among others.

This book focuses on the policy, law and data management practice aspects of open scientific data.

The primary approach adopted in the present study is the problem–solution method. Firstly, I identify and analyse the relevant policies, theoretical concepts and international legal mechanisms adopted in support of open scientific data (Chapters 2 and 3). This is followed by a detailed examination of data management practices and by identifying issues that arose in putting into practice the adopted instrument (Chapters 4–7). The final two chapters propose a solution in the form of a staged model for open scientific data that addresses the issues identified in the preceding chapters. The proposed model, along with eight recommendations, presents a roadmap towards more achievable and sustainable open scientific data.

The main objective of the above methodological framework was to develop a greater degree of shared understanding of the legal, policy and conceptual framework that is appropriate for the current level of technological development; the experience of researchers with the digital sharing of scientific outputs; and the proclaimed social and policy objectives of open science.

This study used both doctrinal and empirical research.

Doctrinal research involved analysis of both legal and non-legal documents—including international declarations and instruments guiding open scientific data; the open data policies of research funders and publishers; statutes and case law governing the notion of data; copyright in data; ownership of data; confidentiality; and privacy of the research subjects. This analysis is supported by secondary sources such as monographs, peer-reviewed articles and reports underpinning the benefits of sharing data and scientific knowledge in the context of the digital revolution.

This research, to make the claim impersonal research goes beyond legal scholarship in that it considers the theories of science and knowledge production and the social functions of science. Historical records regarding the emergence of early data-sharing initiatives and projects have provided the context and background. This combination of multiple literature sources provides a more comprehensive and insightful discussion of the challenges arising in the implementation of open scientific data.

The staged model for open scientific data presented in this book was developed with a view to presenting a pragmatic and achievable approach to open scientific data—taking into account resource and technology constraints, established culture, current research practice and legal impediments to open data. The proposed model is less ambitious in its scope than the current open data mandates of research funders and publishers. Yet, if adopted, the proposed model would provide a basis for improved online data sharing among both researchers and non-researchers and also serve as a springboard for realising the vision for improved access to research data across all scientific disciplines.

1.6 Matters beyond the scope of this book

This book is not an exhaustive review of the current practices in making available open scientific data. It examines specific developments in the area of research data mandates and how these are implemented. To keep within reasonable bounds and for the sake of cohesion, the research in this study focuses on open scientific data as implemented by public research organisations that conduct applied and basic scientific research, including medical research institutes. In this

context, a research organisation is considered to be publicly funded if it receives more than 50% of its income from public sources. The primary focus is on research institutes, and not on universities, mainly because universities are relatively late entrants and have far less experience with research data management and data sharing than data-intensive research organisations.

This book is not intended to cover the experiences with data sharing in the private sector. However, I draw on certain experiences of private sector organisations in regulating access to data from clinical trials to examine how they manage the ethical, privacy and research integrity issues arising in data sharing in digital formats.

While open data now covers all domains of research, including humanities and, to some extent, social sciences, this study only considers experiences with open data in the STEM disciplines—science, technology, engineering and medicine. Specifically, this book focuses on particle physics and clinical trial data and, to a lesser extent, geospatial sciences.

Open scientific data spans many jurisdictions, with rules emanating from different sources of law. As such, open data can be governed by various regimes. This book does not study the transnational operability of open data. However, Chapter 7 considers some cross-jurisdictional issues arising in the reuse and mining of open scientific data across national boundaries. The legal definitions of ‘data’ and ‘databases’ are analysed in selected English-speaking jurisdictions, including Australia and the United States, and within the context of the European Union’s General Data Protection Regulation.²⁵

Finally, there are many issues arising in the context of open scientific data, and every one of them can probably lead to a separate monograph. This book covers only selected issues associated with open scientific data, focusing on research data mandates and their implementation, especially with regard to research data management. The study also focuses on the legal issues arising in the release and reuse of open data.

The technical parameters and infrastructures that make open data findable, accessible, interoperable and reusable are not substantially covered in this book.

Nor is the question of the economics of open scientific data covered extensively, even though recent research on this matter is touched upon in this study. The economics of open data is a nascent area and the relevant economic models for quantifying the components of open data infrastructures and the methods for evaluating the benefits and costs of open scientific data are just starting to emerge.

The legal and other developments discussed in this book are those available to me as at 1 December 2017, but significant changes that have occurred after this date have been included where possible.

1.7 The structure of this book

This book consists of nine chapters, including *Introduction* (Chapter 1) and *Conclusion* (Chapter 9).

²⁵ Regulation (EU) 2016/679 of the European Parliament of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal* L 119/2, 04/052015, 1–88. The European data protection framework is complemented by Directive 2002/58/EC on privacy and electronic communication [42–44].

Chapter 2. The Case for Open Scientific Data: Theory, Benefits, Costs and Opportunities

Chapter 2 examines why the idea of open data has been taken up by research funders, research organisations, the broader scientific community and civil society. It first considers whether the activities of these actors constitute a social movement seeking to mobilise open scientific data and whether the move towards openness in science is fostering a transition to a knowledge economy. In order to assess such broad questions, this chapter reviews the theories underpinning open knowledge-based and knowledge-based society as well as the role of open access in fostering the dissemination and reuse of scientific data. It also discusses the ways in which scientific knowledge has been produced historically and how it is produced today.

Chapter 3. The Current Policies of Research Funders and Publishers

This chapter analyses the policies mandating open access to scientific data. These policies are primarily focused on research funders and vary in their scope and format, which make any comparative analysis difficult. None of the policies under evaluation are more than 7 or 8 years old. Despite these limitations, the preliminary analyses are positive. The chapter finds that in addition to early entrants—that is, the English-speaking world—more than one third of the 28 European Union members had mandates for open scientific data in place at the end of 2017. Many Latin American countries have also adopted open access to research data, and similar mandates were under consideration in countries with large scientific output such as China. Some European countries—notably Italy, Spain, Germany and the Netherlands—have legislated open access to scientific data.

Chapter 4. The Unclear Meaning of Open Scientific Data

The problematic notion of ‘data’ is further examined in Chapter 4. Each term such as ‘open data’ and ‘research data’ has multiple meanings in the open data mandates and practice. Key concepts are often conflated or used interchangeably. ‘Data’ mean different things to different stakeholders, all at the same time, this chapter concludes. Researchers often like to share ‘datasets’, another contested term. Settling on criteria that define ‘data’ raises more questions. Chapter 4 provides background on each of these terms although every one of them warrants a lengthy study.

Chapter 5. Research Data Management at CERN

This key chapter deals with some of the evolving aspects of research data management. It examines data-driven experiments at the European Organization for Nuclear Research (CERN), acknowledging that it is not feasible to address, within the purview of a single chapter, all unfolding issues associated with the curation and reuse of open scientific data. The experiences with open data at CERN demonstrate efforts to reconcile the interests of all relevant parties. A major finding is that the data management process is a continuously evolving process and that this and the thinking around data preservation, curation and open sharing have proved to be challenging in most organisations, including CERN.

Finally, Chapter 5 explores efforts to define data in operational terms, based on levels of data processing and granularity and concludes by offering a working definition of the stages of open scientific data that is further canvassed in Chapter 6 and then embedded in the staged model proposed in Chapter 8.

Chapter 6. Open Sharing of Clinical Trial Data

Chapter 6 outlines the research data management process in the context of sharing data that results from clinical trials. The principal focus is on the specific

protocols for data sharing—particularly data quality protocols at various levels of data processing, the risks in sharing data and the approaches used to mitigate the risks. This chapter argues that the risks arising from data sharing (as opposed to non-sharing) can be addressed through controls of data access. This exploration reveals that the application of soft tools, such as professional code of conduct and data use agreements, can be a means for reducing concerns that create disincentives for sharing clinical trial data as open data. Finally, Chapter 6 considers the motivation of researchers to share their own data and their willingness to reuse data produced by others.

Chapter 7. Key Legal Issues Arising in Open Data Release and Reuse

This chapter discusses the legal issues arising at two critical stages, namely, data release and data reuse.

The first part examines the legal issues arising in data release. The focus is on intellectual property rights, especially copyright in data and databases. This is followed by consideration of the uncertainty around data ownership—identified as the cause of subsequent problems affecting data licencing—along with a potential lack of interoperability and unclear conditions governing data reuse. There is an examination of some relevant licencing issues.

The second part is dedicated to analysis of different types of data reuse, such as linking or mining, and whether these types of reuse can infringe copyright. Other issues that need specific attention in the reuse of open scientific data include ensuring the privacy of research subjects, the ethics of research, and managing the risks associated with possible disclosure of confidential information.

Chapter 8. The Staged Model for Open Scientific Data

Drawing on the findings of the three preceding chapters, Chapter 8 proposes a model to address problems arising in their implementation.

This chapter consists of three main parts. It first outlines the policy setting within which the policies mandating open access to scientific data have emerged. This is followed by an overview of the main features of the mandates and identification of their drawbacks. The final section discusses the shortcomings in more detail and introduces a staged model for open scientific data, along with eight recommendations.

It is argued that the open data mandates have created a momentum for data release globally. At the same time, the mandates alone are insufficient to effectively drive open data into the future because the digital curation of research data for public release poses many challenges. The open data mandates, as they stand today, fail to acknowledge and address these challenges and thus should be revised. The recommendations in the proposed model suggest options for dealing with the issues arising in implementation so as to ensure sustainability of open scientific data into the future.

Chapter 9. Conclusion: Towards Achievable and Sustainable Open Scientific Data

The final chapter answers the research questions posed in this study. It summarises the core findings and contributions. The chapter concludes with a call to revise the open data mandates and so initiate changes in data management practices across different scientific disciplines to move towards the staged model.

IntechOpen

IntechOpen

Author details

Vera J. Lipton
Zvi Meitar Institute for Legal Implications of Emerging Technologies,
Harry Radzyner Law School, IDC Herzliya, Israel

*Address all correspondence to: vera.lipton@bigpond.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. 