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

5,200
Open access books available

129,000
International authors and editors

150M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

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

Foreword

Foreword

Our lives and our societies are being transformed, thanks to digital technologies, social media, data, algorithms and artificial intelligence. Most of us use the Internet, mobile phones and computers, and most of us are familiar with Facebook, Amazon and Google. These and other technologies are changing forever how we do business and engage in politics, how we conduct academic research, produce and disseminate science and how we teach and learn in online spaces.

As digital technologies have become increasingly integrated into our society, they are generating unprecedented quantities of data, sometimes called big data. Such data is now available more widely and through ever-faster access points. Such data have greater coverage and they include new types of collections and measurements that were previously not available [1]. For example, in medicine alone, the annual rate of increase of digital healthcare data has stood at around 50% over the past 5 years, according to Stanford University.[1] It is becoming clear that digital data is a dynamic force that can drive science forward, especially if the data is openly available.

Much has been written about enabling open access to research data for further study, and researchers are not alone in thinking about how to share research data in electronic formats. They have been recently joined by research funders, librarians and science policymakers. At the same time, many issues surrounding digital scientific data remain unknown, and new problems and questions are emerging as we address the old ones—issues such as the legal, ethical, cultural and technical impediments to open data sharing. Some of the questions that arise are the following: What would motivate researchers to share their own data and to reuse data produced by others? How can the data be curated and shared efficiently and cost-effectively? Who should be responsible for data sharing? What are the optimal ways of enabling access while protecting research integrity and confidentiality? How can we balance the commercial interest of universities with providing open access to research data? Who should pay for open research data?

This book offers pragmatic answers to many of these questions. Rather than emphasising the future ideal of open scientific data, it deals with the current realities and practical aspects of open data sharing. Rather than focusing on data across all scientific disciplines, it draws on lessons learned at CERN and in the use of open clinical trial data. Rather than dwelling on the expectations of funders and publishers, it deals with the challenges posed to researchers—challenges that also present many opportunities. These are clearly articulated in the 'staged model' for open scientific data that the author proposes as a 'flexible template' for future data sharing in other scientific disciplines.

---

[1] Almost 2500 exabites (one exabite = one million gigabites) will be produced in 2020, according to the Stanford University School of Medicine. Compare this with just 153 exabites produced in 2013. See Ref. [2].

It often happens that a research publication is strong on analysis but weak on the proposals. Not this one. The 'staged model' for open scientific data is illustrated in detail in Chapter 8; but this is really the full enunciation of multiple, and complementary, strands of research and data sharing principles.

The principles emerge in logical progression over several chapters—the identification of the ways in which the open data present challenges different from those in open publications (Chapter 7) and raise reuse issues that, in turn, are entirely different from the issues presented by access (Chapter 1); how the open data mandates, adopted on the wings of success of open access publications, fall short of dealing with the specificities of the issued raised by data (Chapter 3); how research data management at CERN (Chapter 5) and in clinical trial data (Chapter 6) offer templates for the 'staged model' that is to be understood as a general frame of reference rather than a single optimal solution. In short, the thoughts and concepts examined in all these chapters seamlessly converge into the 'staged model' introduced in Chapter 8 and is then summarised again from the perspective of providing answers to the research questions posed in the concluding Chapter 9.

The contribution of this book to the open scientific data debate is not only original but is also well-timed. Much of the zeal towards extending the success of open access publishing to open research data often fails to take into account the peculiarities of the latter field. This book takes a fresh look at the main critical points— including the importance of the criterion of data reuse over the goal of access, the multiple fine implementation issues, the importance of getting right the incentives for the different stakeholders such as researchers. The lessons learnt from CERN practice in addressing these issues are especially welcome.

I highly commend the determination of the author and IntechOpen to make this book available as an open access publication. This will, no doubt, enable an expansion of its readership outside the science, research and policy spheres. Yet, as the author rightly points out, for open scientific data to be successful, it needs to be first embraced by researchers. Only then, the general public can follow and engage in open science projects.

I hope this book will bring many illuminating insights into open scientific data to all.

Professor Marco Ricolfi
Chair, Intellectual Property, Department of Law, University of Turin
Co-founder, The NEXA Center for Internet and Society
IntechOpen profile: https://www.intechopen.com/profiles/318786/Marco-Ricolfi