OPEN DATA FOR GLOBAL SCIENCE: A REVIEW OF RECENT DEVELOPMENTS IN NATIONAL AND INTERNATIONAL SCIENTIFIC DATA POLICIES AND RELATED PROPOSALS

FOREWORD

The development of laws and public policies always lags behind the emergence of new technologies. That is certainly the case with the digital technology and network revolution. Governments have been trying to keep pace with the rapid changes in the digital age and to regulate the implementation and uses of these technologies at the national and international levels.

Digital scientific data in the public sector are a significant part of this struggle by the public policy community to catch up. On the one hand, advances in digital technologies have made possible tremendous improvements in the quantity and quality of scientific data. It is not hyperbolic to state that the changes in data collection, management, dissemination, and their potential applications are creating a paradigm shift in the way research is done in most fields of science. On the other hand, much of that potential remains unexploited or underutilized. These shortcomings have resulted in large part from the lack of focus and funding by the science policy, funding, and management communities in support of the scientific data infrastructure and on the new research opportunities based on data science. Most of the innovations and models thus far have been established from the bottom up at the individual investigator and institutional level, rather than from the top down, although, as some of the articles in this volume demonstrate, there is now considerable activity and understanding of these issues among research managers and decision-makers.

This special volume of the CODATA Data Science Journal presents four of these significant data policy developments at the intergovernmental and national levels in Part One. The first two articles address international data policy developments, while the latter two discuss national initiatives. The first paper, by Drs. Dirk Pilat and Yukiko Fukasaku, describes the recent Principles and Guidelines for Access to Research Data from Public Funding of the Organisation for Economic Cooperation and Development (OECD). Dr. Pilat is Head of the Science and Technology Policy Division at the OECD in Paris, France, and Dr. Fukasaku is the Managing Director of Innovmond, a science policy consulting firm also in Paris, and a former OECD staff member responsible for this data policy project. While the OECD Principles and Guidelines are not binding on the Member States, they have great moral force and signify high-level attention to the importance accorded by a major intergovernmental policy organization to promoting data access and sharing among researchers, research institutions, and national research agencies.

The next article, by Prof. Robert Clark of the School of Law at University College Dublin in Ireland, critiques the European Union's 1996 Directive on the Legal Protection of Databases. Although this Directive was not specifically motivated to protect public scientific databases *per se*, it has been roundly criticized over the years for its potentially negative effects on public-sector research. Prof. Clark describes several recent high-level reviews of the Directive that drew negative conclusions about its implementation, including that of the European Commission, which originally promulgated it, and of the British Academy Review Group of Copyright and the Gowers Review of Intellectual Property in the United Kingdom. Prof. Clark concludes that the adoption of an unfair competition law regarding database activities in the United Kingdom and Ireland could be an appropriate ancillary legal solution to an otherwise very unsettled implementation of a flawed Directive.

Part One concludes with two examples of national data policies, one in the process of implementation and the other recommended for action. Dr. Guan-Hua Xu, the Minister of China's Ministry of Science and Technology in Beijing, provides an overview of the status of implementation and future plans for the Scientific Data Sharing Program in that country. Minister Xu concludes that "data sharing will become an effective way to promote science and innovation, and thus become an inevitable choice for progress in the information age." The second of the two articles on national data policies presents the results of the 2004-2005 Canadian National Consultation on Access to Scientific Research Data. The authors, Prof. Michel

Sabourin of the University of Montreal, and Bernard Dumouchel, until recently the Director General of the Canada Institute for Scientific and Technical Information at the National Research Council Canada, describe this National Consultation process and the principal recommendations for action, which have not yet been implemented.

As someone who has been involved in various capacities in all four of these data policy activities, I believe they signify a major shift in the science policy establishment of most countries toward the greater valuation of the role of publicly funded digital data in research. They also augur a concomitant recognition that improved access and re-use conditions of these digital resources will help improve the return on the public investment that originally generated them.

Whereas Part One describes specific data policy initiatives at the governmental and inter-governmental levels, the articles in Part Two analyze various data policy developments. The first of these, by myself and Peter Schröder, of the Data Archiving and Networked Services in the Netherlands, provides an overview of the role of openness as the default principle in managing publicly funded data, including the underlying legal, economic, institutional, and science policy considerations. The article reviews some of the opportunities and challenges to the global science system associated with establishing an open data policy.

The second piece, by Drs. Belinda Seto and James Luo, both of the U.S. National Institutes of Health, explores several of the underlying technical and management requirements in implementing a biomedical data sharing policy. They highlight the networking, interoperability, privacy and security issues, and conclude with an overview of the significance of open-source software in bioinformatics.

Prof. Harlan Onsrud and James Campbell, both of the University of Maine in the United States, outline in the next article an innovative approach that would enable individual researchers to contribute their data sets to the digital commons for much broader use. They describe five key requirements for implementing a highly distributed, voluntary, online infrastructure for sharing such “small science” data, including common-use licensing, metadata generation, provenance tracking, archiving, and peer review and evaluation.

The volume concludes with an essay by Peter Schröder, who points out that large-scale data projects can have an unwelcome homogenizing effect on research directions and output, thereby potentially slowing the progress of science by limiting the opportunities for unconventional, but more innovative, research. Creating global digital commons is necessary and important, but not the only value that needs to be considered. Data resources should be assessed not only for their quality, accessibility, and sustainability but also for their diversity, in order to avoid the restrictive effects on research.

I would like to express my gratitude to all the other authors for their strong contributions to this collection, which provides a valuable snapshot of the current status and emerging trends in scientific data policy. I also would like to thank the editors and staff of the CODATA *Data Science Journal* for their assistance in publishing our work.

Paul F. Uhlir

Editor of this special volume on *Open Data for Global Science*
Washington, DC

CONTENTS