Annals of Library and Information Studies
Vol. 65, December 2018, pp. 271-272

BOOK REVIEW

Data Stewardship for Open Science: Implementing FAIR Principles**,** by Barend Mons, CRC Press, London, 2018, 226 pages, hardback, £46.99, ISBN: 9780815348184.

In the book, Barend Mons of the Leiden University Medical Centre examines the changing paradigms of the scientific data management and an introduction of FAIR principles in data-driven knowledge societies embracing latest information and communication technologies (ICTs). In a dynamic era of data-driven and evidence-based research, the research data principles are bound to change to accommodate the dynamic knowledge productions in trans-disciplinary and inter-disciplinary research. The scientific data are frequently used, reused and shared amongst the collaborative research partners, in order to make the data useful in the downstream of research. The scientific community is also cared for validation of shared or archived data to ensure their public scrutiny and citizen participation in data-driven decision making. Thus, a scientific community engages with a data steward who primarily focuses on the implementation of FAIR data principles in their respective research centres and laboratories.

FAIR Principles are the guiding force for the scientific community in the 21st century, while FAIR stands for the Findable (independently by machines and humans), Accessible (under well-defined conditions), Interoperable (again, independently by machines and humans), and Reusable (under properly defined licenses and properly cited) research objects. FAIR principles apply to all research objects, including the research data, and associated tools and services.

In the Introduction chapter, author Barend Mons defines data stewardship as: "The process and attitude that makes one deal responsibly with one's own and other peoples' data throughout and after the initial scientific creation and discovery cycle" (p. 36). In the book, the author provides several one-liners for conveying the take-home messages. For example, Chapter 1 includes the apt one-liners related to data stewardship: "At the very basis of the needed revolution in scientific methods lies good data stewardship"; "Generating research data without an executable data stewardship plan is scientific malpractice"; "Open science cannot develop without machine-readable data and good data stewardship"; "A good data steward publishes data with a supplementary article (Data(+))"; "Both the human-readable outputs and machine-readable outputs have their rightful place in modern scholarly communication". Thus, the book attempts to serve the purpose of making the readers fully aware of the importance of good data stewardship in open science.

The book is organized into seven data cycle steps, namely, (1) Design of Experiment, (2) Data Design and Planning, (3) Data Capture (Equipment Phase), (4) Data Processing and Curation, (5) Data Linking and Integration, (6) Data Analysis, Interpretation, and (7) Information and Insight Publishing. These seven steps are explained in Chapters from two to eight.

In Chapter 2 "Data Cycle Step 1: Design of Experiment", Mons inspects how a data object is being integrated into a new research experiment. The researcher has to identify if there is pre-existing data from the collections of other people's data and associated services (OPEDAS). Some of the datasets might be reusable that can help the researcher reducing the generation of new datasets. Then she has to check their interoperability. In case not interoperable, she has to extract, transform and load (ETL) them to make them reusable for her purpose. The chapter further guides the researcher how to make use of OPEDAS following a certain number of fact checks, data formatting, and reusability norms.

In Chapter 3 "Data Cycle Step 2: Data Design and Planning", Mons delves into the actual design of data formats, basic software and hardware choices as required in the planned experiments. Here data stewards play a vital role in guiding the research team to make data optimally usable and reusable. For that, a strong metadata norm has to be applied. The author suggests here always use community-compliant,

supported, and sustainable data formats. One has to understand new terminology systems set to have controlled vocabularies.

In Chapters from 4 to 7, the author describes the essential intermediary steps for making the research data open in a FAIR format. In the final chapter, "Data Cycle Step 7: Information and Insight Publishing", the author narrates available options of publishing research data in a FAIR format. Here he appraises publishing practices of a good data steward. He opines, a good data steward never publishes data 'hidden behind an article'. She "publishes data in FAIR format and adds a 'supplementary article' to make data understandable and reusable for people as well" (p. 196).

The book refers to a number of resources and exemplary smart data management plans for FAIR Open Science on accompanying websites http://dmp.fairdata.solutions/resources/ and https://ds-wizard.org.

The scope of this book is not limited to a geographic boundary, while a broader global perspective is presented with easy to grasp dos and don'ts. However, the book was published in the pre-GDPR regime. Thus, the key aspects of data privacy are not presented here. The book helps us in understanding the complex matters related to the implementation of FAIR data principles in a scientific research project. It also confirms the conceptual framework of the responsible research and innovation (RRI) – that includes an element of open science practices in the scientific experiments. This book, thus, is recommended to the scholars engaged in academic research, the future data scientists, data archivists, and open science practitioners.

Anup Kumar Das
Centre for Studies in Science Policy, Jawaharlal Nehru University, India
ORCID: 0000-0001-9490-7938
Email: anup_csp@jnu.ac.in