

A Practical Guide to Communicate Open Data with Ethical Considerations: Annotated

Bibliography

Siqi Zheng, Liza Bolton

University of Toronto

Abstract

This annotated bibliography is a guide to explore various aspects of open data communication from theory to practice with ethical considerations. This annotated bibliography is not meant to be comprehensive but suggestive. All selected sources contain further readings and fruitful resources for different types of target audience. This bibliography contains two sources for the general audience: Open Knowledge Foundation (2022) presents a self-contained guide about the ethical issues of open data for all researchers and York University (2019) illustrates the concept of open data in an accessible way. For institutions, this bibliography includes a paper about the theory and history of open data services and communication (Shaw, 2015) and a typical website of sharing open data under certain ethical and legal regulations by the City of Toronto (2018). We also offer two technical sources (Research Data Management Services @ MIT Libraries, 2016; Wickham, 2014) about open data cleaning and storage. For each source, we present the key focuses and the limitations in two paragraphs.

Keywords: communication, open data, ethics

Annotated Bibliography for Open Data

General Guide

Open Knowledge Foundation. (2022). *How to Open up Data*. Retrieved April 13, 2022, from <http://opendatahandbook.org/guide/en/how-to-open-up-data/>

This website describes the steps to open data with three core rules: simplify dataset, engage frequently with data, and respond to concerns efficiently. It elaborates on the openness of data from a legal perspective (open license) and a technical perspective (means to share data online). Further readings about license can also be found on the website. This guide has language options which allows researchers from different countries/regions to benefit from the handbook. **General audience** can gain a holistic view of open data.

The handbook is comprehensive, but it only has limited discussion of data storage. Hence I suggest another source from MIT that tackles this problem specifically.

York University. (2017, August 29). *What is open data?* Open Access and Open Data Steering Committee. Retrieved April 14, 2022, from <https://www.library.yorku.ca/web/open/overview/open-data/>

York University also illustrates the concept of open data in an accessible way. The website is more friendly with individual researchers given that it covers many aspects of open data. The resources at the end of the page are appropriate for the **general audience**. Unfortunately, some links on the page are not working now. This webpage is not supposed to be a comprehensive list of topics related to open data communication. Thus, it is more appropriate as a starting point.

Institution Guide

City of Toronto. (2018, January 11). *What is Open Data?* Retrieved April 13, 2022, from

<https://www.toronto.ca/city-government/data-research-maps/open-data/what-is-open-data/>

Open Data supported by the City of Toronto is a typical website for open datasets related to Toronto. It provides basic rules to use open datasets effectively and legally. It serves as an exemplar for **institutions** that would like to build similar platforms for data sharing.

This website is also friendly to **researchers** from diverse disciplines.

The datasets on this page cover a wide range of topics in different formats. It is true that some special data (e.g. geographical data) have to be stored in specific formats, the inconsistency of data formats may create difficulties for researchers who would like to work on various datasets at the same time. Furthermore, some datasets are out-of-date.

This raises question of the timeliness of open data.

Shaw, E. (2015). *IMPROVING SERVICE AND COMMUNICATION WITH OPEN DATA: A HISTORY AND HOW-TO*. DataSmart City Solutions.

<https://hwpi.harvard.edu/files/datasmart/files/improvingserviceandcommunication.pdf?m=1630076789>

This paper discusses the concept of open data as publicly available content and a way to publish data. It focuses on why and how governments open data, so **administrators** who get started with open data will largely benefit from the detailed discussion in this report. Furthermore, it suggests a list of general guidelines and existing portals about open data for further reading. In general, this is a good starting point for **people who are interested in the brief history of open data communication**.

This paper, however, is not a step-by-step practical guide. Therefore, researchers who expect some hands-on examples should check out other sources in this bibliography.

Practical Guide

Research Data Management Services @ MIT Libraries. (2016). *Research Data Management:*

Strategies for Data Sharing and Storage [Slides]. MIT OpenCourseWare.

https://ocw.aprende.org/resources/res-str-002-data-management-spring-2016/workshop-materials/MITRES_STR002S16_DSStorage.pdf

If data would be shared publicly, the choice of data storage can also be an ethical decision since the data should be reliable and transparent to researchers. In this presentation slides, the authors show multiple ways to understand data management and practical guidelines in an academic setting. This presentation is helpful for **researchers who would like to publish their collected data**.

This presentation, however, does not provide many theories about the transparency and credibility of open data. Since it is a practical guideline, it should be used as supplementary materials of other readings in this bibliography.

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10).

<https://doi.org/10.18637/jss.v059.i10>

To communicate the data effectively, researchers should pay attention to tidying the open datasets. This paper presents a way to make data tidy for researchers and demonstrates how to deal with different messy datasets and convert them into tidy format with R. This paper is especially helpful to **R users**.

This paper does not target the definition of open data in general, since it only illustrates one possible way to present the final data to the public.