IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 001, LIR041
DOI: https://doi.org/10.52731/lir.v001.041

Current Status and Issues of University Education-related Data in TOKYO OPEN DATA

Mio Tsubakimoto*

Abstract

In this report, the necessary challenges for the openness of educational data in universities are described regarding "TOKYO OPEN DATA". It pointed out the lack of textual data and that data exist but are not open in universities. In addition, the report expressed how to solve these problems. The recommendations also included issues related to open data and data science.

Keywords: data science, education-related data, open data, TOKYO OPEN DATA

1 Introduction

Tokyo, the capital of Japan, has a population of approximately 14 million [1]. As of April 2022, about 11% of Japan's total population is estimated to reside in Tokyo. Tokyo comprises 23 wards, 26 cities, 5 towns, and 8 villages, having 160 universities (14 national, 2 public, and 144 private) [2] that account for about 20% of the national total. Thus, Tokyo is a huge city with a high concentration of population and higher education institutions.

The Tokyo Metropolitan Government (TMG) has established "TOKYO OPEN DATA," an open data catalog site for checking various data and information on the metropolitan government and municipalities [3]. The site is published using CKAN, an open-source data platform. As of April 2022, 4,376 datasets were registered, each categorized into 14 groups involving important TMG issues (Figure 1).

"TOKYO OPEN DATA" contains limited data related to higher education and the 14 group names do not include "education." However, data on primary and secondary education and school boards can be found mixed in the "Other" category. Contrarily, the 207 data published by the Tokyo Metropolitan Government Board of Education include many surveys and reports on libraries and reading as well as school statistics. While all of these education-related data are valuable, the number of open data on university education in Tokyo is scarce. Japan's population of students enrolled in university continues to decline due to a nationwide drop in birth rates. In Tokyo, it is expected that by producing data (including text data) on various surveys and learning behavior at universities that are publicly available and utilizable, policies and services related to higher education can be created through collaboration between industry, government, and academia.

This report describes the status and challenges of education open data, using TOKYO OPEN DATA as an example, for using university education-related data as open data.

* University Education Center, Tokyo Metropolitan University, Tokyo, Japan

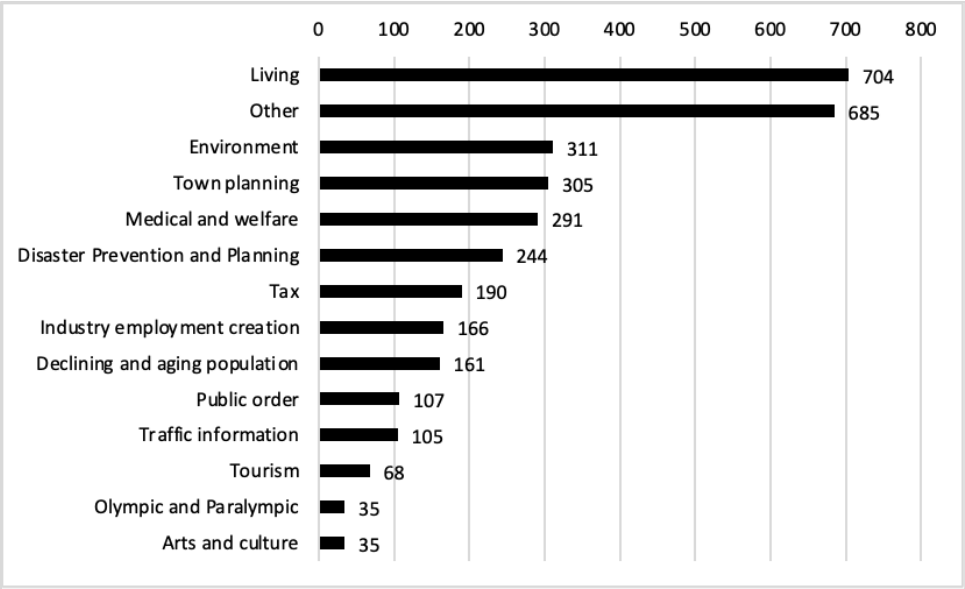

Figure 1: 14 group labels and number of datasets

2 Open Data Related to University Education

This section discusses the current status and issues of open data related to university education using TOKYO OPEN DATA as the subject matter.

2.1 Data Are Gathered but Not Open

In most cases, education-related data obtained at universities are not open. However, multiple type education surveys are conducted annually at most universities and although vast amounts of data are obtained, they are closed within those universities. One possible reason involves that many education-related data in universities are specific to the university context and are not necessarily considered to be open. Contrarily, it would be good if items that are not context-specific (e.g., learning time outside class time, motivation to learn in online classes, etc.) could be implemented and made as open data with an assortment of item names and scales. However, many universities have not progressed in raising awareness among their members regarding the need to open such data or develop the necessary procedures. As the hurdles to openness are very high for questionnaire and performance data, it is advisable to start with educational data of a highly public nature (e.g., neighborhoods usage rates of university libraries, usage rates of lifelong learning courses, etc.).

2.2 Collecting Data in a Machine-readable Format

Figure 2 shows the data formats and their numbers in all datasets of "TOKYO OPEN DATA." Of the data collected, 56% were collected in CSV format and the second most common format was PDF. PDF is not a machine-readable format and is considered unstructured data [4]. PDF data are often scanned straight from brochures or conference papers and cannot be used for analysis. The

Ministry of Internal Affairs and Communications has proposed rules for the notation of machine-readable data [5]. Those in charge of handling educational data at universities would learn these rules and know how to hold these data for easy opening and analysis. This necessity may require data literacy education for faculty and staff.

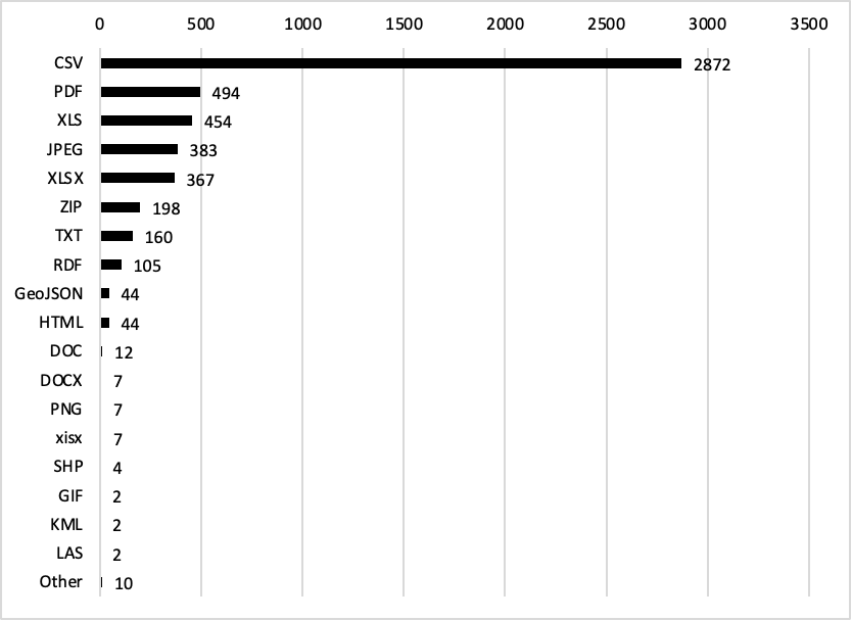

Figure 2: File formats of datasets

2.3 Maintaining and Integrating Data

Even with similar content items and survey forms, data is submitted by different municipalities, resulting in scattered data and inconsistent formats. It may be necessary to standardize the wording of questions and data format in higher organizations in individual universities to make the data "usable." Data cleaning work may also be required. If having personnel at each university taking charge of format unification and data cleaning is difficult, ways to promote data maintenance may involve external personnel participating as "data volunteers" or crowdsourcing. Data volunteering and crowdsourcing can also promote lifelong learning and employment.

2.4 Combining Government, Industry, and University Data

"TOKYO OPEN DATA" still shows little data from universities and industries. Similar to administrative level data, data is also available at the university level and from education providers. Obtaining an additional three-dimensional view of a single phenomenon and noticing many things related to education/industry are possible by combining several data sets. The benefits of actively publishing data from universities and industry on the same platform also exist for three parties: industry, academia, and government.

2.5 Enriching Text Data Regarding Both Quality and Quantity

Universities generate several textual data on education, such as free-text descriptions in class evaluation and graduation surveys, interview data, and public relations publications. Not all data can be made available to the public, but even the parts that can be made open should be made public for limited purposes. Free statements often contain detailed content, which can be analyzed using various text mining methods for sentiment and content. There are multiple uses of data depending on the analysis purpose, such as analyzing text data alone or with numerical data. Even when content that cannot be published is included, it may be possible to deal with it through data cleaning. Therefore, it is better to actively discuss the publication of text data. In addition, it is necessary to steadily promote separate open science initiatives for text data collected in the research.

3 Conclusion

In this report, the necessary challenges for the openness of educational data in universities are described regarding "TOKYO OPEN DATA." It identified the lack of textual data and that data exist but are not open in universities. In addition, the report also expressed how to solve these problems. The recommendations also include issues related to open data and data science. Japanese government nowadays requires universities to develop data science human resources. Specialized educational courses are being developed at many universities in Tokyo. Further, the use of open data reminds people of the collaboration between industry and government. However, it seems necessary for data-holding governments to actively collaborate with the higher education community, aiming both to develop human resources involved in data science and, simultaneously, to create new knowledge through open data.

Acknowledgement

We would like to thank Editage (www.editage.com) for English language editing.

References

[1]Statistics Division, Tokyo Metropolitan Government, Statistics of Tokyo; https://www.toukei.metro.tokyo.lg.jp/jsuikei/js-index.htm, 2022.

[2] Ministry of Education, Culture, Sports, Science and Technology, School Basic Servey 2021. https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00400001&tstat=000001011528, 2022

[3] Tokyo Metropolitan Government, TOKYO OPEN DATA; https://portal.data.metro.tokyo.lg.jp, 2022.

[4] Glossary, OPEN DATA HANDBOOK; https://opendatahandbook.org/glossary/en/, 2015.

[5] Ministry of Internal Affairs and Communications, Notation for producing machine-readable data in statistical tables. https://www.soumu.go.jp/main_content/000723626.pdf, 2020 [In Japanese].