# Comparison of Digitized Book Index among Japan, the United States, and the United Kingdom

Teru Agata
Asia University
Japan
agata@asia-u.ac.jp

Shuichi Ueda
Professor Emeritus
Keio University
Japan
ueda@keio.jp

## ABSTRACT

Some reports of publishers' sales data on trends in e-books have been released, but no attempts have been made to precisely indicate the particular titles of the printed versions that were digitized. Therefore, the Digitized Book Index (DBI) has been developed, comprising titles in the national catalog databases of Japan, the United States, and the United Kingdom. The DBI is composed of titles listed at major e-book sites, on e-book platforms, and in electronic libraries. It was found that: 1) the DBI is generally increasing in volume; 2) there are differences among the three countries; and 3) the DBIs are different according to the classification item.

## CCS CONCEPTS

• Applied computing → Computers in other domains → **Digital libraries and archives**

## KEYWORDS

E-books, Bibliometrics, Library catalogs, Open data, Google

## 1 INTRODUCTION

The label "e-book" refers to electronic books, and it is an ambiguous term that encompasses several concepts, forming a definition that changes over time. The *ALA Glossary of Library and Information Science 4th ed.* defines an "e-book" as "an electronic version of a printed book that can be read on a computer or handheld device such as an electronic book reader." This study used the ALA definition to define e-books and, therefore, books that originated in a digital form ("born digital") are not included.

Project Gutenberg, which started its service about 50 years ago in 1971, was the first provider of e-books [1]. Initially, studies for e-books were focused on reading devices, but attention has recently shifted to e-books in everyday life [2]. However, to study the uses of e-books, it is necessary to first clarify the publication numbers and the types of print books that are being converted into e-books. Because publishing data on e-books are not available, this study developed a method for comparing e-book titles in the databases of national libraries using the titles on the main e-book platforms, and it thereby proposes a general index of digitized books and a method for digitizing them: the Digitized Book Index (DBI). DBI is defined as the number of digitized books published in a year divided by the number of books published in a year. This index gives an objective perspective to the future of the chaotic publishing market, and it is also possible to compare internationally.

## 2 METHODS

Publishers usually convert paper books into e-books that are then posted on digital platforms such as Google, but cross-platform search tools for e-books have not been realized because e-book services, publishers' websites, joint platforms of multiple publishers, electronic libraries, and the other relevant platforms are not linked. To conduct this study, bibliographical records with International Standard Book Numbers (ISBN) were extracted from the Online Public Access Catalog (OPAC) database of the legal deposit library, which comprehensively collects domestic publications. These data comprised 5,000 book titles for every five-year period from 1985 through 2010, and 5,000 book titles for every year from 2014 to 2017. These catalog records were matched by ISBN to the titles in the bibliographical database on the platforms listing e-books.

First, the feasibility of the study's approach was confirmed using the catalog database of the National Library in Japan. The holdings of the National Diet Library (NDL) include almost all Japanese publications. Therefore, book titles with bibliographical information were extracted from the NDL catalog database using the NDL search application program interface (API). The retrieved titles were matched to the titles in the following e-book platforms: amazon.com, kinokuniya.com, books.google.com (Google Books), kw.maruzen.co.jp (Maruzen eBook Library), and dl.ndl.go.jp (NDL Digital Collection). In the United States, the Library of Congress (LC) is the national legal repository for copyright protection and registration. In 2017, the LC began providing 25 million catalog records from 1968 to 2014 as open data. Book titles with bibliographical information were extracted from the LC open data and merged with the collected titles using the LC Z39.50 API. These titles were matched to titles in the LC, amazon.com, and Google Books e-book platforms using ISBNs. Similarly, the United Kingdom's bibliographical data were extracted from the open data of the British Library (BL) catalog and collated with amazon.com and Google Books databases. Calculate example of DBI in 2017 is shown in Table 1.

IEEE
computer
society

**Table 1: Examples of DBI (2017)**

|  | Japan (NDL) | U.S.(LC) | U.K.(BL) |
|---|---|---|---|
| Total ISBNs | 56,217 | 60,038 | 82,724 |
| Sample Books | 5,000 | 5,000 | 5,000 |
| E-books | 1,843 | 3,127 | 3,074 |
| DBI | 36.9% | 61.5% | 62.5% |

## 3 RESULTS

Figure 1 shows the DBI in the three countries under observation from 1985 through 2017. The DBI's volume increases across time in Japan. In 2017, about 36.9% of the titles of the print versions were converted into e-books.
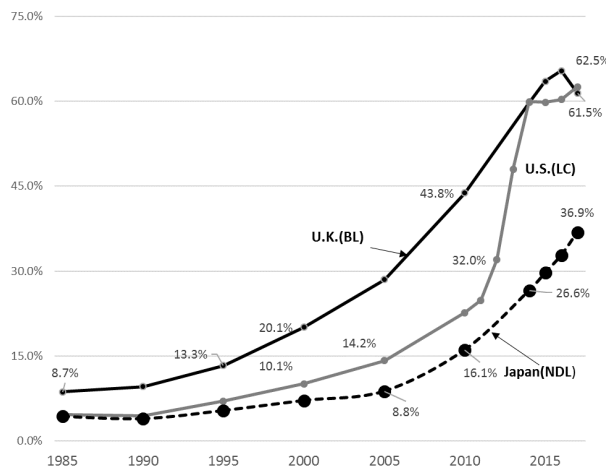


**Figure 1: DBI in Japan, the US, and the UK (1985–2017)**

Similarly, the more recent the publication year, the larger the DBI in the United States. The DBI was over 50% in 2014 and was around 60% afterward. The DBI of the US is larger than that of Japan. However, its increases leveled off during 2015–2017, which might relate to delays in cataloging new books at the LC.

The publication of e-books in the United Kingdom has historically surpassed the other two countries, but the United States has recently exceeded it. Although the United Kingdom's DBI in 2017 was small, there is a time lag in converting newly published paper books into e-books. About twenty percent of the e-books in the LC also are in the British Library catalog.

Figure 2 compares the distributions of Japanese e-books classified according to the Nippon Decimal Classification (NDC) system, to those of the US and UK, which are classified according to the Dewey Decimal Classification (DDC) system, in 2014, 2015, 2016, and 2017. In the US and UK, the differences in the DBI by classification are small compared to Japan, where the DBI varied greatly by type. In Japan, the e-book proportions were high in the arts, which includes Japanese comic books, and in literature, which includes novels, whereas the philosophy and general works categories in the United States, which includes the computer sciences, had the largest proportions of print books converted to e-books.

## 4 DISCUSSION

This study developed a DBI that could be used in many countries. Valery Giroux predicted in 2015 that e-book sales as a percentage of total book sales, which was about 12% in 2013, would more than double by 2018[3]. According to this study's DBI, these three countries had exceeded 24% by 2017. One limitation of our DBI is that it cannot be used on books that do not have an ISBN. However, so long as there is no cross-platform search tool for e-books or a union catalog of e-books, this DBI is effective.

This study was found that: 1) the DBI are generally increasing in volume; 2) there are differences among the three countries; and 3) the DBI are different according to the subjects represented by classification table.

## REFERENCES

[1] Project Gutenberg. https://www.gutenberg.org/wiki/Gutenberg/
[2] Agata, T.; Ueda, S. (2018). Current status of e-book conversion in Japan. Proceedings of Japan Society of Library and Information Science annual meeting, (p.111-114). Tokyo: Waseda University.
[3] Giroux, Valery. E-book sales as a percentage of total book sales worldwide in 2013 and 2018. The Statistics Portal. 2018. https://electricnews.fr/us-e-books-statistics/
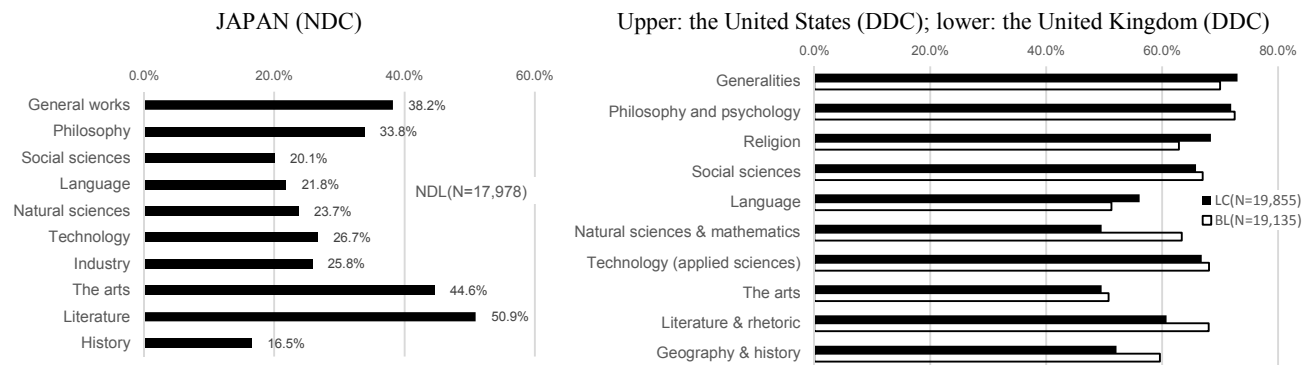
**Figure 2: DBI by classification (2014–2017)**