

# Comparing the access to and legibility of Japanese language texts in Massive Digital Libraries

Andrew Weiss

Oviatt Library

California State University, Northridge (CSUN)

Los Angeles, CA, United States

andrew.weiss@csun.edu

Ryan James

Hamilton Library

University of Hawai'i at Manoa

Honolulu, HI, United States

rsjames@hawaii.edu

*Abstract* — In previous studies, Weiss and James have examined the impact of Massive Digital Libraries (MDLs) on the development of libraries in terms of copyright, metadata, accessibility and diversity. This paper continues these investigations by presenting the results of a study conducted in 2013-2014 that examines the coverage and accessibility of Japanese language books in two MDLs, *Google Books* and *HathiTrust*. A random sample of 5000 Japanese-language books with publication dates prior to 1943 was extracted from the OCLC WorldCat database; of these another 800 were randomly selected and 400 titles were examined. The titles were queried in both *Google Books* and *HathiTrust*. The texts were then examined for their level of typical user access, their accuracy in metadata and their quality of scans. Despite their likely public domain status within Japan and in the United States, 0.2% (N=1) of the sampled texts were visible in *Google Books* as full texts. While 12.5% (N=50) of the sample were visible in *HathiTrust*. Within the full view texts, errors in scanning and metadata were identified, including problems with legibility ("moji tsuture") in 68% of visible texts; distorted content (including slanted and upside-down pages) in 90%; motion or blur of turning pages captured by digital cameras in 48%; extra-textual objects (3-D items not part of text; i.e. fingers, hands, book holders, etc.) in 94%; and use of heavily-defaced, dirty or fragile source material in 28%. The most common metadata errors were missing bibliographic information, especially missing page numbers (in 18% of texts) and incomplete tables of contents (in 22%); and problems associated with poor OCR, especially unusable keywords and common phrases (in 50% of texts) that appear to be random words, articles, and unpronounceable symbols.

**Keywords**—Massive Digital Libraries (MDLs); *HathiTrust*; *Google Books*; metadata aggregation; mass digitization; public domain; copyright

## I. INTRODUCTION

### A. Examining Massive Digital Libraries (MDLs)

Weiss and James in their research propose Massive Digital Libraries (MDLs) as a separate class of digital libraries [1][2]. This class can be defined in part as collections of mass-digitized print books rivaling the size of traditional libraries.

They comprise "extremely large digital collections (>1,000,000) of digitized print book content aggregated from multiple mass-digitization projects" [1] The reliance on mass digitization and its industrialized and automated, quantity-over-quality approach to content creation is a defining characteristic but also a source of well-documented problems with metadata accuracy [3][4][5], scanning quality [6], and lack of diversity in collection development.[7] Quality control for digitized books remains a constant concern and reminder that mass-digitization and mass content-aggregation coupled with extracting the content from the original physical container comes with a price.

*Google Books*, with an estimated 30 million digitized books, and *HathiTrust*, with 13,467,158 (5 million openly available), remain the two most well-known mass digitization projects. [8] Other projects, such as the *Internet Archive* and *Europeana* have significant digital book collections, but they also focus on alternative types of digital media content, including software programs, audio recordings, videos, and still images.

*Google Books* and *HathiTrust's* history is a shared one, as much of the works were scanned by Google in partnership with many of the current *HathiTrust* member institutions, including University of Michigan, Harvard, Yale, University of California campuses and many others. As a result, when discussing either MDL, there are many similarities in terms of collection content, metadata development and member organizations. However, each one's organizational focus and motivation have diverged over time. Google is a for-profit information technology company concerned with shareholders, profit margins, and market positioning, while *HathiTrust*, which is rooted in academic and public libraries, remains concerned with the educational mission of its member organizations. Distinct differences appear in their approaches to content development, metadata creation and accuracy, and overall access and approaches to copyright compliance.

The development of MDLs as major influences on libraries and information science as well as the publishing industry is well-documented, especially with regards to two high-profile lawsuits initiated by Authors Guild. [9][10] The resolutions of these cases in the favor of MDLs only further

solidifies their foundations and ensures that their disruptions against traditional publishing industries and their distribution systems will remain in place. The rulings hinged on analyses of Fair Use in US law and the creation of a digital corpus, which allows researchers as well as those with physical disabilities to search through millions of texts.

Ultimately, the number of texts digitized was not at issue, the court reasoned. Instead, it was examined whether a digitized corpus of copyrighted and orphan works books was a transformative use of the original content. Presiding judges in both cases reasoned that the creation of a digital corpus for enhanced searching and aggregated metadata was indeed changed enough from the original to merit Fair Use. It also didn't matter whether it was one book or one million: Fair Use was Fair Use *regardless of the scope*. Currently, appeals are pending, but it seems that MDLs do have sufficient legal standing to continue mass-digitizing print books. [11] They are here to stay.

### B. Diversity, Accessibility, and Open Access in MDLs

The major flaws associated with MDLs are generated in part by their reliance upon mass-digitization and mass-aggregation efforts to develop the collections. In previous studies regarding the diversity of texts related to non-standard English collections of materials, including Hawaii and Pacific collections, and Spanish-language collections, the authors demonstrated that a lack of coherent policy negatively impacts the overall quality of the MDL's diversity. [2][7]

Spanish language collections were found to be limited in *HathiTrust*, for example, when examined by Weiss and James in 2013. Spanish is the second-most spoken language after English in the United States, but is underrepresented in the *HathiTrust*. This is likely a result of the historical focus of the academy upon French and German publications. [2] The reliance on similar research institutions – especially those designated as "R1" in the United States and described by the *HathiTrust* as "an international community of research libraries," (i.e. Harvard, Yale, Cornell, UCLA, University of Michigan) – further distorts the online record.

Additionally, full-text accessibility has also diminished somewhat since the projects began. This is likely an impact of the Authors' Guild lawsuits. The *HathiTrust* provides access to all texts for its member users, but not those outside of the consortium. *Google Books* takes a more conservative approach for access and provides fewer fully-accessible texts to researchers. However, it also provides its digital corpus in the form of searchable datasets and an online data-visualization tool called the Ngram viewer. While the full texts are generally not visible, unless determined to be in the public domain, these enhancements for searching and text/data mining provide significant advantages to the researcher.

### C. Problem Scope

After finding and documenting problems with Hawaiian and Pacific materials, and subsequently problems with coverage of Spanish-language texts, the researchers asked whether the same levels of coverage from mass-digitization and digital full-text collection access in MDLs

would apply to non-European languages -- especially for those languages that do not use Roman letters as the basis of their writing systems.

A few candidates were considered, including Cyrillic-based languages, Greek, Chinese and other East- or South-East Asian languages. However, Japanese was ultimately chosen due to the complexity of the written language, which often utilizes up to four different writing systems (*Kanji*, *hiragana*, *katakana*, *romaji*). Furthermore, the physical materials used in many of the traditionally published works are unique and provide distinct differences between the physical properties of Western books and Japanese texts. Finally, the source institutions for the projects provides a significant pool of texts, roughly 228,251 (3% of its collection) from which to derive general trends in the mass-digitization of Japanese-language books.[8]

## II. METHODOLOGY

A random data sample containing the metadata for 5,000 Japanese language texts published prior to 1943 and held with the *HathiTrust* was extracted from OCLC WorldCat. Using this as a master list, 800 texts were then chosen using a random number generator to ensure that no texts were selected based on their physical properties. The first 409 texts in the list (with ascending OCLC numbers) were queried for their level of access in both *Google Books* and *HathiTrust*. Each item in both MDLs was examined using titles, author names and publication dates in order to determine exact editions of a text.

Item records found in *Google Books* and *HathiTrust* were evaluated for their level of access, ranging from 'No Record'; 'Record Only'; 'Partial View'; 'Preview'; and 'Full Text'. 'No Record' designates that the item record was not retrievable using the available metadata. 'Record Only' designates that the MDL provides only metadata about item queried. 'Partial View' designates that small parts of the text, known as 'Snippets' in *Google Books*, are available to view; or, that queries of word frequency inside a text are available to users. 'Preview' designates that several pages or multiple chapters in various sections of a text are available for view. 'Full Text' designates that texts are fully searchable and all pages can be viewed without restriction.

In the cases where the texts were fully visible, notes on the quality of the scans were compiled. Scans were examined for several factors, including legibility of source text (i.e. *moji tsubure*, distorted auto-correction, etc.), existence of extra-textual objects captured in the digital images (i.e. hands, fingers, clamps, etc.), placement of pages (i.e. slanted, upside down, backwards, mirrored), color correction, and motion/blur of pages captured in digital images. Finally, the quality of source text was noted, especially the physical properties (i.e. types of paper used, or binding type), and overall condition (i.e. markings in pen or pencil, stains, and dirt).

As a side note, *moji tsubure* is defined as the blurring of the characters' strokes. The meaning derived from *Kanji* is dependent upon subtle changes in stroke order to create their structures. When the characters are too blurred, the ink has bled, or the ink gradation has been removed to increase contrast, information is lost. At its best, it is a minor inconvenience to piece together words and concepts. At its worst, the text becomes incomplete and illegible. [See Image 1 below]



Image 1: Detail of scan from Chōsen kōshō jikenroku 朝鮮交渉事件録 [OCLC WorldCat #297314060] in HathiTrust showing "moji tsubure," a condition where the Kanji or other characters, which should comprise of clearly-defined strokes, have lost their definition and appear blurred or globular. This results in lost legibility and incomplete digital texts.

### III. RESULTS

The following levels of access (N=409), margin of error at 4.64% at .95 level of confidence, were found in *Google Books* and *HathiTrust*:

| Access Level | Google Books | HathiTrust | % GB / HT     |
|--------------|--------------|------------|---------------|
| No record    | 5            | 0          | 1.2% / 0.0%   |
| Record only  | 71           | 0          | 17.4% / 0.0%  |
| Partial      | 331          | 359        | 80.9% / 87.8% |
| Preview      | 1            | 0          | .2% / 0.0%    |
| Full text    | 1            | 50         | .2% / 12.2%   |

Table 1: Access levels in *Google Books* and *HathiTrust*.

All the books were found in the *HathiTrust*. However, five were not found in *Google Books*. 71 of the books in *Google Books* were available only as metadata records, suggesting that, along with the 5 books with no findable records, 76 (18.2%) Japanese-language books from this list had never been digitized by Google. Overall the levels of access for partial views, which include a searchable text mining feature in *HathiTrust*, are quite close. However, significant difference in the access to full-text versions is noticeable. Only one book was accessed as full text in Google while 50 were accessed in the *HathiTrust*. [See Figure 1]

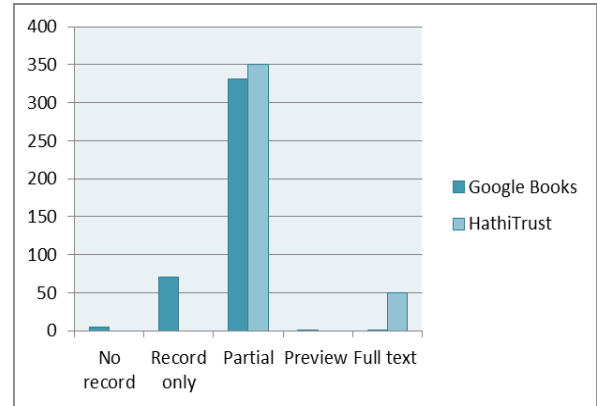


Figure 1: Access levels in *Google Books* and *HathiTrust*.

The following frequency of scanning errors was noted among the 50 full-text accessible items in the *HathiTrust*:

| SCAN ERROR TYPE                                      | N (50) | %   |
|--|--------|-----|
| moji tsubure   | 34     | 68% |
| extra-textual objects (holders, hands/fingers, etc.) | 47     | 94% |
| slanted pages  | 45     | 90% |
| upside down pages                                    | 4      | 8%  |
| blur/motion/page turn                                | 24     | 48% |
| inconsistent image correction                        | 17     | 34% |
| distorted/curved pages                               | 19     | 38% |
| dirty/marked pages                                   | 14     | 28% |

Table 2: Scan error types in full-text digital books

Despite a high margin of error (13%) from this small subset of fully-accessible texts, clear trends were observed. *Moji tsubure* [See Image 1] was found in 68% of the visible texts (N=34). Additionally, extra-textual objects (i.e. hands, fingers, book clamps, miscellaneous objects) were found in 94% (N=47) of the texts. These break down further to the following: book holder/clamps holding pages found in 66% of texts (N=33); hands or fingers were found in 56% of texts (N=28); and miscellaneous objects (i.e. library cards, inserted papers or clippings, etc.) were found in 48% (N=24) of texts. [See Image 2] Poorly displayed pages (i.e. slanted, upside-down, backwards, etc.) were found in 90% of books (N=45).

Inconsistent color correction, resulting in some images black and white and some in color, was found in 34% of texts (N=17). Finally, distortion of the page itself, likely due to computer auto-correction of curving pages, was noticed in 19 texts. [See Image 3]

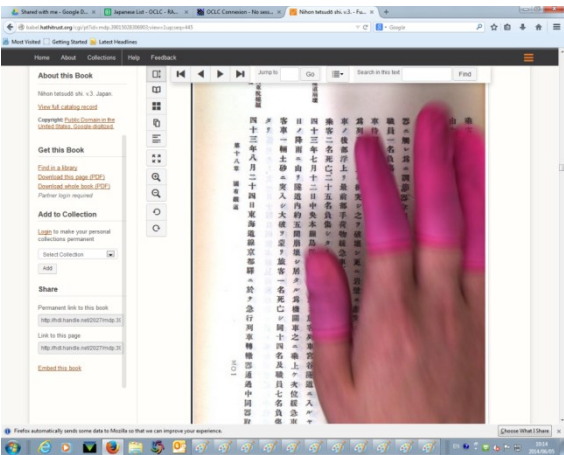


Image 2: Screenshot from Nihon tetsudō shi [OCLC WorldCat #551239900] in HathiTrust showing an extra-textual hand in the image. Hands and fingers were found in 28 of 50 (56%) sampled texts.

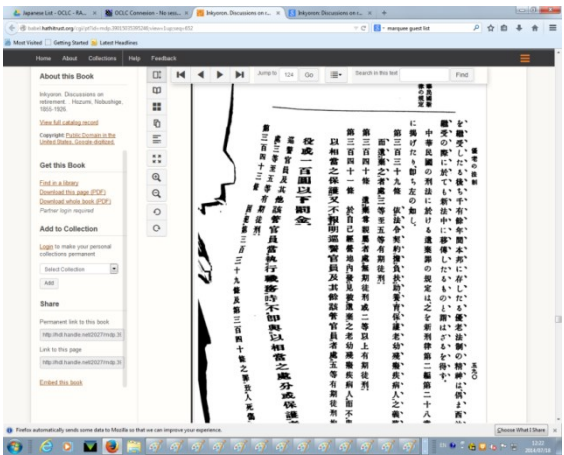


Image 3: Screenshot from Inkyoron [OCLC WorldCat #551596587] in HathiTrust showing extreme distortion of text. This type of text distortion was found in 19 of 50 (38%) sampled texts.

The following metadata issues were tabulated from the sample (N=409), margin of error at 4.64% at .95 level of confidence:

| METADATA ERROR               | N (409) | %      |
|------------------------------|---------|--------|
| Incomplete table of contents | 89      | 21.70% |
| Missing page numbers         | 72      | 17.60% |
| Unusable keywords            | 197     | 49.90% |

Table 3: Metadata errors in Google Books.

From a random sample of 409 texts, 49% (N=197) showed problems with unusable keywords. [See Image 4] Additionally, tables of contents as well as pages numbers were missing from the metadata records in 21.70% (N=89) and 17.6% (N=72) respectively.

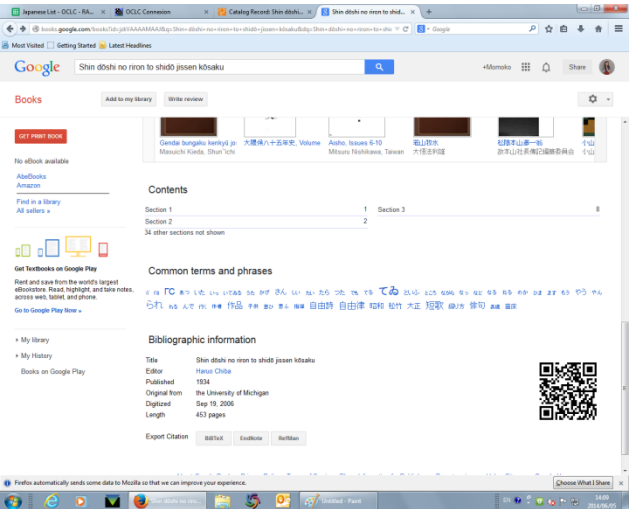


Image 4: Screenshot from Shin dōshi no riron to shidō jissen kōsaku [OCLC WorldCat #551246326] in Google Books showing unusable keywords,( i.e.: ところ, ながら, なっ, など, なる, ねる, のか, ひま, ます, もう, やう, やん, られ, れる, んで, et al.) which are found in 197 of 409 (49%) sampled texts.

#### IV. DISCUSSION

##### A. Access to texts in Google Books and HathiTrust

The results show only slight differences between Google Books and HathiTrust with regard to the overall access of sampled books. Both MDLs provide roughly the same amount of texts (331 / 81% and 359 / 88% respectively) in partial view. Google allows users to see snippets of the text while HathiTrust allows non-member users to search for frequency of terms in order to determine whether texts are relevant to their information needs. Though these are not exact equivalents in terms of access, both are limitations upon the visibility of the digitized text far beyond the typical in-library access to a print book. In neither case are users allowed to see the original text. This is actually less useful than trying to track down the physical book through inter-library loan or through a book vendor.

The main difference, though, occurs in the access to public domain books. Google Books provides access to merely one text from the sample. HathiTrust, in contrast, provides 50 (12%) texts. Why Google Books has chosen not to provide these texts when they are well within their rights to do so is not clear. Perhaps they consider the problem too small to address. Perhaps the complexity of the issue makes it not worth the time to fix.

Additionally, the problem of international copyright comes up as well. This raises issues of whether an international collection of all the books in the world can be provided by a US-based corporation guided by and operating within the laws of the United States. Google appears to be avoiding the issue completely by not providing the sampled Japanese texts as full texts. *HathiTrust*, for its part, appears to be providing texts that were published prior to 1923, which is the cut-off date in the United States for public domain works. However, it should be noted that there are numerous texts that have been published after 1923 and just as likely in the public domain. An automatic cut-off date is a useful rule of thumb but should not replace a comprehensive solution.

In the case of books published in Japan before 1943 (which is the entire sample), it could be argued that most, if not all, of the books in the sample are likely in Japan's public domain as well as the United States'. The law prior to the US 1976 copyright act required authors to actively provide registration as well as indication within the text of copyright protection. Without it, the works fell into the public domain. It would be possible to argue that the *HathiTrust* might consider adding these as full text books given the unlikelihood of litigation for these works. The risk seems minimal, and a takedown policy would likely resolve the problem.

#### B. Scan quality of Japanese texts in the HathiTrust

By far the most common problem observed in the sampled texts was the consistently poor quality of the digital imaging. These problems are not merely results of the age, quality or even language of the source material. Instead, most of the poor scans are a direct result of the mass digitization process itself. The emphasis upon quantity rather than quality has resulted in many of the scanned texts having backwards, upside down and slanted pages. The appearance of extra-textual objects within the books is a result of scanning too fast and not cropping the resulting images. Hands and fingers appearing in the texts as well as the blurred pages captured mid-turn suggests that the scan speeds are too high to avoid human interference. [See Image 5] Slower scanning to work at speeds more attuned to human working conditions or more automated scan processes would improve these basic errors. Additionally, attempts at fixing the problem have resulted in poor imaging results as well. [See Image 6]

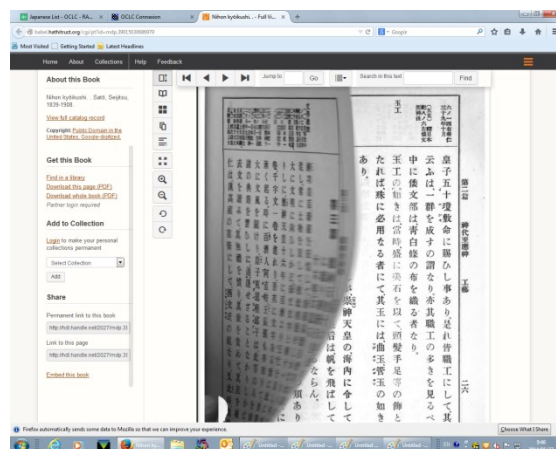


Image 5: Screenshot from *Nihon kyōikushi* [OCLC WorldCat #551383937] showing page caught in the act of scanning. Blur and interference caused from too rapidly scanning texts.

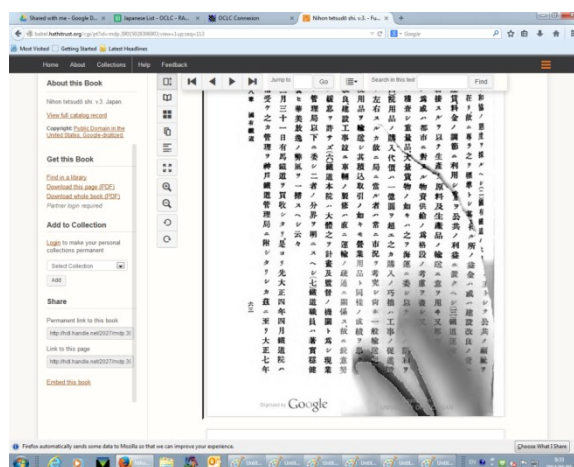


Image 6: Screenshot from *Nihon tetsudō shi* [OCLC WorldCat #551239900] page with poorly attempted correction of extra-textual object (hand).

Auto-corrected pages are another chronic source of poor scan quality. Color correction goes wrong fairly often within these texts as does the post-scan generation of text-alignment. [See Image 7] The software creating these distortions would need to be upgraded to help improve this situation. As it is, the results are largely unusable for readers. The distorted text may be readable to some, but it interferes with the optical character recognition and further reduces the legibility of the text.



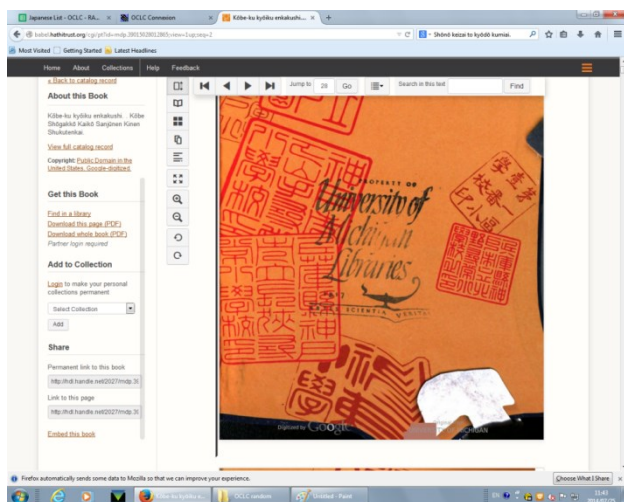


Image 7: Screenshot from *Kōbe-ku kyōiku enkakushi* 神戸区教育沿革史 [OCLC WorldCat #551662901] showing poor color-correction as well as overall distortion of the image.

The frequency of these scan errors raises the question of what kind of digital library do users deserve? If this is the best that MDLs can do, it is well worth questioning the overall endeavor. Many of the books are simply illegible, incomplete, and unusable. Given the amount of time and money spent on these projects there should be improved control of the output quality. Otherwise, users will turn to future alternatives that resolve these problems.

Some of the problems with scanning Japanese books likely stem from assumptions that their physical makeup, their content, and their presentation would be exactly like books published in the United States or Europe. Not all books in Japan are bound in the same way as books in the United States. Many books in Japan contain bibliographic information in the margin (called the *gyoubi*). Many books start at the opposite ends of the bound pages. Many books are written with the words moving top to bottom and right to left. [See Image 8] These customs are usually not copied in Western books or books published in the United States.

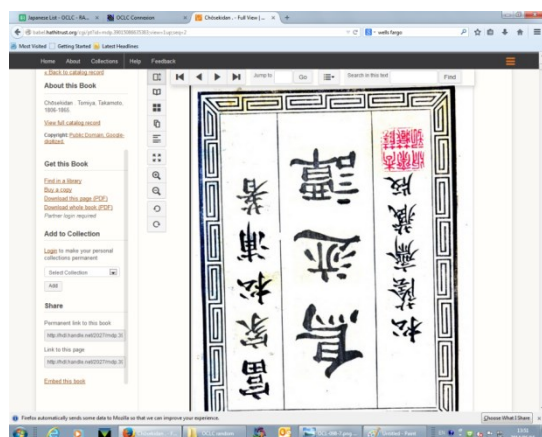


Image 8: Screenshot from *Chōsekidan* [WorldCat #551313119] showing title page with text scanned upside down.

### C. Metadata quality of Japanese books in Google Books

Metadata has been one of *Google Books* significant problems, as documented by several studies, including James and Weiss (2012), Nunberg in 2009, and others. [3][4][5][12] In the sampled texts for this study, the problems noted are mostly a result of the problem with OCR software for Japanese not being up to par as English.

The researchers examined how well the available text and keywords were legible to readers. Regarding the legibility of visible OCR text within *Google Books* records, nearly half of the books contained errors in legibility of the visible sample texts provided in a Google record.

Additionally, the problems of OCR extend to the development of tables of contents as well as keyword terms. Keywords are especially problematic and are mostly rendered unusable due to the problems with OCR. Many words repeat and the inclusion of grammatical articles and clauses (neither nouns nor verbs) such as *ところ*, *ながら*, *なっ*, *な*, etc., does not help readers find the topics they are searching for.

## V. CONCLUSIONS

### A. Quality control and cultural misunderstandings

The evidence suggests that quality control needs to play a more central role in the development of mass-digitized collections. While it is not always possible to examine every page of every book to ensure that they are perfect, the near-universal occurrence of errors found in the scans of these books suggests that there is truly a problem with quality control. If 94% of texts contain an added, non-textual object such as a finger or hand, or if 90% of texts contained pages that were scanned incorrectly (slanted, etc.), then there is a true problem with the workflows, methods and equipment used in the projects. Some of the problems stem from trying to do too much too fast. Some of the problems stem from not being mindful of the unique constructions of books. Not all books can be scanned in the same way, especially as book size, age, binding, paper material and size of font/text create variables that are impossible to anticipate for every book.

Additionally, the problems of native English-speaking digitization staff dealing with the Japanese language and the attitudes surrounding books in Japanese culture could also contribute to scanning problems. It is difficult to imagine someone familiar with the Japanese language allowing a book to be scanned backwards or upside down, as was seen in Image 8.

As far as access is concerned, it is likely that most, if not all, of the pre-1943 books sampled are in the public domain in Japan. The "life of author + 50 years" copyright policy, as it is currently enforced in Japan, allows books from authors who died in 1965 to be in the public domain. Obviously, *HathiTrust* is following US copyright law as all of the full text books in their collection were published pre-1923.

However, it is not clear that litigation will ever occur over the access status of a book in the public domain in a foreign country, especially if the books were probably never registered with copyright offices in the United States – a strict requirement of the time period. *HathiTrust* may be acting overly cautious in terms of liability. Google, on the other hand, is just missing out completely. There is no reason for books clearly in the public domain in both the United States and in Japan to be missing from their collection. Again, this hints at poor overall quality control and undefined collection development policies.

Metadata errors remain a consistent problem as well. Most of the problems found in this sample were likely the result of underdeveloped OCR software. This is not entirely Google's or the *HathiTrust*'s fault. The Japanese language is more complex due to its multiple writing systems. Designing an OCR program to meet all of the wrinkles found in *Kanji*, *hiragana* and *katakana* will be difficult. The technology has simply not yet advanced enough. This is evident in the 48% of sampled texts that had problems with keywords, tables of contents and page-numbers derived from optical character recognition outputs.

### B. Limitations

This study examined only a tiny fraction of the overall corpus of digitized books. Examining larger full-text samples (increasing from 50 to 400, for example) would provide more accurate rates of accessibility, rates of scan and metadata errors. At the same time, it is unclear without manually searching both Google and *HathiTrust* which of these appear both in Google and in the *HathiTrust*.

### C. Future Directions

Future studies will examine larger samples of the Japanese language text corpus in order to get a more accurate picture of the rates of scan and metadata error that appear in both the *HathiTrust* and *Google Books*. For *HathiTrust*, the problems are mainly in the area of poor scan quality, the result of the mass-digitization process itself. For *Google Books*, the problems are mainly in the area of poorly aggregated keywords and OCR text. Studies regarding the copyright status of international books within Massive Digital Libraries should be undertaken to better provide overall coverage of public domain books without fear of litigation.

This article is also meant to indicate where major flaws in the mass digitization of books occur and provide a warning for future MDL projects. In particular, the problem areas of scan quality control, the use of non-Western bound books and non-English OCR should be examined with an eye to improve upon the digitization output so that books in

languages that use non-Roman-letter based materials can be better preserved and accessed.

The issue ultimately becomes a wider philosophical one as well. The digital text becomes "detached from its library, detached from its...past, it escapes from any necessary context, becoming... ephemeral, ubiquitous, insubstantial, available, valueless, free." [13] These missing or erased contexts may be destroying history and the physical book along with it. The texts themselves, sampled with all their evident mistakes and frayed tethers to physical reality, seem to support this assertion.

### ACKNOWLEDGMENTS

Andrew Weiss thanks Luiz Mendes, for his role in facilitating the generation of metadata lists from OCLC; and Momoku Uchimura, for her assistance in searching and tabulating results of book searches in *HathiTrust* and *Google Books*.

### REFERENCES

- [1] A. Weiss and R. James, *Examining Massive Digital Libraries: A LITA Guide*, Chicago: Neil-Schuman Publishing, 2014.
- [2] A. Weiss and R. James, "An Examination of Massive Digital Libraries' Coverage of Spanish Language Materials: Issues of Multi-lingual Accessibility in a Decentralized, Mass-Digitized World," *Culture and Computing (Culture Computing)*, 2013 International Conference on , vol., no., pp.10,14, 16-18 Sept. 2013 doi: 10.1109/CultureComputing.2013.10
- [3] G. Nunberg, *Google Books: A Metadata Train Wreck. Language Log*. 2009. Retrieved from <http://languageblog.ldc.upenn.edu/nll/?p=1701>
- [4] G. Nunberg, "Google's Book Search: A Disaster for Scholars." *The Chronicle Review*. 2009. Retrieved from <http://chronicle.com/article/Googles-Book-Search-A/48245/>
- [5] R. James and A. Weiss, "An assessment of *Google Books*' metadata," *Journal of Library Metadata*, vol. 12, Iss:1, pp. 15-22, February 2012.
- [6] R. James, "An assessment of the legibility of *Google Books*," *Journal of Access Services*, Vol.7, Iss:4, October 2010.
- [7] A. Weiss and R. James, "Assessing the coverage of Hawaiian and Pacific books in the *Google Books* digitization project," *OCLC Systems and Services*, vol. 29 Iss:1, pp.13-21, January 2013.
- [8] *HathiTrust Languages. HathiTrust*. 2015. Retrieved from [http://www.hathitrust.org/visualizations\\_languages](http://www.hathitrust.org/visualizations_languages)
- [9] *Authors Guild v. Google*, 2d Cir. 2013. Retrieved from <https://www.documentcloud.org/documents/834877-google-books-ruling-on-fair-use.html>
- [10] *Authors Guild v. HathiTrust*, 2d Cir. 2014. Retrieved from [https://www.eff.org/files/hathitrust\\_decision\\_copy\\_2.pdf](https://www.eff.org/files/hathitrust_decision_copy_2.pdf)
- [11] M. Barclay and C. McSherry, "Digitizing Books Is Fair Use: Author's Guild v. *HathiTrust*", Electronic Frontier Foundation, 2012. Retrieved from <https://www.eff.org/deeplinks/2012/10/authors-guild-v-hathitrustdecision>
- [12] J. Jeanneney, *Google and The Myth of Universal Knowledge*, Chicago, University of Chicago Press, 2006.
- [13] R. Szpiech, "The Dagger of Faith in the Digital Age: A vitriolic medieval manuscript illuminates how Google is destroying the act of reading," October 7, 2014. *Tablet* Retrieved from <http://tabletmag.com/jewish-arts-and-culture/books/183443/dagger-digital-age?all=1>