

# Analysis of the entire HathiTrust Digital Library Collection by Source Institution

---

March 4, 2020

*by Tom Teper and Vera Kuipers*

Library Faculty Meeting



Office of Collections and Technical Services  
University Library at the University of Illinois at Urbana-Champaign

# Research Questions



Dataset



Tools



Data VIZs

a) What **Institutions** deposited their digitized volumes?

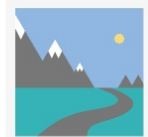
*Number of deposited volumes by Institution, State, and Country*

b) In which **Region** and **Country / State** were the majority of the digitized volumes published?

*Overall / by Institution*

c) In which **Languages** were the majority of the digitized volumes published?

*Overall / by Institution*



# Approach

## general workflow of the research

- Obtain the dataset
- Import the dataset to the database
- Decide on what metadata is needed for the analysis
- Query data to evaluate how much data cleaning needs to be done
- Decide whether additional data ("dictionaries") are needed & where to obtain them
- Perform data cleaning & manipulation (SQL)
- Design various data visualizations and dashboards (Tableau)
- Present the result -> get a feedback -> make corrections -> publish VIZ on Tableau Server for everyone to explore

# Dataset Overview

Collection of Digitized Items:  
17.1 M / 847 K records datasets

The screenshot shows the HathiTrust website home page. The header includes the HathiTrust logo, a search bar, and navigation links. The main content area features a 'Hathifiles' section with a description of the collection and a list of files provided below. The 'Hathifiles' section is highlighted with an orange box. The 'Files provided below' section is also highlighted with an orange box. The 'Our Digital Library' section is visible on the right side of the page.

**Hathifiles**

The "Hathifiles" are tab-delimited text files that describe every item in the HathiTrust collection. They include information derived from the bibliographic record (e.g., title, publisher, language, commonly used identifiers, etc.), rights and access codes, and information about the source of the item.

A description of the fields included in the hathifiles as well as potential use cases is provided in the "Hathifiles Description" page.

**Files provided below**

A **monthly file** is uploaded on the first of every month with a row for every item that is in the HathiTrust collection at the moment the file is created. The filename begins with "hathi\_full.". These files tend to be large and may be difficult to open with standard spreadsheet software or text editors. You may need to work with the files programmatically (e.g., using Python to extract desired data).

An **update file** is uploaded every day and contains a row for every item that has changed in the previous 24 hours. The filename begins with "hathi\_upd.". Items are included in the update files if any of the following has occurred: the item was newly deposited into the collection, a new copy of the digital item overrode the previous copy, the rights and access status has changed, or a new bibliographic record was provided by the contributor.

**Our Digital Library**

- Searching, Reading, Building Collections, and Accessibility
- Technology, Standards, and Specifications
- Getting Content into HathiTrust
- HathiTrust Metadata Management System
- Data Availability and APIs
  - HathiTrust Bibliographic API
  - HathiTrust Data API
- Hathifiles
- Hathifiles Description
- Datasets
- HathiTrust OAI feed

The screenshot shows the 'Files provided below' section of the HathiTrust website. It contains a table listing the files provided, including their display name, creation date, size, modification date, and mime type. The table is highlighted with an orange box. The 'Files provided below' section is also highlighted with an orange box. The 'Our Digital Library' section is visible on the right side of the page.

**Files provided below**

A **monthly file** is uploaded on the first of every month with a row for every item that is in the HathiTrust collection at the moment the file is created. The filename begins with "hathi\_full.". These files tend to be large and may be difficult to open with standard spreadsheet software or text editors. You may need to work with the files programmatically (e.g., using Python to extract desired data).

An **update file** is uploaded every day and contains a row for every item that has changed in the previous 24 hours. The filename begins with "hathi\_upd.". Items are included in the update files if any of the following has occurred: the item was newly deposited into the collection, a new copy of the digital item overrode the previous copy, the rights and access status has changed, or a new bibliographic record was provided by the contributor.

A **"header" file** is also included below. This file contains one row of labels for the data elements included in the hathifiles. It can be combined with the regular hathifiles for ease of working with the data. This header file is only updated when a new data element is added to the hathifiles.

Display name	created	size	modified	Mime type
<a href="#">hathi_full_list.txt</a>	February 1, 2020	307 bytes	February 1, 2020	text/plain
<a href="#">hathi_full_list.json</a>	February 8, 2020	12.4 KB	February 8, 2020	application/octet-stream
<a href="#">hathi_full_20200101.txt.gz</a>	January 1, 2020	986.26 MB	January 1, 2020	application/octet-stream
<a href="#">hathi_full_20200201.txt.gz</a>	February 1, 2020	992.52 MB	February 1, 2020	application/octet-stream
<a href="#">hathi_upd_20200101.txt.gz</a>	January 2, 2020	1.19 MB	January 2, 2020	application/octet-stream
<a href="#">hathi_upd_20200102.txt.gz</a>	January 3, 2020	1.17 MB	January 3, 2020	application/octet-stream
<a href="#">hathi_upd_20200103.txt.gz</a>	January 4, 2020	1.75 MB	January 4, 2020	application/octet-stream

**Our Digital Library**


- Searching, Reading, Building Collections, and Accessibility
- Technology, Standards, and Specifications
- Getting Content into HathiTrust
- HathiTrust Metadata Management System
- Data Availability and APIs
  - HathiTrust Bibliographic API
  - HathiTrust Data API
- Hathifiles
- Hathifiles Description
- Datasets
- HathiTrust OAI feed



Hathifiles | www.hathitrust.org

hathitrust.org/hathifiles

HomeAboutCollectionsHelpFeedbackLOG IN



Search words about or within the items

All Fields

Search HathiTrust

☒ Full-text☐ Catalog☒ Full view only

[Advanced full-text search](#)[Advanced catalog search](#)[Search tips](#)

[Our Membership](#)[Our Digital Library](#)[Our Collaborative Programs](#)[Our Research Center](#)[Help](#)  
[Welcome to HathiTrust!](#)

About > Hathifiles

## Hathifiles

The "Hathifiles" are tab-delimited text files that describe every item in the HathiTrust collection. They include information derived from the bibliographic record (e.g., title, publisher, language, commonly used identifiers, etc.), rights and access codes, and information about the source of the item.

A description of the fields included in the hathifiles as well as potential use cases is provided in the "[Hathifiles Description](#)" page.

### Files provided below

A **monthly file** is uploaded on the first of every month with a row for every item that is in the HathiTrust collection at the moment the file is created. The filename begins with "hathi\_full\_". These files tend to be large and may be difficult to open with standard spreadsheet software or text editors. You may need to work with the files programmatically (e.g., using Python to extract desired data).

An **update file** is uploaded every day and contains a row for every item that has changed in the previous 24 hours. The filename begins with "hathi\_upd\_". Items are included in the update files if any of the following has occurred: the item was newly deposited into the collection, a new copy of the digital item overrode the previous copy, the rights and access status has changed, or a new bibliographic record was provided by the contributor.

Search websiteFind

### Our Digital Library

- Searching, Reading, Building Collections, and Accessibility
  - Technology, Standards, and Specifications
- Getting Content Into HathiTrust
  - HathiTrust Metadata Management System
- Data Availability and APIs
  - HathiTrust Bibliographic API
  - HathiTrust Data API
  - Hathifiles
  - Hathifiles Description
  - Datasets
  - HathiTrust Catalog

<https://www.hathitrust.org/filebrowser/download/269539>

Hathifiles | www.hathitrust.org








hathitrust.org/hathifiles

### Files provided below

A **monthly file** is uploaded on the first of every month with a row for every item that is in the HathiTrust collection at the moment the file is created. The filename begins with "hathi\_full\_". These files tend to be large and may be difficult to open with standard spreadsheet software or text editors. You may need to work with the files programmatically (e.g., using Python to extract desired data).

An **update file** is uploaded every day and contains a row for every item that has changed in the previous 24 hours. The filename begins with "hathi\_upd\_". Items are included in the update files if any of the following has occurred: the item was newly deposited into the collection, a new copy of the digital item overrode the previous copy, the rights and access status has changed, or a new bibliographic record was provided by the contributor.

A **"header" file** is also included below. This file contains one row of labels for the data elements included in the hathifiles. It can be combined with the regular hathifiles for ease of working with the data. This header file is only updated when a new data element is added to the hathifiles.

Display name	created	size	modified	Mime type
 <a href="#">hathi_field_list.txt</a>	February 1, 2020	307 bytes	February 1, 2020	text/plain
 <a href="#">hathi_file_list.json</a>	February 8, 2020	12.4 KB	February 8, 2020	application/octet-stream
 <a href="#">hathi_full_20200101.txt.gz</a>	January 1, 2020	986.26 MB	January 1, 2020	application/octet-stream
 <a href="#">hathi_full_20200201.txt.gz</a>	February 1, 2020	992.52 MB	February 1, 2020	application/octet-stream
 <a href="#">hathi_upd_20200101.txt.gz</a>	January 2, 2020	1.19 MB	January 2, 2020	application/octet-stream
 <a href="#">hathi_upd_20200102.txt.gz</a>	January 3, 2020	1.17 MB	January 3, 2020	application/octet-stream
 <a href="#">hathi_upd_20200103.txt.gz</a>	January 4, 2020	1.75 MB	January 4, 2020	application/octet-stream

- Management System
- Data Availability and APIs
  - HathiTrust Bibliographic
    - API
    - HathiTrust Data API
    - Hathifiles
    - Hathifiles Description
    - Datasets
    - HathiTrust OAI feed
- Policies
- Statistics and Visualizations
- HathiTrust Personas

# Tools

---

## Data

**Manipulation**

**Cleaning**

**Analysis**

**& Visualization**



**MS SQL Server**

---



**SQLiteStudio**

---



**MS Excel / Office 360**

---



**OpenRefine**

---



**Tableau Desktop**

---


# Data Cleaning & Manipulation

## Publication Place & Language

MARC Code List for Countries: x +

loc.gov/marc/countries/

The Library of Congress >> Especially for Librarians and Archivists >> Standards >> MARC 21



**MARC Code List for COUNTRIES**  
Library of Congress  
Network Development and MARC Standards Office

**2003 Edition**

[Last updated April 4, 2008 to incorporate 12 additions]

---

**Table of Contents**

- [Introduction](#) (PDF)
  - General Information
  - Structure and Assignment of Codes
  - Arrangement of the List
  - Changes in the Codes
  - MARC 21 Formats and Fields in which Codes are Used
  - Code Maintenance
- [Name Sequence](#) (HTML)
- [Code Sequence](#) (HTML)
- [Regional Sequence](#) (HTML)
- [Changes in 2003 printed edition](#) (PDF)
- [Changes to online list since 2003 printed edition](#) (HTML)

An XML version is available for use in applications at: [www.loc.gov/standards/codelists/countries.xml](http://www.loc.gov/standards/codelists/countries.xml)


---

[MARC 21 HOME](#) >> MARC Code List for Countries

The Library of Congress >> Especially for Librarians and Archivists >> Standards (02/24/2012)

MARC Code List for Language: x +

loc.gov/marc/languages/



**MARC Code List for LANGUAGES**  
Library of Congress  
Network Development and MARC Standards Office

**2007 Edition**

[Updated continuously to incorporate additions and changes (HTML only).]

---

**Table of Contents**

- [Introduction](#) (PDF)
  - Changes in 2007 Edition
  - Structure of the Codes
  - Relationship to ISO 639-2
  - Individual Language Codes
  - Language Group Codes
  - Names of Languages and Dialects
  - Special Codes for Special Situations
  - Arrangement of the List
  - Use of Language Names in Cataloging
  - MARC 21 Fields in which Codes are Used
  - Publication Background
  - Code Maintenance
  - Related MARC 21 Documents
  - Other Related Documents
- [Name Sequence](#) (HTML)
- [Code Sequence](#) (HTML)
- [Changes in 2007 edition](#) (PDF)
- [Changes since the 2007 edition](#): Summarizes code changes, additions, deletions and revisions to language captions.

An XML version is available for use in applications at: [www.loc.gov/standards/codelists/languages.xml](http://www.loc.gov/standards/codelists/languages.xml)

An RDF version is also available at: [id.loc.gov/vocabulary/languages](http://id.loc.gov/vocabulary/languages)

---

[MARC 21 HOME](#) >> MARC Code List for Languages

The Library of Congress >> Especially for Librarians and Archivists >> Standards (09/08/2011)

[The Library of Congress](#) >> [Especially for Librarians and Archivists](#) >> [Standards](#) >> [MARC 21](#)



## 2003 Edition

[Last updated April 4, 2008 to incorporate 12 additions]

### Table of Contents

- [Introduction](#) (PDF)
  - General Information
  - Structure and Assignment of Codes
  - Arrangement of the List
  - Changes in the Codes
  - MARC 21 Formats and Fields in which Codes are Used
  - Code Maintenance
- [Name Sequence](#) (HTML)
- [Code Sequence](#) (HTML)
- [Regional Sequence](#) (HTML)
- [Changes in 2003 printed edition](#) (PDF)
- [Changes to online list since 2003 printed edition](#) (HTML)

An XML version is available for use in applications at: [www.loc.gov/standards/codelists/countries.xml](http://www.loc.gov/standards/codelists/countries.xml)

[MARC 21 HOME](#) >> MARC Code List for Countries

[The Library of Congress](#) >> [Especially for Librarians and Archivists](#) >> [Standards](#)  
( 02/24/2012 )

- [Introduction](#)
  - Chan
  - Struc
  - Relat
  - Indiv
  - Langu
  - Name
  - Speci
  - Arran
  - Use o
  - MAR
  - Publi
  - Code
  - Relat
  - Other
- [Name Sequen](#)
- [Code Sequen](#)
- [Changes in 2](#)
- [Changes sinc](#)

An XML version is


An RDF version is

[MARC 21 HOME](#) :

[The Library of Con](#)  
( 09/08/2011 )

MARC Code List for Languages x +

← → ↻ ⓘ Not Secure | loc.gov/marc/languages/



**MARC Code List for  
LANGUAGES**  
*Library of Congress  
Network Development and MARC Standards Office*

**2007 Edition**  
  
[Updated continuously to incorporate additions and changes (HTML only).]

**Table of Contents**

- [Introduction](#) (PDF)
  - Changes in 2007 Edition
  - Structure of the Codes
  - Relationship to ISO 639-2
  - Individual Language Codes
  - Language Group Codes
  - Names of Languages and Dialects
  - Special Codes for Special Situations
  - Arrangement of the List
  - Use of Language Names in Cataloging
  - MARC 21 Fields in which Codes are Used
  - Publication Background
  - Code Maintenance
  - Related MARC 21 Documents
  - Other Related Documents
- [Name Sequence](#) (HTML)
- [Code Sequence](#) (HTML)
- [Changes in 2007 edition](#) (PDF)
- [Changes since the 2007 edition](#): Summarizes code changes, additions, deletions and revisions to language captions.

An XML version is available for use in applications at: [www.loc.gov/standards/codelists/languages.xml](http://www.loc.gov/standards/codelists/languages.xml)

An RDF version is also available at: [id.loc.gov/vocabulary/languages](http://id.loc.gov/vocabulary/languages)

[MARC 21 HOME](#) >> MARC Code List for Languages

[The Library of Congress](#) >> [Especially for Librarians and Archivists](#) >> [Standards](#)  
( 09/08/2011 )




MARC Code List for Countries: x

+

loc.gov/marc/countries/countries\_code.html

The Library of Congress >> Librarians, Archivists >> Standards

 **MARC STANDARDS**

MARC 21 HOME >> MARC 21 Code List for Countries >> Code Sequence

MARC Code List for Countries

Part 2: Code Sequence

A B C D E F G H I J K L M N O P Q R S T U V

Discontinued codes are identified by a hyphen preceding the first letter in the code string.

aa	Albania
abc	Alberta
-ac	Ashmore and Cartier Islands
aca	Australian Capital Territory
ae	Algeria
af	Afghanistan
ag	Argentina
-ai	Anguilla
ai	Armenia (Republic)
-air	Armenian S.S.R.
aj	Azerbaijan
-ajr	Azerbaijan S.S.R.
aku	Alaska
alu	Alabama
am	Anguilla
an	Andorra
ao	Angola
aq	Antigua and Barbuda
aru	Arkansas
as	American Samoa
at	Australia
au	Austria
aw	Aruba
ay	Antarctica
azu	Arizona

MARC Code List for Countries: x

+

loc.gov/marc/countries/countries\_name.html

Arizona [azu]

Arkansas [aru]

Armenia (Republic) [ai]

[Coded [air] (Armenian S.S.R.) before June 1992]

UF Armenian S.S.R.

Armenian S.S.R.

[Coded [air] (Armenian S.S.R.) before June 1992]

USE Armenia (Republic)

Armo (Atoll)

USE Marshall Islands

Arquipélago dos Bijagós

USE Guinea-Bissau

Aruba [aw]

[Coded [na] (Netherlands Antilles) before Mar. 1988]

UF Netherlands Antilles

Ascension Island (Atlantic Ocean)

USE Saint Helena

Ascension Island (Micronesia)

USE Micronesia (Federated States)

Ashanti

USE Ghana

Ashmore and Cartier Islands

[Coded [ac] (Ashmore and Cartier Islands) before Jan. 1978]

USE Australia

Asuncion Island

USE Northern Mariana Islands

Atafu Atoll

USE Tokelau

Atauro, Ilha de

[Coded [pt] (Portuguese Timor) before Jan. 1978]

USE Timor-Leste

Austral Islands

USE French Polynesia

Australia [at]

UF Ashmore and Cartier Islands

Macquarie Island

Australian Capital Territory [aca]

[Coded [at] (Australia) before Sept. 2005]

Austria [au]

Azerbaijan [aj]

[Coded [ajr] (Azerbaijan S.S.R.) before June 1992]

UF Azerbaijan S.S.R.

Azerbaijan S.S.R.

[Coded [ajr] (Azerbaijan S.S.R.) before June 1992]

USE Azerbaijan

Azores

[The Library of Congress](#) >> [Librarians, Archivists](#) >> [Standards](#)

## MARC STANDARDS

[MARC 21 HOME](#) >> [MARC 21 Code List for Countries](#) >> **Regional Sequence**

### MARC Code List for Countries Part 3: Regional Sequence

- [Africa](#)
- [Asia](#)
- [Atlantic Ocean](#)
- [Australasia](#)
- [Central America](#)
- [Europe](#)
- [Indian Ocean](#)
- [North America](#)
- [Pacific Ocean](#)
- [South America](#)
- [West Indies](#)
- [Other](#)

#### Africa

ae	Algeria
ao	Angola
dm	Benin
bs	Botswana
uv	Burkina Faso
bd	Burundi
cm	Cameroon
cx	Central African Republic
cd	Chad
cf	Congo (Brazzaville)
cg	Congo (Democratic Republic)
iv	Côte d'Ivoire
ft	Djibouti
ua	Egypt
eg	Equatorial Guinea
ea	Eritrea
et	Ethiopia
go	Gabon
gm	Gambia
gh	Ghana
gv	Guinea
pg	Guinea-Bissau
kn	Kenya

#### Atlantic Ocean

bm	Bermuda Islands
bv	Bouvet Island
cv	Cabo Verde
fa	Faroe Islands
fk	Falkland Islands
xj	Saint Helena
xs	South Georgia and the South Sandwich Islands

#### Australasia

aca	Australian Capital Territory
at	Australia
nz	New Zealand
nx	Norfolk Island
qea	Queensland
tma	Tasmania
vra	Victoria
wea	Western Australia
xga	Coral Sea Islands Territory
xna	New South Wales
xoa	Northern Territory
xra	South Australia

#### Central America

bh	Belize
cr	Costa Rica
es	El Salvador
gt	Guatemala
ho	Honduras
nq	Nicaragua
pn	Panama

#### Europe

aa	Albania
an	Andorra
au	Austria
bw	Belarus
be	Belgium
bn	Bosnia and Herzegovina
bu	Bulgaria
ci	Croatia
xr	Czech Republic
..	..



## Data Cleaning and Manipulation using SQL

1. Created 3 dictionaries based on MARC Code Lists for Countries and Languages:

- for countries
- for states
- for languages

```
select * from Dictionary_Lang
select * from Dictionary_Country
select * from Dictionary_State
```

Lang_Code	Lang_Name
aar	Afar
abk	Abkhaz
ace	Achinese
ach	Acoli
ada	Adangme
ady	Adyghe
afa	Afroasiatic (Other)
afh	Afarsi (Afghani language)
afk	Afrikaans
ain	Ainu

Country_Code	Country	Region
al	Albania	Europe
ae	Algeria	Africa
af	Afghanistan	Asia
ag	Argentina	South America
ai	Armenia	Asia
az	Azerbaijan	Asia
am	Angola	West Indies
an	Andorra	Europe
ao	Angola	Africa
aq	Antigua and Barbuda	West Indies

Country	State_Code	State	Region
United States	alu	Alabama	North America
United States	aku	Alaska	North America
United States	aru	Arkansas	North America
United States	azu	Arizona	North America
United States	cau	California	North America
United States	cou	Colorado	North America
United States	dou	District of Columbia	North America
United States	deu	Delaware	North America
United States	ctu	Connecticut	North America
United States	flu	Florida	North America
United States	hlu	Hawaii	North America
United States	lau	Iowa	North America
United States	idu	Idaho	North America
United States	gau	Georgia	North America

Query executed successfully.

2. Assigned new codes to the discontinued ones

3. Removed spaces and hashes at the of the codes in the dataset

```
select htid,
pub_place,
pub_place_clean,
lang,
lang_clean
from HashTrust_full_20200101
where pub_place != pub_place_clean or
lang != lang_clean
```

htid	pub_place	pub_place_clean	lang	lang_clean
unv.x000538617	cs	sr	ger	ger
slug.3011209591532	cs	sr	cze	cze
slug.3011209591540	cs	sr	cze	cze
pat.000002255316	cs	sr	cze	cze
unv.x000400566	cs	sr	cze	cze
unv.319510014970537	xx#	xx	eng	eng
unv.x004219569	xx#	xx	eng	eng
unv.31951002437443	fr	fr	fre	fre
unv.x001397369	fr	fr	fre	fre
unv.x030683214	ru	ru	slv	slv
coo.31924022068823	ge	gw	ger	ger
coo.31924022068831	ge	gw	ger	ger
coo.31924022068849	ge	gw	ger	ger
coo.31924022068856	ge	gw	ger	ger
unv.x004618911	ge	gw	ger	ger
unv.3195100132223b	ge#	gw	ger	ger
unv.31951002327619a	ge#	gw	ger	ger
unv.31951002327619b	ge#	gw	ger	ger
unv.31951002327616a	ge#	gw	ger	ger
unv.x000384153	ge#	gw	ger	ger
unv.x001227455	cs	sr	cze	cze
unv.x002675591	gw	gw	slv	slv
unv.x000680106	cs	sr	lat	lat
coo.31924026978290	cs	sr	cze	cze
unv.x000175784	cs	sr	cze	cze
unv.x000327253	uk	gg	eng	eng
unv.x000376024	cs	sr	ger	ger
unv.x000376023	cs	sr	ger	ger
unv.x001704892	cs	sr	cze	cze
hvd.hvdgdy	ru	ru	rus	rus
unv.x001017695	ru	ru	rus	rus
unv.31951002050563	fr	fr	fre	fre

Query executed successfully.

```

select * from Dictionary_Lang
select * from Dictionary_Country
select * from Dictionary_State

```

100 %

Results Messages

	Lang_Code	Lang_Name
1	aar	Afar
2	abk	Abkhaz
3	ace	Achinese
4	ach	Acoli
5	ada	Adangme
6	ady	Adygei
7	afa	Afroasiatic (Other)
8	afh	Afrihili (Artificial language)
9	afz	Afrikaans
10	ain	Ainu

	Country_Code	Country	Region
1	aa	Albania	Europe
2	ae	Algeria	Africa
3	af	Afghanistan	Asia
4	ag	Argentina	South America
5	ai	Armenia	Asia
6	aj	Azerbaijan	Asia
7	am	Anguilla	West Indies
8	an	Andorra	Europe
9	ao	Angola	Africa
10	aq	Antigua and Barbuda	West Indies

	Country	State_Code	State	Region
1	United States	alu	Alabama	North America
2	United States	aku	Alaska	North America
3	United States	aru	Arkansas	North America
4	United States	azu	Arizona	North America
5	United States	cau	California	North America
6	United States	cou	Colorado	North America
7	United States	dcu	District of Columbia	North America
8	United States	deu	Delaware	North America
9	United States	ctu	Connecticut	North America
10	United States	flu	Florida	North America
11	United States	hiu	Hawaii	North America
12	United States	iau	Iowa	North America
13	United States	idu	Idaho	North America
14	United States	gau	Georgia	North America

Query executed successfully.

```

select
    htid,
    pub_place,
    pub_place_clean,
    lang,
    lang_clean
from HathiTrust_full_20200101
where pub_place != pub_place_clean or
      lang != lang_clean

```

	htid	pub_place	pub_place_clean	lang	lang_clean
1	uva.x000538617	cs	xr	ger	ger
2	uiug.30112093591532	cs	xr	cze	cze
3	uiug.30112093591540	cs	xr	cze	cze
4	pst.000002255316	cs	xr	cze	cze
5	uva.x000400566	cs	xr	cze	cze
6	umn.319510014978537	xx#	xx	eng	eng
7	uva.x004215959	xx#	xx	eng	eng
8	umn.31951002437443i	fr#	fr	fre	fre
9	uva.x001397369	fr#	fr	fre	fre
10	uva.x030683214	yu	rb	slv	slv
11	coo.31924022068823	ge	gw	ger	ger
12	coo.31924022068831	ge	gw	ger	ger
13	coo.31924022068849	ge	gw	ger	ger
14	coo.31924022068856	ge	gw	ger	ger
15	uva.x004618911	ge	gw	ger	ger
16	umn.31951p00132223o	ge#	gw	ger	ger
17	umn.31951002327619k	ge#	gw	ger	ger
18	umn.31951002327618m	ge#	gw	ger	ger
19	umn.31951002327616q	ge#	gw	ger	ger
20	uva.x000384153	ge#	gw	ger	ger
21	uva.x001227455	cs	xr	cze	cze
22	uva.x002675591	gw	gw	scc	srp
23	uva.x000680106	cs	xr	lat	lat
24	coo.31924026978290	cs	xr	cze	cze
25	uva.x000175784	cs	xr	cze	cze
26	uva.x000327253	ulk	gg	eng	eng
27	uva.x030376924	cs	xr	ger	ger
28	uva.x030376923	cs	xr	ger	ger
29	uva.x001704892	cs	xr	cze	cze
30	hvd.hwbgy	rur	ru	rus	rus
31	uva.x001017695	rur	ru	rus	rus
32	umn.319510020050569	fr#	fr	fre	fre

Query executed successfully.

# Data Cleaning & Manipulation

## Source Institution

select \* from HathiTrust\_full\_20200101

id	access	rights	h_bib_key	description	source	source_bib_num	oclc_num	isbn	ismn	icon	title	imprint
1	hvd:mfjg	allow	pd	009777331	NULL	HVD	001437965	7787758	NULL	98000841	Friendship, by Hugh Black. With an introductory...	F.H. Revell, 1898.
2	uc2.afk./1396016zw1d09e	allow	pd	009777331	NULL	UK	b-12063116a	7787758	NULL	98000841	Friendship, by Hugh Black. With an introductory...	F.H. Revell, 1898.
3	uva.x030803302	allow	pd	009777332	NULL	UVA	u1170865	18232559	NULL	01011878	The greatness of patience.	T. Y. Crowell & co. [c-1901]
4	uva.x000305165	allow	pd	009777333	v 2 1916	UVA	u1170887	692267559	NULL	16008093	Character, how to strengthen it.	Funk & Wagnalls company, 1916.
5	hvd:hmgfc	allow	pd	009777334	NULL	HVD	001345146	1436826	NULL	16008097	Timidity, how to overcome it, by Yortomo-Taehi...	Funk & Wagnalls company, 1916.
6	uva.x004490145	allow	pd	009777334	v 3 1916	UVA	u1170888	1436826	NULL	16008097	Timidity, how to overcome it, by Yortomo-Taehi...	Funk & Wagnalls company, 1916.
7	uva.x0045080509	allow	pd	009777335	v 10 1916	UVA	u1170895	15225962	NULL	16008095	Personality, how to build it.	Funk & Wagnalls company, 1916.
8	hvd:hn1o1	allow	pd	009777335	NULL	HVD	001580549	15225962	NULL	16008095	Personality, how to build it.	Funk & Wagnalls company, 1916.
9	uva.x030803147	allow	pd	009777336	NULL	UVA	u1170922	669874229	NULL	08013240	The gentle life.	S. Low, Manston, Searle, and Rivington, 1
10	uva.x000241378	allow	pd	009777337	NULL	UVA	u1170927	703550607	NULL	20005340	Every-day topics.	Scibner, Armstrong and company, 1876.
11	uva.x030803150	allow	pd	009777338	NULL	UVA	u1170930	7250765	NULL	36015857	Gold foil, hammered from popular proverbs.	Scibner, Armstrong & co., 1875.
12	uva.x002140318	allow	pd	009777339	NULL	UVA	u1170933	5217299	NULL	unk-82007186	The egotist...	1880.
13	uva.x030803148	allow	pd	009777340	NULL	UVA	u1170935	1872243	NULL	10012266	The making of a man.	Cassell publishing company [1892]
14	uva.x000529100	allow	pd	009777341	NULL	UVA	u1170936	11922503	NULL	10012404	How to observe. Morals and manners.	C. Knight and co., 1838.
15	hvd:hwdff	allow	pd	009777342	NULL	HVD	002244590	3330557	0837071798 (microfiche)	12037081	The building of character / by J.R. Miller.	Thomas Y. Crowell, c1894.
16	loc.afk./1396019q257w2w	allow	pd	009777342	NULL	LOC	7763936	3330557	0837071798 (microfiche)	12037081	The building of character / by J.R. Miller.	Thomas Y. Crowell, c1894.
17	uva.x002403914	allow	pd	009777342	NULL	UVA	u1170937	3330557	0837071798 (microfiche)	12037081	The building of character / by J.R. Miller.	Thomas Y. Crowell, c1894.
18	uva.x004613363	allow	pd	009777343	NULL	UVA	u1170942	4155754	NULL	12037570	Through my spectacles.	T. Y. Crowell & company [c1898]
19	uva.x030803182	allow	pd	009777344	NULL	UVA	u1170947	5206521	NULL	15003385	When thou hast shut thy door. A book for the still...	Plimng H. Revell company [c1895]
20	hvd:sh4oge	allow	pd	009777345	NULL	HVD	002566721	845862	NULL	10030166 / ...	The Trinity, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1910.
21	uva.x002027892	allow	pd	009777345	v 4	UVA	u1171010	845862	NULL	10030166 / ...	The Trinity, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1910.
22	yale.39002029677235	allow	pd	009777346	NULL	YALE	3629339	509607	NULL	15023098	The incarnation, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
23	hvd:sh4ogh	allow	pd	009777346	NULL	HVD	002565926	509607	NULL	15023098	The incarnation, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
24	uva.x030741716	allow	pd	009777346	v 6	UVA	u1171011	509607	NULL	15023098	The incarnation, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
25	uva.x030735932	allow	pd	009777347	NULL	UVA	u1171025	3808443	NULL	25004656	Elements of divinity.	Morton & Griswold, 1847.
26	uva.x030735934	allow	pd	009777348	NULL	UVA	u1171029	13579294	NULL	02062079	Some features of faith; a popular discussion of ce...	The Young churchman co., 1902.
27	uva.x000531376	allow	pd	009777349	NULL	UVA	u1171034	614135288	NULL	08030158	Outlines of systematic theology.	American Baptist publication society, 1901
28	uva.x030735955	allow	pd	009777350	NULL	UVA	u1171036	3010360	NULL	39007848	Theological colloquies.	Lewis & Coleman, 1837.
29	uva.x030735956	allow	pd	009777351	NULL	UVA	u1171040	669874243	NULL	39007041	Outlines of Christian theology.	T. Whittaker, 1894.
30	uva.x030804130	allow	pd	009777352	NULL	UVA	u1171065	3760124	NULL	01029439	The theological compend: containing a system of...	Publishing house of the Methodist Episcopal...
31	uva.x030804135	allow	pd	009777353	NULL	UVA	u1171082	1823813	NULL	06009692	Christian doctrines: a compendium of theology.	American Baptist publication society 1878
32	uva.x030808059	allow	pd	009777354	NULL	UVA	u1171102	181822160	NULL	07028761	Old and new theology.	T. & T. Clark, 1885.
33	uva.x000736296	allow	pd	009777355	NULL	UVA	u1171144	70350644	NULL	32017455	The Bible, the church, and the reason.	C. Scribner's sons, 1893.
34	uva.x000430790	allow	pd	009777356	NULL	UVA	u1171171	7490799	NULL	13002018	The evolution of the kingdom.	C. C. Cook, [etc. etc. c1913]
35	uva.x030806027	allow	pd	009777357	NULL	UVA	u1171178	12852810	NULL	21000463	Why do things happen?	The Nunc licet press [c1920]
36	uva.x030809116	allow	pd	009777358	NULL	UVA	u1171287	51675160	08011	48190829	Microcosm and Macrocosm: lectures deliv...	Clark, Dover & co., 1916

Created a dictionary for the Source code (original source of content). Assigned Institution name, city, state, and country to the codes.

select \* from Dictionary\_Institution

Source_Code	Source_Name	Source_Country	Source_State	Source_City
1	AEU	University of Alberta	Canada	Alberta
2	AZTES	Arizona State University	United States	Arizona
3	AuBU	University of Queensland	Australia	Queensland
4	CHI	University of Chicago	United States	Illinois
5	CMALG	Getty Research Institute	United States	California
6	COO	Cornell University	United States	New York
7	OU	University of Connecticut	United States	Connecticut
8	DUL	Duke University	United States	North Carolina
9	DeU	University of Delaware	United States	Delaware
10	FMU	University of Miami	United States	Florida
11	FU	University of Florida	United States	Florida
12	GEU	Emory University	United States	Georgia
13	HVD	Harvard University	United States	Massachusetts
14	IEU	Northwestern University	United States	Illinois
15	IND	University of Notre Dame	United States	Indiana
16	INU	Indiana University	United States	Indiana
17	IaU	University of Iowa	United States	Iowa
18	JTKU	Keio University	Japan	Tokyo
19	LOC	Library of Congress	United States	District of Columbia
20	MCHB	Boston College	United States	Massachusetts
21	MDBJ	Johns Hopkins University	United States	Maryland
22	MDL	Minnesota Digital Library	United States	Minnesota
23	MMeT	Tufts University	United States	Massachusetts
24	MU	University of Massachusetts	United States	Massachusetts
25	MWICA	Steele and Francine Clark Art Institute	United States	Massachusetts
26	MdU	University of Maryland	United States	Maryland
27	MEM	Michigan State University	United States	Michigan
28	MIU	University of Michigan	United States	Michigan
29	MoU	University of Missouri	United States	Missouri
30	NCU	North Carolina State University	United States	North Carolina
31	NCWSW	Wake Forest University	United States	North Carolina
32	NJP	Princeton University	United States	New Jersey
33	NNC	Columbia University	United States	New York
34	NYP	New York Public Library	United States	New York
35	OU	Ohio State University	United States	Ohio
36	PHC	Haverford College	United States	Pennsylvania
37	PST	Pennsylvania State University	United States	Pennsylvania
38	PU	University of Pennsylvania	United States	Pennsylvania

Query executed successfully.



select \* from HathiTrust\_full\_20200101

100 %

Results Messages

	htid	access	rights	ht_bib_key	description	source	source_bib_num	oclc_num	isbn	issn	lcn	title	imprint
1	hvd.rsfgi	allow	pd	009777331	NULL	HVD	001437685	7787758	NULL	NULL	98000841	Friendship, by Hugh Black. With an introductory ...	F.H. Revell, 1898.
2	uc2.ark:/13960/t6zw1d09s	allow	pd	009777331	NULL	UC	b12068116x	7787758	NULL	NULL	98000841	Friendship, by Hugh Black. With an introductory ...	F.H. Revell, 1898.
3	uva.x030803302	allow	pd	009777332	NULL	UVA	u1170865	18232559	NULL	NULL	01011878	The greatness of patience,	T. Y. Crowell & co. [c1901]
4	uva.x000306166	allow	pd	009777333	v.2 1916	UVA	u1170887	692267559	NULL	NULL	16008093	Character, how to strengthen it,	Funk & Wagnalls company, 1916.
5	hvd.hnmg6z	allow	pd	009777334	NULL	HVD	001345146	1436826	NULL	NULL	16008097	Timidity, how to overcome it, by Yoritomo-Tashi, ...	Funk & Wagnalls company, 1916.
6	uva.x004490145	allow	pd	009777334	v.3 1916	UVA	u1170888	1436826	NULL	NULL	16008097	Timidity, how to overcome it, by Yoritomo-Tashi, ...	Funk & Wagnalls company, 1916.
7	uva.x004508509	allow	pd	009777335	v.10 1916	UVA	u1170895	15225962	NULL	NULL	16008095	Personality, how to build it,	Funk & Wagnalls company, 1916.
8	hvd.hn1ci1	allow	pd	009777335	NULL	HVD	001580549	15225962	NULL	NULL	16008095	Personality, how to build it,	Funk & Wagnalls company, 1916.
9	uva.x030803147	allow	pd	009777336	NULL	UVA	u1170922	669874229	NULL	NULL	08013240	The gentle life;	S. Low, Marston, Searle, and Rivington, 1
10	uva.x000241378	allow	pd	009777337	NULL	UVA	u1170927	703550607	NULL	NULL	20005340	Every-day topics;	Scribner, Armstrong and company, 1876.
11	uva.x030803150	allow	pd	009777338	NULL	UVA	u1170930	7250765	NULL	NULL	36015857	Gold-foil, hammered from popular proverbs,	Scribner, Armstrong & co, 1875.
12	uva.x002140318	allow	pd	009777339	NULL	UVA	u1170933	5217299	NULL	NULL	unk82007186	The egotist...	1880.
13	uva.x030803148	allow	pd	009777340	NULL	UVA	u1170935	1872243	NULL	NULL	10012266	The making of a man;	Cassell publishing company [1892]
14	uva.x000529100	allow	pd	009777341	NULL	UVA	u1170936	11922503	NULL	NULL	10012404	How to observe. Morals and manners.	C. Knight and co., 1838.
15	hvd.hwdif	allow	pd	009777342	NULL	HVD	002244590	3330557	0837071798 (microfiche)	NULL	12037081	The building of character / by J.R. Miller.	Thomas Y. Crowell, c1894.
16	loc.ark:/13960/t9q257w2w	allow	pd	009777342	NULL	LOC	7763936	3330557	0837071798 (microfiche)	NULL	12037081	The building of character / by J.R. Miller.	Thomas Y. Crowell, c1894.
17	uva.x002403914	allow	pd	009777342	NULL	UVA	u1170937	3330557	0837071798 (microfiche)	NULL	12037081	The building of character / by J.R. Miller.	Thomas Y. Crowell, c1894.
18	uva.x004613363	allow	pd	009777343	NULL	UVA	u1170942	4155754	NULL	NULL	12037570	Through my spectacles,	T. Y. Crowell & company [c1898]
19	uva.x030803182	allow	pd	009777344	NULL	UVA	u1170947	9206921	NULL	NULL	15003385	When thou hast shut thy door. A book for the still...	Fleming H. Revell company [c1895]
20	hvd.ah4qge	allow	pd	009777345	NULL	HVD	002566721	845862	NULL	NULL	10030166 /...	The Trinity, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1910.
21	uva.x002027892	allow	pd	009777345	v.4	UVA	u1171010	845862	NULL	NULL	10030166 /...	The Trinity, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1910.
22	yale.39002029677235	allow	pd	009777346	NULL	YALE	3629239	509607	NULL	NULL	15023098	The incarnation, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
23	hvd.ah4qgb	allow	pd	009777346	NULL	HVD	002565926	509607	NULL	NULL	15023098	The incarnation, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
24	uva.x030741716	allow	pd	009777346	v.6	UVA	u1171011	509607	NULL	NULL	15023098	The incarnation, by the Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
25	uva.x030735932	allow	pd	009777347	NULL	UVA	u1171025	3808443	NULL	NULL	25004656	Elements of divinity;	Morton & Griswold, 1847.
26	uva.x030735934	allow	pd	009777348	NULL	UVA	u1171029	13579294	NULL	NULL	02026079	Some features of faith; a popular discussion of ce...	The Young churchman co., 1902.
27	uva.x000531376	allow	pd	009777349	NULL	UVA	u1171034	614135288	NULL	NULL	08030158	Outlines of systematic theology.	American Baptist publication society, 1908
28	uva.x030735955	allow	pd	009777350	NULL	UVA	u1171036	3010360	NULL	NULL	39007848	Theological colloquies;	Lewis & Coleman, 1837.
29	uva.x030735956	allow	pd	009777351	NULL	UVA	u1171040	669874243	NULL	NULL	39007041	Outlines of Christian theology.	T. Whittaker, 1894.
30	uva.x030804130	allow	pd	009777352	NULL	UVA	u1171065	3760124	NULL	NULL	01029439	The theological compend: containing a system of...	Publishing house of the Methodist Episcop
31	uva.x030804135	allow	pd	009777353	NULL	UVA	u1171082	1823813	NULL	NULL	06009692	Christian doctrines: a compendium of theology.	American Baptist publication society 1878
32	uva.x030800859	allow	pd	009777354	NULL	UVA	u1171102	181822160	NULL	NULL	07028761	Old and new theology.	T. & T. Clark, 1885.
33	uva.x000736296	allow	pd	009777355	NULL	UVA	u1171144	703550644	NULL	NULL	32017455	The Bible, the church, and the reason;	C. Scribner's sons, 1893.
34	uva.x000430790	allow	pd	009777356	NULL	UVA	u1171171	7490799	NULL	NULL	13002018	The evolution of the kingdom,	C. C. Cook; [etc., etc., c1913]
35	uva.x030806027	allow	pd	009777357	NULL	UVA	u1171178	12852810	NULL	NULL	21000463	Why do things happen?	The Nunc licet press [c1920]
36	uva.x030803115	allow	pd	009777358	NULL	UVA	u1171257	216421600	NULL	NULL	46040063	Mohammed and Mohammedanism; lectures deliv...	Smith, Elder & co, 1926

	imprint
lack. With an introductory ...	F.H. Revell, 1898.
lack. With an introductory ...	F.H. Revell, 1898.
ince,	T. Y. Crowell & co. [c1901]
ngthen it,	Funk & Wagnalls company, 1916.
ome it, by Yontomo-Tashi, ...	Funk & Wagnalls company, 1916.
ome it, by Yontomo-Tashi, ...	Funk & Wagnalls company, 1916.
ld it,	Funk & Wagnalls company, 1916.
ld it,	Funk & Wagnalls company, 1916.
	S. Low, Manston, Searle, and Rivington, 1
	Scribner, Armstrong and company, 1876.
xm popular proverbs,	Scribner, Armstrong & co, 1875.
	1880.
	Cassell publishing company [1892]
ls and manners.	C. Knight and co., 1838.
ter / by J.R. Miller.	Thomas Y. Crowell, c1894.
ter / by J.R. Miller.	Thomas Y. Crowell, c1894.
ter / by J.R. Miller.	Thomas Y. Crowell, c1894.
s,	T. Y. Crowell & company [c1898]
hy door. A book for the still...	Fleming H. Revell company [c1895]
r. Francis J. Hall.	Longmans, Green, and Co., 1910.
r. Francis J. Hall.	Longmans, Green, and Co., 1910.
e Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
e Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
e Rev. Francis J. Hall.	Longmans, Green, and Co., 1915.
	Morton & Griswold, 1847.
; a popular discussion of ce...	The Young churchman co., 1902.
theology,	American Baptist publication society, 1906
;	Lewis & Coleman, 1837.
eology,	T. Whittaker, 1894.
end: containing a system of...	Publishing house of the Methodist Episcop
compendium of theology,	American Baptist publication society 1878
.	T. & T. Clark, 1885.
and the reason;	C. Scribner's sons, 1893.
ingdom,	C. C. Cook; [etc., etc., c1913]
y?	The Nunc licet press [c1920]
meditation: lectures deliv...	Smith, Elder & co, 1976

select * from Dictionary_Institution					
100 %					
Results Messages					
	Source_Code	Source_Name	Source_Country	Source_State	Source_City
1	AEU	University of Alberta	Canada	Alberta	Edmonton
2	AZTES	Arizona State University	United States	Arizona	Tempe
3	AuBrU	University of Queensland	Australia	Queensland	Brisbane
4	CHI	University of Chicago	United States	Illinois	Chicago
5	CMALG	Getty Research Institute	United States	California	Los Angeles
6	COO	Cornell University	United States	New York	Ithaca
7	CtU	University of Connecticut	United States	Connecticut	Storrs
8	DUL	Duke University	United States	North Carolina	Durham
9	DeU	University of Delaware	United States	Delaware	Newark
10	FMU	University of Miami	United States	Florida	Coral Gables
11	FU	University of Florida	United States	Florida	Gainesville
12	GEU	Emory University	United States	Georgia	Atlanta
13	HVD	Harvard University	United States	Massachusetts	Cambridge
14	IEN	Northwestern University	United States	Illinois	Evanston
15	IND	University of Notre Dame	United States	Indiana	Notre Dame
16	INU	Indiana University	United States	Indiana	Bloomington
17	IaU	University of Iowa	United States	Iowa	Iowa City
18	JTKU	Keio University	Japan	Tokyo	Tokyo
19	LOC	Library of Congress	United States	District of Columbia	Washington, D.C.
20	MChB	Boston College	United States	Massachusetts	Chestnut Hill
21	MDBJ	Johns Hopkins University	United States	Maryland	Baltimore
22	MDL	Minnesota Digital Library	United States	Minnesota	Minneapolis
23	MMeT	Tufts University	United States	Massachusetts	Medford
24	MU	University of Massachusetts	United States	Massachusetts	Amherst
25	MWICA	Sterling and Francine Clark Art Institute	United States	Massachusetts	Williamstown
26	MdU	University of Maryland	United States	Maryland	College Park
27	MEM	Michigan State University	United States	Michigan	East Lansing
28	MiU	University of Michigan	United States	Michigan	Ann Arbor
29	MoU	University of Missouri	United States	Missouri	Columbia
30	NCSU	North Carolina State University	United States	North Carolina	Raleigh
31	NCWSW	Wake Forest University	United States	North Carolina	Winston-Salem
32	NJP	Princeton University	United States	New Jersey	Princeton
33	NNC	Columbia University	United States	New York	New York City
34	NYP	New York Public Library	United States	New York	New York City
35	OU	Ohio State University	United States	Ohio	Columbus
36	PHC	Haverford College	United States	Pennsylvania	Haverford
37	PST	Pennsylvania State University	United States	Pennsylvania	University Park
38	PU	University of Pennsylvania	United States	Pennsylvania	Philadelphia
Query executed successfully.					

# Data Visualization 1

## Number of Deposited Volumes to HathiTrust Digital Library by Original Source of Content

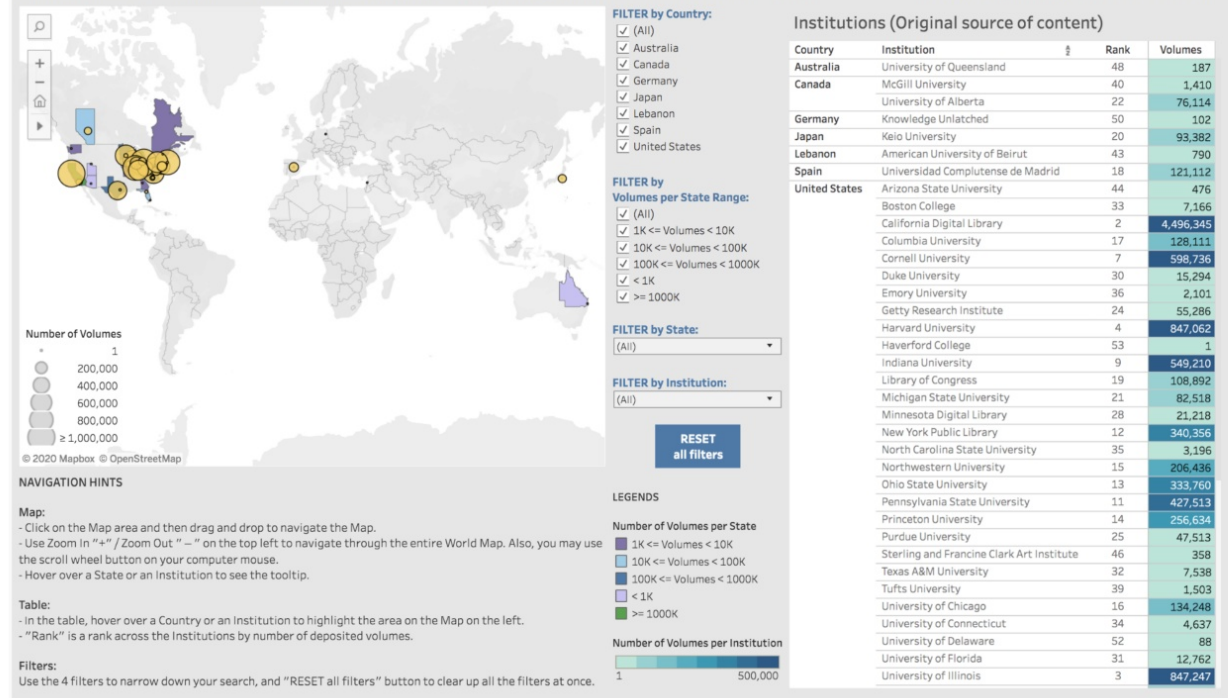
Data as of January 1, 2020

Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606 from 53 institutions representing 7 countries.

by Vera Kuipers

Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign





## Number of Deposited Volumes to HathiTrust Digital Library by Original Source of Content

Data as of January 1, 2020

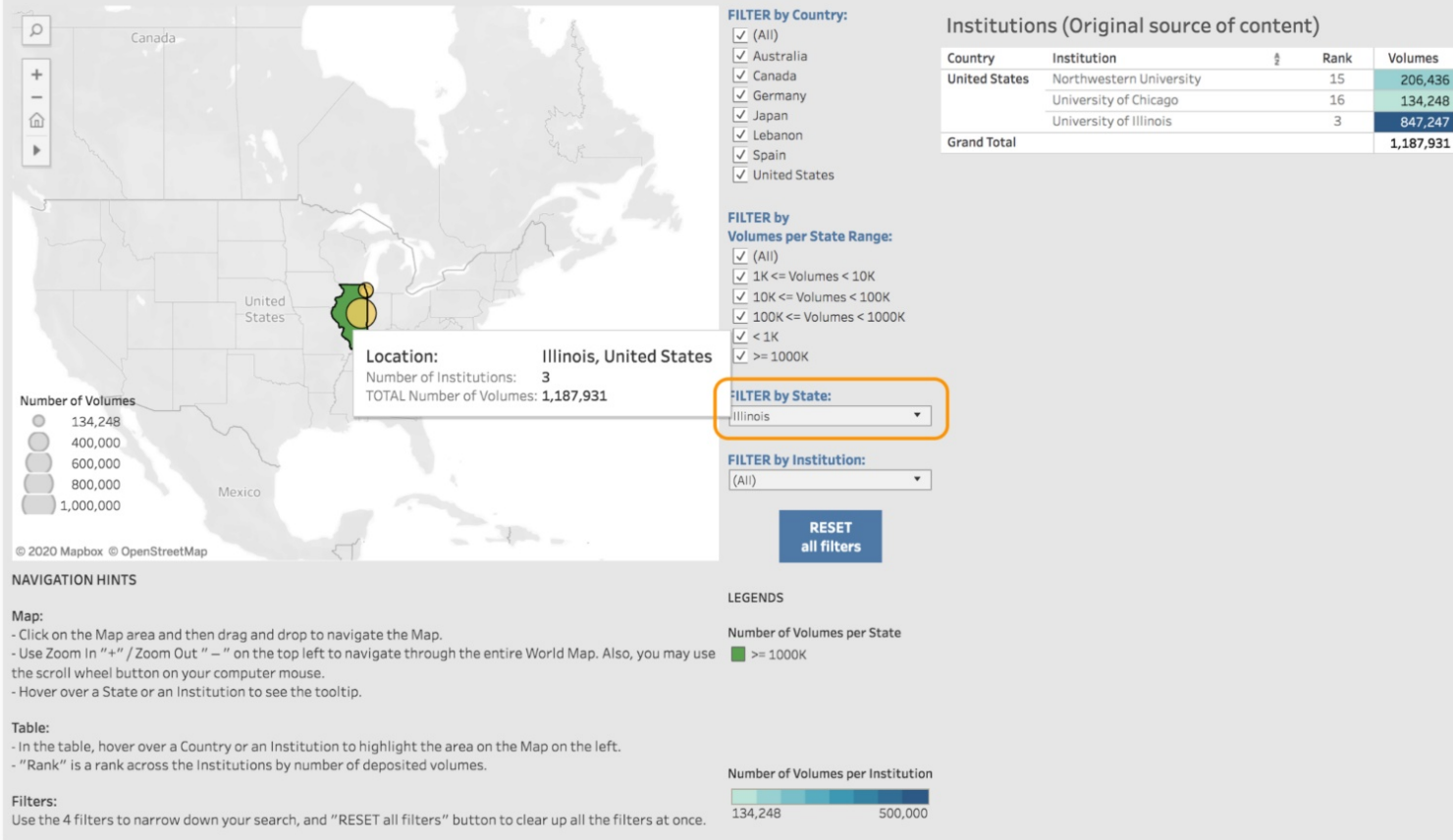
Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606 from 53 institutions representing 7 countries.

by Vera Kuipers

Office of Collections and Technical Services

University Library, University of Illinois at Urbana-Champaign





## Number of Deposited Volumes to HathiTrust Digital Library by Original Source of Content

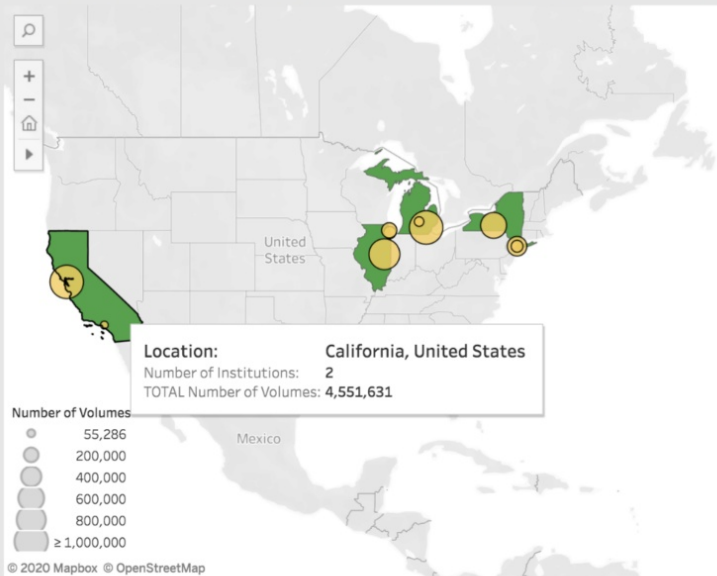
Data as of January 1, 2020

Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606 from 53 institutions representing 7 countries.

by Vera Kuipers

Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign



### FILTER by Country:

- ☒ (All)
- ☒ Australia
- ☒ Canada
- ☒ Germany
- ☒ Japan
- ☒ Lebanon
- ☒ Spain
- ☒ United States

### FILTER by

#### Volumes per State Range:

- ☐ (All)
- ☐ 1K <= Volumes < 10K
- ☐ 10K <= Volumes < 100K
- ☐ 100K <= Volumes < 1000K
- ☐ < 1K
- ☒ >= 1000K

### FILTER by State:

(All)

### FILTER by Institution:

(All)

RESET  
all filters

### Institutions (Original source of content)

Country	Institution	#	Rank	Volumes
United States	California Digital Library		2	4,496,345
	Columbia University		17	128,111
	Cornell University		7	598,736
	Getty Research Institute		24	55,286
	Michigan State University		21	82,518
	New York Public Library		12	340,356
	Northwestern University		15	206,436
	University of Chicago		16	134,248
	University of Illinois		3	847,247
	University of Michigan		1	4,910,335
Grand Total				11,799,618

### NAVIGATION HINTS

#### Map:

- Click on the Map area and then drag and drop to navigate the Map.
- Use Zoom In "+" / Zoom Out "-" on the top left to navigate through the entire World Map. Also, you may use the scroll wheel button on your computer mouse.
- Hover over a State or an Institution to see the tooltip.

#### Table:

- In the table, hover over a Country or an Institution to highlight the area on the Map on the left.
- "Rank" is a rank across the Institutions by number of deposited volumes.

#### Filters:

Use the 4 filters to narrow down your search, and "RESET all filters" button to clear up all the filters at once.

### LEGENDS

#### Number of Volumes per State

>= 1000K

#### Number of Volumes per Institution

55,286 500,000

## Number of Deposited Volumes to HathiTrust Digital Library by Original Source of Content

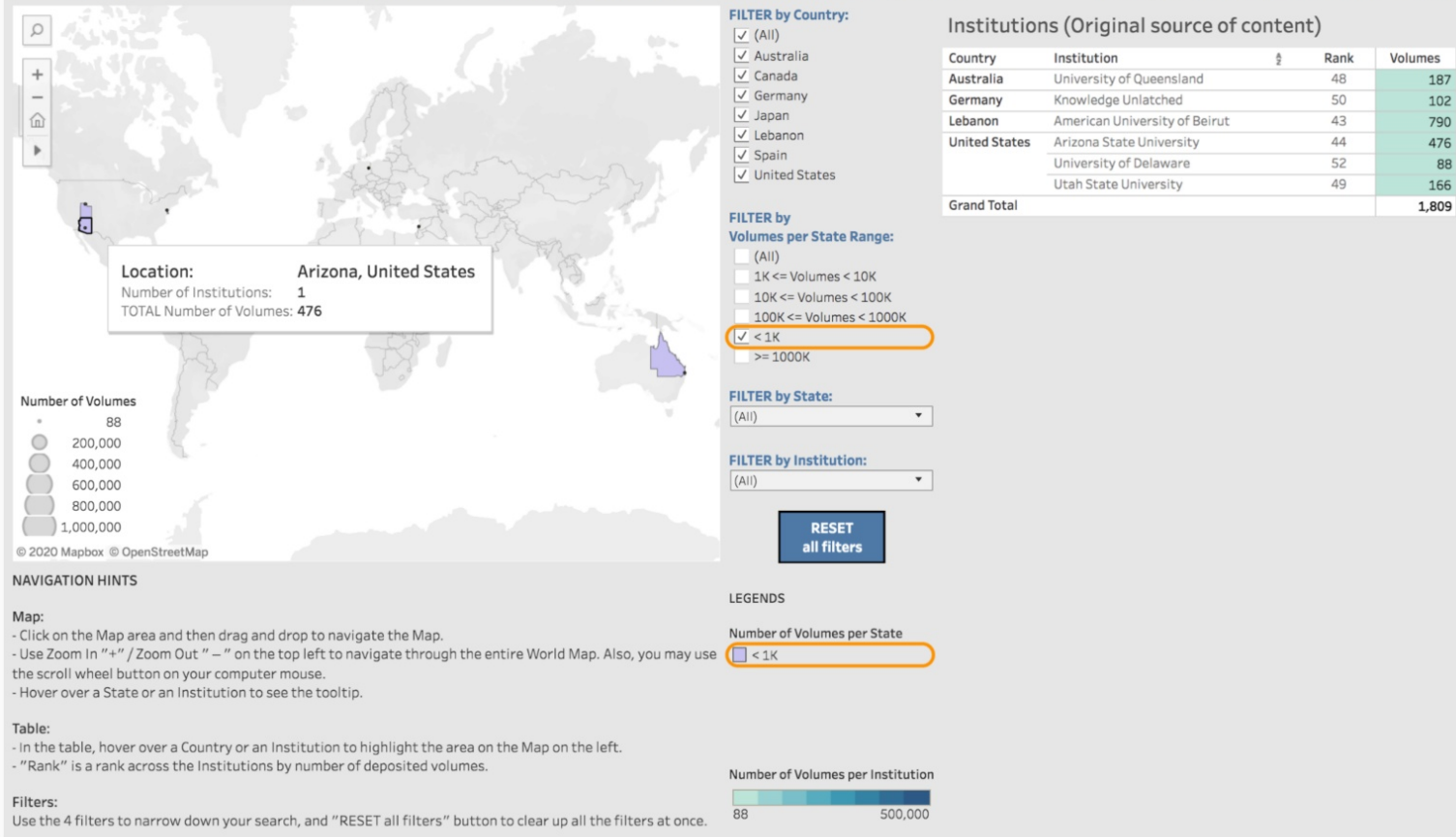
Data as of January 1, 2020

Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606 from 53 institutions representing 7 countries.

by Vera Kuipers

Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign



## Data Visualization 2

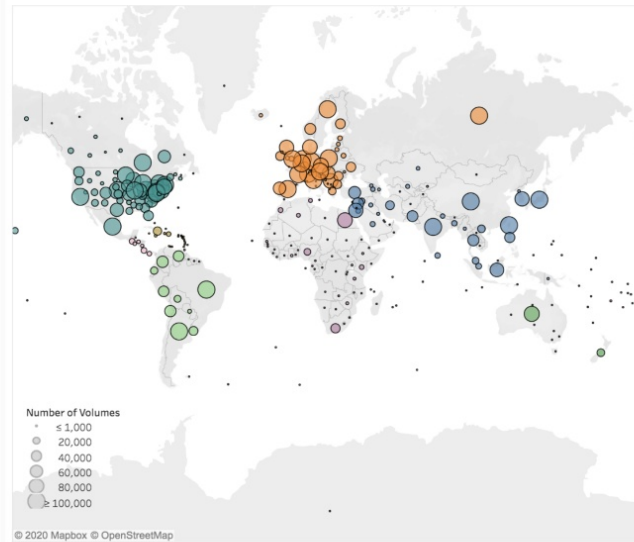
### Publication Place and Language of Volumes Deposited to HathiTrust Digital Library by Institution

Data as of January 1, 2020

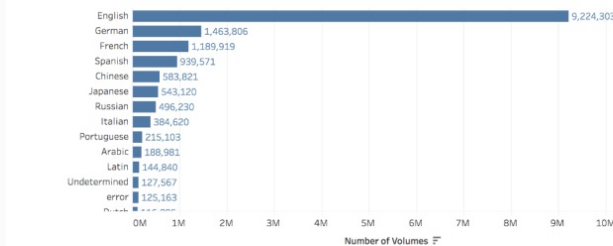
Data source: <https://www.hathitrust.org/hathitrust>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606.

Place of Publication:



Language of Publication:



FILTER by Institution (Original source of content)

(All)

TOTAL STATS for 53 Institution(s):  
17,153,606 deposited volumes published in 329 various places  
and in 432 languages

Click on Publication REGION  
to highlight it and apply filter:

Region	Count
Africa	202,020
Asia	2,114,771
Atlantic Ocean	969
Australasia	103,485
Central America	49,093
Europe	6,281,802
Indian Ocean	2,284
North America	6,313,728
Other	1,257,497
Pacific Ocean	9,715
South America	532,317
West Indies	49,037
Null	184,301
error	52,827
Grand Total	17,153,606

LEGEND

Region	Color
Africa	Blue
Asia	Orange
Atlantic Ocean	Green
Australasia	Yellow
Central America	Purple
Europe	Red
Indian Ocean	Dark Blue
North America	Light Blue
Pacific Ocean	Pink
South America	Light Green
West Indies	Light Yellow
Other	Grey
Null	White
error	Dark Grey

NAVIGATION HINTS

Data:  
"Null" means that no publication place/language was indicated in the dataset for the volume.  
"error" means that publication place/language code is incorrect.

Filters:  
- To RESET Filter/Highlight, click on the selected Region or Country/State again.  
- In Country/State table, use a scroll bar to see all the places of publication and the total number of volumes.

Bar chart:  
- Use a scroll bar to see all the languages of publication.

by Vera Kuipers  
Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign

Click on COUNTRY/STATE  
to highlight it and apply filter:

Country/State	Count
New York	1,544,207
Germany	1,309,341
England	1,301,020
No place, unknown, or un...	1,247,956
District of Columbia	1,070,956
France	987,141
Japan	544,616
Italy	422,252
Russia	405,089
India	396,832
Massachusetts	395,376
Illinois	378,249
California	367,822
China	364,327
Spain	281,029
Pennsylvania	260,573
Netherlands	203,619
Michigan	193,190
Taiwan	173,143
Ontario	170,750
Switzerland	160,602
Brazil	158,231
United States	155,071
Mexico	152,573
Maryland	142,824
Argentina	131,151
Poland	126,751
Ohio	119,266
Austria	119,144
Belgium	111,455
Sweden	108,838
New Jersey	100,728
Minnesota	94,595
Connecticut	82,286
Australia	78,923
Czech Republic	78,344
Wisconsin	77,433
Virginia	77,414
Egypt	76,074
Serbia	75,803
Denmark	73,868
Scotland	71,543
Israel	71,040
Korea (South)	70,560

## Publication Place and Language of Volumes Deposited to HathiTrust Digital Library by Institution

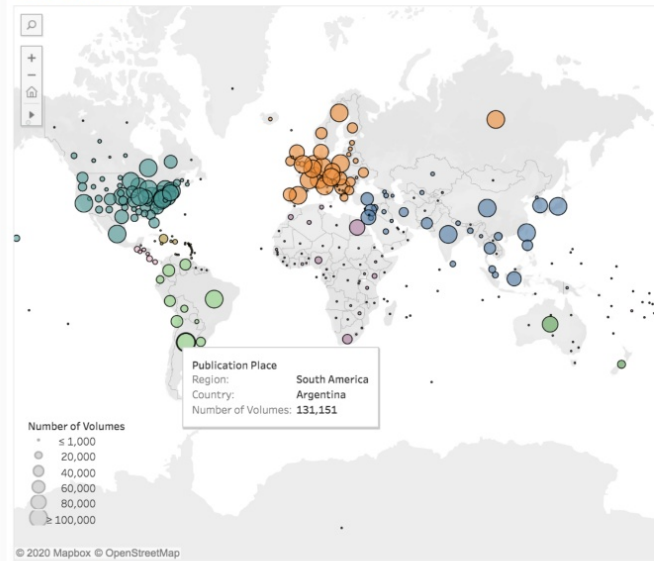
Data as of January 1, 2020

Data source: <https://www.hathitrust.org/hathifiles>

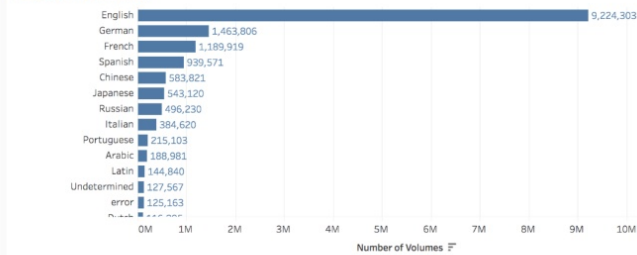
As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606.

by Vera Kuipers  
Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign

### Place of Publication:



### Language of Publication:



### FILTER by Institution (Original source of content)

(All)

TOTAL STATS for 53 Institution(s):  
17,153,606 deposited volumes published in 329 various places  
and in 432 languages

Click on Publication REGION  
to highlight it and apply filter:

Region	
Africa	202,020
Asia	2,114,771
Atlantic Ocean	969
Australasia	103,485
Central America	49,093
Europe	6,281,802
Indian Ocean	2,284
North America	6,313,728
Other	1,257,257
Pacific Ocean	9,715
South America	532,317
West Indies	49,037
Null	194,301
error	52,827
Grand Total	17,153,606

### LEGEND

Region	
Africa	North America
Asia	Pacific Ocean
Atlantic Ocean	South America
Australasia	West Indies
Central Ame...	Other
Europe	Null
Indian Ocean	error

### NAVIGATION HINTS

Data:  
"Null" means that no publication place/language was indicated in the dataset for the volume.  
"error" means that publication place/language code is incorrect.

### Filters:

- To RESET Filter/Highlight, click on the selected Region or Country/State again.  
- In Country/State table, use a scroll bar to see all the places of publication and the total number of volumes.

### Bar chart:

- Use a scroll bar to see all the languages of publication.

Country/State	
New York	1,544,207
Germany	1,309,341
England	1,301,020
No place, unknown, or un...	1,247,956
District of Columbia	1,070,956
France	987,141
Japan	944,616
Italy	422,252
Russia	405,089
India	396,832
Massachusetts	395,376
Illinois	378,249
California	367,822
China	364,327
Spain	281,029
Pennsylvania	260,573
Netherlands	203,619
Michigan	193,190
Taiwan	173,143
Ontario	170,750
Switzerland	160,602
Brazil	158,231
United States	155,071
Mexico	152,573
Maryland	142,824
Argentina	131,151
Poland	126,751
Ohio	119,266
Austria	119,144
Belgium	111,455
Sweden	108,838
New Jersey	100,728
Minnesota	94,595
Connecticut	82,286
Australia	78,923
Czech Republic	78,344
Wisconsin	77,433
Virginia	77,414
Egypt	76,074
Serbia	75,803
Denmark	73,868
Scotland	71,543
Israel	71,040
Korea (South)	70,560

## Publication Place and Language of Volumes Deposited to HathiTrust Digital Library by Institution

Data as of January 1, 2020

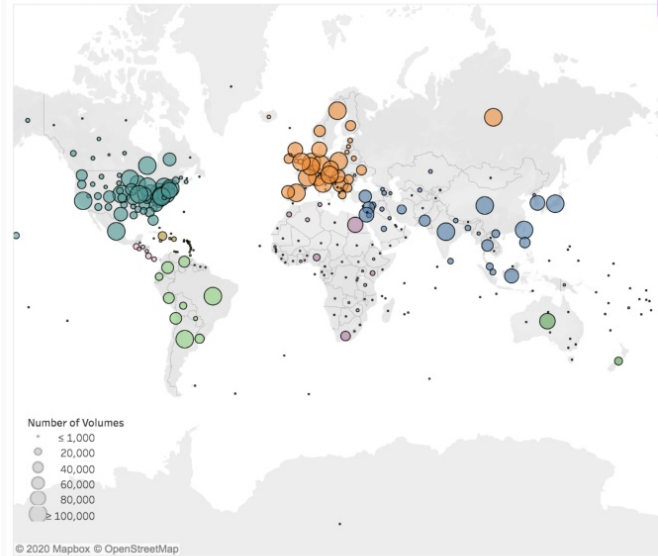
Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606.

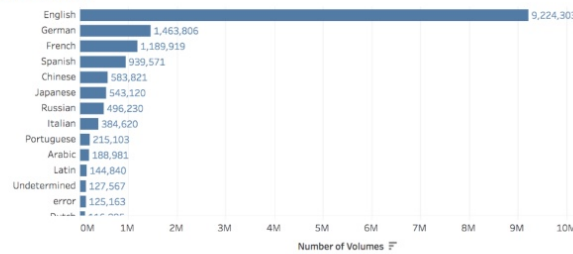
by Vera Kuipers

Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign

Place of Publication:



Language of Publication:



FILTER by Institution (Original source of content)

(All)

- ☒ (All)
- ☒ American University of Beirut
- ☒ Arizona State University
- ☒ Boston College
- ☒ California Digital Library
- ☒ Columbia University
- ☒ Cornell University
- ☒ Duke University
- ☒ Emory University
- ☒ Getty Research Institute
- ☒ Harvard University
- ☒ Haverford College
- ☒ Indiana University
- ☒ Keio University
- ☒ Knowledge Unlatched
- ☒ Library of Congress
- ☒ McGill University
- ☒ Michigan State University
- ☒ Minnesota Digital Library
- ☒ New York Public Library
- ☒ North Carolina State University
- ☒ Northwestern University
- ☒ Ohio State University
- ☒ Pennsylvania State University
- ☒ Princeton University
- ☒ Purdue University
- ☒ Sterling and Francine Clark Art Institute
- ☒ Texas A&M University
- ☒ Tufts University

NAVIGATION HINTS

Data:  
"Null" means that no publication place/language was indicated in the dataset for the volume.  
"error" means that publication place/language code is incorrect.

Filters:

- To RESET Filter/Highlight, click on the selected Region or Country/State again.  
- In Country/State table, use a scroll bar to see all the places of publication and the total number of volumes.

Bar chart:

- Use a scroll bar to see all the languages of publication.

Ohio	119,266
Austria	119,144
Belgium	111,455
Sweden	108,838
New Jersey	100,728
Minnesota	94,595
Connecticut	82,286
Australia	78,923
Czech Republic	78,344
Wisconsin	77,433
Virginia	77,414
Egypt	76,074
Serbia	75,803
Denmark	73,868
Scotland	71,543
Israel	71,040
Korea (South)	70,560

## Publication Place and Language of Volumes Deposited to HathiTrust Digital Library by Institution

Data as of January 1, 2020

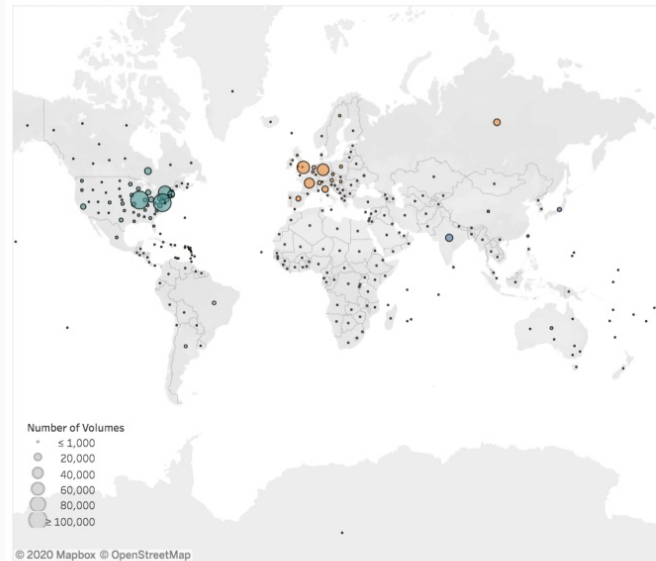
Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606.

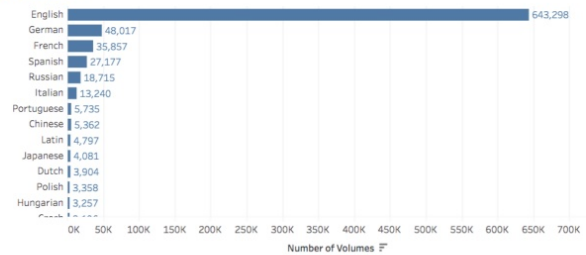
by Vera Kuipers

Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign

### Place of Publication:



### Language of Publication:



### FILTER by Institution (Original source of content)

University of Illinois

TOTAL STATS for 1 Institution(s):  
847,247 deposited volumes published in 284 various places  
and in 187 languages

Click on Publication REGION  
to highlight it and apply filter:

Region	
Africa	6,905
Asia	37,649
Atlantic Ocean	35
Australasia	5,048
Central America	1,590
Europe	219,955
Indian Ocean	181
North America	531,859
Other	22,554
Pacific Ocean	71
South America	16,175
West Indies	1,701
Null	2,506
error	1,019
Grand Total	847,247

### LEGEND

Region	
Africa	North America
Asia	Pacific Ocean
Atlantic Ocean	South America
Australasia	West Indies
Central America	Other
Europe	Null
Indian Ocean	error

### NAVIGATION HINTS

Data:  
"Null" means that no publication place/language was indicated in the dataset for the volume.  
"error" means that publication place/language code is incorrect.

Filters:  
- To RESET Filter/Highlight, click on the selected Region or Country/State again.  
- In Country/State table, use a scroll bar to see all the places of publication and the total number of volumes.

Bar chart:  
- Use a scroll bar to see all the languages of publication.

Country/State	
New York	57,671
Germany	44,287
England	50,647
No place, unknown, or un...	22,341
District of Columbia	123,026
France	32,692
Japan	5,790
Italy	14,525
Russia	14,913
India	17,536
Massachusetts	16,007
Illinois	130,305
California	9,947
China	3,066
Spain	9,076
Pennsylvania	13,430
Netherlands	5,182
Michigan	9,119
Taiwan	3,013
Ontario	15,850
Switzerland	5,580
Brazil	4,841
United States	2,027
Mexico	3,469
Maryland	8,402
Argentina	3,936
Poland	4,082
Ohio	9,910
Austria	4,557
Belgium	3,031
Sweden	3,057
New Jersey	5,952
Minnesota	4,990
Connecticut	4,402
Australia	3,632
Czech Republic	3,817
Wisconsin	5,152
Virginia	7,761
Egypt	730
Serbia	2,011
Denmark	2,078
Scotland	2,321
Israel	429
Korea (South)	1,176



## Publication Place and Language of Volumes Deposited to HathiTrust Digital Library by Institution

Data as of January 1, 2020

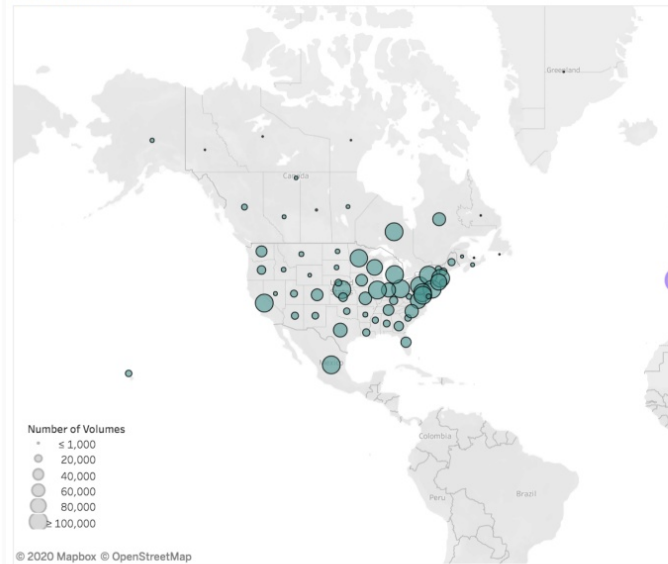
Data source: <https://www.hathitrust.org/hathifiles>

As of January 1, 2020, the total number of deposited volumes to HathiTrust Digital Library is 17,153,606.

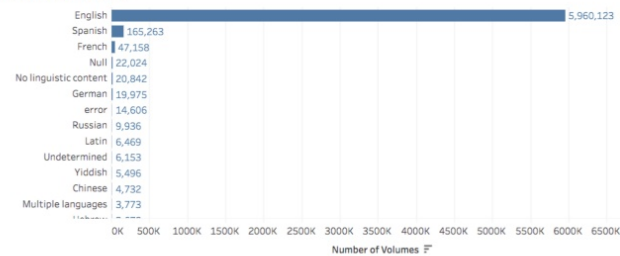
by Vera Kuipers

Office of Collections and Technical Services  
University Library, University of Illinois at Urbana-Champaign

### Place of Publication:



### Language of Publication:



### FILTER by Institution (Original source of content)

(All)

TOTAL STATS for **53** Institution(s):  
**17,153,606** deposited volumes published in **329** various places  
 and in **432** languages

Click on Publication REGION  
to highlight it and apply filter:

Region	
Africa	202,020
Asia	2,114,771
Atlantic Ocean	969
Australasia	103,485
Central America	49,093
Europe	4,918,327
Indian Ocean	2,999
North America	6,313,728
Other	1,257,457
Pacific Ocean	9,715
South America	532,317
West Indies	49,037
Null	184,301
error	52,827
Grand Total	17,153,606

### LEGEND

Region  
 North America

Click on COUNTRY/STATE  
to highlight it and apply filter:

Country/State	
New York	1,544,207
District of Columbia	1,070,956
Massachusetts	395,376
Illinois	378,249
California	367,822
Pennsylvania	260,573
Michigan	193,190
Ontario	170,750
United States	155,071
Mexico	152,573
Maryland	142,824
Ohio	119,266
New Jersey	100,728
Minnesota	94,595
Connecticut	82,286
Wisconsin	77,433
Virginia	77,414
Indiana	63,087
Texas	60,634
North Carolina	58,159
Quebec	53,904
Missouri	51,400
Iowa	44,798
Colorado	44,714
Washington	39,364
Tennessee	38,045
Georgia	29,362
Florida	34,337
Kansas	24,846
Oregon	24,424
Kentucky	22,417
New Hampshire	18,917
Louisiana	17,955
Rhode Island	17,397
Arizona	17,160
Maine	17,147
Utah	16,332
Alabama	15,956
Nebraska	15,612
New Mexico	15,473
South Carolina	14,595
Oklahoma	14,576
Hawaii	14,273
Vermont	14,018

### NAVIGATION HINTS

Data:  
 "Null" means that no publication place/language was indicated in the dataset for the volume.  
 "error" means that publication place/language code is incorrect.

Filters:  
 - To RESET Filter/Highlight, click on the selected Region or Country/State again.  
 - In Country/State table, use a scroll bar to see all the places of publication and the total number of volumes.

Bar chart:  
 - Use a scroll bar to see all the languages of publication.

# Thank You!