

Good Data = Trustworthy Findings: Metrics for Validating Quality in Large-Scale Digitization Paul Conway, Associate Professor

ABSTRACT

This project is exploring quality issues in the largescale digitization of books by third-party vendors, such as Google and the Internet Archive. Metrics developed in the project test a model of digitization error in random samples of digital book-surrogates representing the full range of source volumes. Systematic sampling of page-images within each volume in the samples produced a study set of over 1 million page images. Using a custombuilt web-enabled database system, trained coders recorded perceived error in these page-images on a severity scale of 0-5 for up to eleven possible errors and five additional errors in whole volumes. Physical inspection of source volumes validate errors detected from scanning and post scan manipulation.



FROM PROJECT GUTENBERG TO GOOGLE BOOKS, the large-scale digitization of books and serials is generating extraordinary collections of intellectual content that are transforming the way society reads and learns.

LITERATURE ON INFORMATION QUALITY is almost silent on how to measure quality attributes of very large collections of digitized books and journals.

WE DEFINE QUALITY as the absence of errors in scanning and post-scan processing relative to expected uses (Conway 2011).

HATHITRUST DIGITAL LIBRARY is the testbed for the project.

Conway, P. (2011). "Archival Quality and Long-term Preservation: A Research Framework for Validating the Usefulness of Digital Surrogates." Archival Science 11 (3): 293-309.

ACKNOWLEDGEMENTS

Partner: University of Minnesota **Partner: HathiTrust Digital Library** Ed Rothman, Professor of Statistics Jackie Bronicki, Project Coordinator Ken Guire, Statistician Ryan Rotter, Systems Programmer Jeremy York, Associate Librarian

School of Information

CONTEXT

REFERENCE

ERROR MODEL

Level of Abstraction

- LEVEL 1: DATA/INFORMATION
- 1.1 Text: thick text [fill, excessive]
- 1.2 Text: broken text [character breakup] 1.3 Illustration: scanner effects [moiré, gridding]
- 1.4 Illustration: tone, brightness, contrast
- 1.5 Illustration: color imbalance, gradient shifts
- LEVEL 2: ENTIRE PAGE
- 2.1 Blur [distortion] 2.2 Warp [text alignment]
- 2.3 Skew [page alignment]
- 2.4 Crop [gutter, text block]
- 2.5 Obscured/cleaned [portions not visible]
- 2.6 Colorization [text bleed, low contrast]
- LEVEL 3: WHOLE VOLUME
- 3.1 Fully obscured [foldouts or objects]
- 3.2 Missing pages [one or more]
- 3.3 Duplicate pages [one or more] 3.4 Order of pages
- 3.5 False pages [not part of Original Content]

PAGE IMAGE DATA



RELIABILITY (Kappa) (0)(1)(2)(3)(4)

	Simple
Gold x Coder 1	0.46
Gold x Coder 2	0.45
Gold x Coder 3	0.48
Gold x Coder 4	0.47
Gold x Coder 5	0.48
Gold x Coder 6	0.46
Gold x Coder 7	0.42

Project Website: http://hathitrust-quality.projects.si.umich.edu/

	Possible Cause of Error]			
	1 ossible Gause of Error		fini Nashata kere		
					MOBY
	Source or post-processing		Many Data Trans 1		-32
	Source or post-processing				N N
	Scanning or post-processing				
3	Scanning, post-processing or source			A REPORT OF	Contraction of the
	Scanning, post-processing or source		FR	BELL-	I BEE
	Scanning or <mark>source</mark> Post-processing		INTERNITY TON INTERNITY TON THE REAL PROPERTY OF THE ALL PROPERTY INTERNITY OF THE ALL PROPERTY OF THE ALL PROPERTY OF THE ALL PROPERTY INTERNITY OF THE ALL PROPERTY OF THE ALL PROPERTY OF THE ALL PROPERTY OF THE ALL PROPERTY OF THE ALL PROPERTY INTERNITY OF THE ALL PROPERTY	10 Instability most first barrier of an an example for any set of any set	Annual Control of the second s
	Scanning, source or post-processing Source or post-processing		The states is the state is trained as the process that it is the state is the stat	The second se	
	Scanning or post-processing		patter and Mich Operation Angles of Inter-F Re- formances, y this Visite of Inter-Face and the Sec- formation of the Visite of Inter-Face and the Sec- formation and the Sector of the Sector Sector and the Sector Sector of the Sector Sector and the Inter-Sector of the Sector Sector and the Inter-Sector of the Sector Sector and the Inter-Sector Sector Sector Sector Sector Sector Inter-Sector Sector Sector Sector Sector Sector Inter-Sector Se	The det and the fit (being the det and	And A start of the
	Source or post-processing			ter state for a	
	Scanning		headening in many dispersion of the set of a set of "Mathematical and a set of a set	. International States	States and
	Source or scanning		The spin strength for a particul all, if and assess both and frame to a lower obtaining and disconting spin strength of the shift of the lower of the spin billion on parameters of the spin of the shift of the spin strength strength strength of the spin of the spin strength of the spin spin strength of the spin strength of	Anna and an and a start of the	
	Source or scanning		This paper would have to have approach to be book. We remain such that experiments to be a topological start of the experiments which is the topological start of the experiments which is the star- tion of the start of the start of the start of the balance of the start of the paper balance of the start is the start of the start of the paper balance of the start is the start of the start of the start is the start of the start is the start of the start of the start of the start is the start of the start of the start of the start is the start of the	And the second s	
	Source or scanning		and start for bar, or meaning to shall all the formula the start of th		A sense of a start of a to be a sense of a start of a start of a sense of a start of a start before a start of a start of a start of a start before a start of a start of a start of a start of a start before a start of a start of a start of a start of a start before a start of a start before a start of a
	Scanning or post-processing		eff. The is a symmetry and transition to starting equation is the transition of the	An effect delivation for analy and well schedule across Rest delivation of the sector and the sector and the Part foreign part were, seen part part and the geld beam interest of entering cards of the spectra sector for the "Mart is the sector delivation of the sector across the foreign cards and the sector across the sector beam is the sector delivation of the sector across the sector of the sector across the sector of the sector have been sector as a sector of the sector is the control more across the sector of the sector is the control more across the sector of the sector is the	the second secon

WHOLE VOLUME DATA



PHYSICAL BOOK DATA

 C A C M B https://rrotter-full.babel.hathitrust.org/qual/physical.html?htid Textblock Missing pages Loose pages Parts of pages missing Text block broken Graphic No help text available. Substantial graphic content Britle Does not break Breaks at 4 double folds Breaks at 2 double folds Breaks at 4 double folds Breaks at 2 double folds Breaks at 3 double folds Breaks at 4 double fo	HathiTi	rust Digital Library - 🗆 🕀
Textblock Missing pages Loose pages Textblock broken Textblock broken Textblock broken Graphic No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Tom, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stary annotation heaved annotated over and over: count it in the results Pencil Ink Highlighter Other	← ⇒ 0	https://rrotter-full.babel.hathitrust.org/qual/physical.html?htid
 Missing pages Loose pages Parts of pages missing Text block broken Graphic No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Tom, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotation noccuus markings. Ignore marks made by librarians in cataloging the volume. One section heavian notated over and over = count it in the results Pencil Ink Highlighter Other 	Textblock	
 Loose pages Parts of pages missing Text block broken Graphic No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, to opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but you able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage in or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heavi annotated over and over= count it in the results Pencil Ink Highlighter Other 	Missing	pages
 Parts of pages missing Text block broken Graphic No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Tom, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotation indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotation indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotation, indicate yes or and over = count it in the results Pencil Ink Highlighter Other 	Loose p	ages
Graphic Graphic No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	Tarts of	pages missing
Graphic No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat Food or drink Annotation Indicate over and over= count it in the results Pencil Inc. Disclose of the inservence of the results Does of the inservence of the results Does of the result	Text Dio	ck broken
No help text available. Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heavi annotated over and over= count it in the results Pencil Ink Highlighter	Graphic	
 Substantial graphic content Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotation indicate over and over= count it in the results Pencil Ink Highlighter Other 	No help te	xt available.
Brittle Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Cutter	Substar	itial graphic content
Does the corner of the page break off after two double folds - i.e., after being folded in one dirction, opposite direction, twice? Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage is or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heavi annotated over and over= count it in the results Pencil Ink Highlighter Other 	Brittle	
 Does not break Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	Does the	corner of the page break off after two double folds - i.e., after being folded in one dirction, firection, twice?
 Breaks at 4 double folds Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise after the printed text or illustration on multiple pages. Ignore stray annotation nearkings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	Opposite t	niecion, wice:
 Breaks at 2 double folds Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotati innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	Breaks	at 4 double folds
Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage i or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter	Breaks	at 2 double folds
Gutter A centimeter is approximately the width of the tip of your pinky finger. Rulers will be supplied, but yo able to judge the gutter after measuring some tight gutters. Image Indicate yes or no to each of the listed types of damage, based on your judgment that the damage or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Image Image Image Indicate yes or no to each of the listed types of damage, based on your judgment that the damage or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Image Image<	555 - 555 - 555 - 555 871 - 65	
 Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage is or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dat significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotation nor otherwise alter the printed text or illustration on multiple pages. Ignore stray annotation hand over and over = count it in the results Pencil Ink Highlighter Other 	Gutter A centime	ter is annrovimately the width of the tin of your ninky finder. Bulers will be sunnlied, but yo
 Tight gutter (1 cm or less) Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage is or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional data significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotation innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	able to jug	Ige the gutter after measuring some tight gutters.
Damage Indicate yes or no to each of the listed types of damage, based on your judgment that the damage or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional da significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Vater damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	🔲 Tight gu	tter (1 cm or less)
Indicate yes or no to each of the listed types of damage, based on your judgment that the damage or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dar significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	Damage	
or systematic. Ignore occasional dog ears, food stains, water damage unless such occasional dat significantly detracts from the usability of the volume. Torn, ripped Animal or insect damage Vater damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	Indicate ye	es or no to each of the listed types of damage, based on your judgment that the damage
 Torn, ripped Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	or system	atic. Ignore occasional dog ears, food stains, water damage unless such occasional dai tv dotracte from the usability of the volume.
 Animal or insect damage Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heavi annotated over and over= count it in the results Pencil Ink Highlighter Other 		ined
 Water damage Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	Animal	or insect damage
 Food or drink Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other 	Water d	amage
Annotation Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	E Food or	drink
Indicate yes or no to each of the possible types of annotation, based on your judgment that such ar may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	Appotation	
may obsure or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotat innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	Indicate ve	es or no to each of the possible types of annotation, based on your judoment that such ar
innocuous markings. Ignore marks made by librarians in cataloging the volume. One section heav annotated over and over= count it in the results Pencil Ink Highlighter Other	may obsu	re or otherwise alter the printed text or illustration on multiple pages. Ignore stray annotal
Pencil Ink Highlighter Other	innocuous	s markings. Ignore marks made by librarians in cataloging the volume. One section heav
Ink Highlighter Other		
Highlighter Other		
Cther	🔲 Highligh	nter
	Other	

REPRESENTATIVENESS

(4)(5)					
leighted	Sample Page	Criteria	Population	Volume	Pages
0.6	1	Google pre-1923 English	1.3 mil	1000	93,858
0.6	2	Google post-1922 English	6.5 mil	1000	86,439
0.0	3	Internet Archive pre-1923 English	850,000	1000	84,539
0.63	4	Non-Roman Scripts	1.29 mil	1000	91,381
0.62	Volume				
0.02	1a	Google pre-1923 English	1.3 mil	1000	397,467
0.62	2a	Google post-1922 English	6.5 mil	1000	294,505
06	Book				
0.0	1b	Google pre-1923 English	1000	906	N/A
0.55	1c	Google post-1922 English Mich	1000	584	N/A

