

Datenbank: Cambridge Histories

Provider: Cambridge University Press

		Cambridge Histories
Access	Web address, API, Dumps, offline back up copy	<ul style="list-style-type: none"> Download from website: https://www.cambridge.org/core/what-we-publish/collections/cambridge-histories Please contact openresearch@cambridge.org before using automated methods to download content No API
Documentation	Web address	<ul style="list-style-type: none"> https://www.cambridge.org/core/what-we-publish/collections/cambridge-histories/about https://www.cambridge.org/core/services/open-research/text-and-data-mining
Distribution		<ul style="list-style-type: none"> Global Monthly updates c. 12 volumes added p.a.
Scope	Content Purpose Field of use	<ul style="list-style-type: none"> 418 books (as at 05/2021 with c. 12 added p.a) c. 200,000 pages Historical reference and research in 10 main subject areas
Time, Place, Language	temporal, local reference	<ul style="list-style-type: none"> Global coverage All historical periods English
Data type	What are the basic data types?	<ul style="list-style-type: none"> XML/HTML on request: openresearch@cambridge.org PDF
Provenance, dependencies, accompanying material	original data source, manufacturer, data collection procedure, dependencies / links to other data sets / online resources, old versions	<ul style="list-style-type: none"> Printed books published from 2008- Includes all titles contained in the Cambridge Histories Series published by Cambridge University Press.
Description Structured text data	Text markup or data structure e.g. TXT, XML, ALTO, TEI, versions	<ul style="list-style-type: none"> Titles published from 2014- are BITS XML and PDF with standard metadata at book and chapter level. Titles pre-2014 generally PDF-only with standard metadata.
Description of databases, tabular data	data tables, existing / recommended data splits (e.g. training / test set)	<ul style="list-style-type: none"> Tables and charts generally captured as images
Description of	as precisely as	<ul style="list-style-type: none"> Various

image formats	possible (e.g. resolution, greyscale / bitonal)	
Standards, vocabularies	as precisely as possible: standards and vocabularies used	n/a
Data quality: OCR; missing, incorrect, redundant data, noise	For example. OCR error rate, OCR process; different raw data available? Used software?	n/a
Administration, cleanups,	e.g. handling of missing data, cutting, rescaling, NLP preprocessing, used software	n/a
Scope /Size	size of data records	<ul style="list-style-type: none"> • 418 books • c. 200,000 pages • c. 14,000 chapters • Content retrieved at chapter level
Metadata	Format/ Standards,	<ul style="list-style-type: none"> • Standard OCLC MARC records
Rights	licenses for metadata, full texts (TDM), rights / use (e.g. on-site, groups, scientific use)	<ul style="list-style-type: none"> • TDM can be carried out by any user with legal rights to access the content without any additional agreements or licences
Ethical Issues	Personal and / or Confidential Information; Bias / representation; offensive / insulting / sensitive content	<ul style="list-style-type: none"> • Scholarly historical works covering potentially sensitive cultural and ethical issues.
Use	Recommendations for use/ not recommended use	<ul style="list-style-type: none"> • Historical reference and research
Text and Data Mining	Additional costs? If so, how much? Trial possible?	<ul style="list-style-type: none"> • None • See https://www.cambridge.org/core/services/open-research/text-and-data-mining • If you need help accessing a large amount of content for a text and data mining project, please contact openresearch@cambridge.org.

Stand 30.3.2022