Datenbank: Romanische Bibliographie

Provider: De Gruyter

		Romanische Bibliographie
Access	Web address, API, Dumps, offline back up copy	• https://www.degruyter.com/document/doi/10.1515/rom/html • API access is not available
Documentation	Web address	• n/a
Distribution		• Continuously
Scope	Content Purpose Field of use	• Only comprehensive specialist bibliography for this discipline, provides electronic access to the entries of the volumes 1965–2019, categorized according to the tried-and-tested structure tree • Approximately 10,000 new entries added continuously every year
Time, Place, Language	temporal, local reference	• Bibliographic entries from the past 55 years • German/English/French
Data type	What are the basic data types?	• XML files
Provenance, dependencies, accompanying material	original data source, manufacturer, data collection procedure, dependencies / links to other data sets / online resources, old versions	• The definitive dataset is kept in a proprietary XML format, formatted by the De Gruyter Bibliography DTD, and from this its online datasets are derived. A DTD file can be provided and the fields are comparable to other standard bibliographic standards.
Description Structured text data	Text markup or data structure e.g. TXT, XML, ALTO, TEI, versions	• Each XML entry contains bibliographic information for a specified work
Description of databases, tabular data	data tables, existing / recommended data splits (e.g. training / test set)	• n/a
Description of image formats	as precisely as possible (e.g. resolution, greyscale / bitonal)	• n/a
Standards, vocabularies	as precisely as possible: standards and vocabularies used	• No standard vocabularies where used but the Romanische Bibliographie has its own classification system that accompanies each entry and can be provided separately and standardized keywords are also included
Data quality: OCR; missing,	For example. OCR error rate, OCR	• n/a

incorrect, redundant data, noise	process; different raw data available? Used software?	
Administration, cleanups,	e.g. handling of missing data, cutting, rescaling, NLP preprocessing, used software	• n/a
Scope /Size	size of data records	• Over 400,000 entries • Around 615 MB of XML files
Metadata	Format/ Standards,	• Bibliographic standard, see above
Rights	licenses for metadata, full texts (TDM), rights / use (e.g. on-site, groups, scientific use)	• Institutions have rights for non-commercial use by Authorised Users of the institutions only.
Ethical Issues	Personal and / or Confidential Information; Bias / representation; offensive / insulting / sensitive content	• n/a
Use	Recommendations for use/ not recommended use	• All purposes of TDM
Text and Data Mining	Additional costs? If so, how much? Trial possible?	• Small cost for delivering the data on hard drives.

Stand 30.3.2022