

**Datenbank:** Germanistik Online

**Provider:** De Gruyter

		Germanistik Online
Access	Web address, API, Dumps, offline back up copy	<ul style="list-style-type: none"> <li>• <a href="https://www.degruyter.com/document/doi/10.1515/ogerm/html">https://www.degruyter.com/document/doi/10.1515/ogerm/html</a></li> <li>• API access is not available</li> </ul>
Documentation	Web address	<ul style="list-style-type: none"> <li>• n/a</li> </ul>
Distribution		<ul style="list-style-type: none"> <li>• Continuously</li> </ul>
Scope	Content Purpose Field of use	<ul style="list-style-type: none"> <li>• Leading international bibliography covering German Language and Literary Studies covering more than 50 years of German philology</li> <li>• Critically reviewed, systematized and keyworded</li> <li>• Close to 400,000 monographic studies, collected volumes and journals from the areas of literature, theatre studies, media studies, cultural history, and linguistics</li> <li>• Continually expanded annually to include the current volume containing about 8,000 new entries</li> </ul>
Time, Place, Language	temporal, local reference	<ul style="list-style-type: none"> <li>• Bibliographic entries from the past 50 years</li> <li>• German/English</li> </ul>
Data type	What are the basic data types?	<ul style="list-style-type: none"> <li>• XML files</li> </ul>
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"> <li>• 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.</li> </ul>
Description Structured text data	Text markup or data structure e.g. TXT, XML, ALTO, TEI, versions	<ul style="list-style-type: none"> <li>• Each XML entry contains bibliographic information for a specified work</li> </ul>
Description of databases, tabular data	data tables, existing / recommended data splits (e.g. training / test set)	<ul style="list-style-type: none"> <li>• n/a</li> </ul>
Description of image formats	as precisely as possible (e.g. resolution, greyscale / bitonal)	<ul style="list-style-type: none"> <li>• n/a</li> </ul>
Standards, vocabularies	as precisely as possible: standards and vocabularies used	<ul style="list-style-type: none"> <li>• No standard vocabularies were used but the Germanistik Online has its own classification system that accompanies each entry and can be provided</li> </ul>

		separately and standardized keywords are also included
Data quality: OCR; missing, incorrect, redundant data, noise	For example. OCR error rate, OCR process; different raw data available? Used software?	<ul style="list-style-type: none"> <li>• n/a</li> </ul>
Administration, cleanups,	e.g. handling of missing data, cutting, rescaling, NLP preprocessing, used software	<ul style="list-style-type: none"> <li>• n/a</li> </ul>
Scope /Size	size of data records	<ul style="list-style-type: none"> <li>• Close to 400,000 entries</li> <li>• Around 500 MB of XML files</li> </ul>
Metadata	Format/ Standards,	<ul style="list-style-type: none"> <li>• Bibliographic standard, see above</li> </ul>
Rights	licenses for metadata, full texts (TDM), rights / use (e.g. on-site, groups, scientific use)	<ul style="list-style-type: none"> <li>• Institutions have rights for non-commercial use by Authorised Users of the institutions only.</li> </ul>
Ethical Issues	Personal and / or Confidential Information; Bias / representation; offensive / insulting / sensitive content	<ul style="list-style-type: none"> <li>• n/a</li> </ul>
Use	Recommendations for use/ not recommended use	<ul style="list-style-type: none"> <li>• All purposes of TDM</li> </ul>
Text and Data Mining	Additional costs? If so, how much? Trial possible?	<ul style="list-style-type: none"> <li>• Small cost for delivering the data on hard drives.</li> </ul>

Stand 30.3.2022