

**Datenbank:** American Periodicals

**Provider:** ProQuest

		American Periodicals
Access	Web address, API, Dumps, offline back up copy	Online access via <a href="https://www.proquest.com/americanperiodicals">https://www.proquest.com/americanperiodicals</a>
Documentation	Web address	Provider Website: <a href="https://proquest.libguides.com/americanperiodicals">https://proquest.libguides.com/americanperiodicals</a>
Distribution		<ul style="list-style-type: none"> <li>Sold in the higher education library market globally</li> </ul>
Scope	Content Purpose Field of use	<ul style="list-style-type: none"> <li>1509 newspapers, magazines and scientific journals from North America (1741-1988), especially:               <ul style="list-style-type: none"> <li>89 journals of the 18<sup>th</sup> century</li> <li>118 periodicals published during the Civil War (1861-1865) and Reconstruction (1865-1877)</li> <li>more than 20 women magazines</li> </ul> </li> <li>Thematic scope:               <ul style="list-style-type: none"> <li>History, Literature, Education, Art</li> <li>History of Science and Medicine</li> <li>Law, Politics, Religion</li> <li>News and Magazines</li> <li>American Studies, Women's Studies</li> </ul> </li> </ul> <p>American Periodicals includes digitized images of the pages of American magazines and journals published from colonial days to the dawn of the 20th century. American Periodicals encompasses <b>two separate collections</b>, which may be acquired individually: American Periodicals Series (APS) Online and American Periodicals from the Center for Research Libraries.</p> <p>Because American Periodicals contains digitized images of periodical pages, researchers can see all of the original typography, drawings, graphic elements, and article layouts exactly as they were originally published. To learn more about the separate collections available in <b>American Periodicals</b>, please see the <a href="#">American Periodicals Construction</a> page.</p>
Time, Place, Language	temporal, local reference	<ul style="list-style-type: none"> <li>All publications published in North America, between 1741 and 1988</li> </ul>
Data type	What are the basic data types?	<ul style="list-style-type: none"> <li>PDF page-image articles, displayed as individual article zones, or as full-page images.</li> <li>Article-level metadata records and metadata giving information about (and providing navigation for) the</li> </ul>

		content at the publication, issue, article, and page (zoning) level, based on underlying XML files for each article. XML files include OCR text with page word coordinates.
Provenance, dependencies, accompanying material	original data source, manufacturer, data collection procedure, dependencies / links to other data sets / online resources, old versions	Metadata created by ProQuest pertaining to the source publications scanned from microfilm or print. No links out in the database to external / alternative data sets
Description Structured text data	Text markup or data structure e.g. TXT, XML, ALTO, TEI, versions	Metadata created to ProQuest specification in the form of underlying XML files for each article.
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)	<ul style="list-style-type: none"> <li>American Periodicals Series contains article-level PDFs consisting of bitonal images scanned from microfilm. American Periodicals from the Center for Research Libraries also presents article-level PDFs, but the images were scanned in colour from print originals.</li> </ul>
Standards, vocabularies	as precisely as possible: standards and vocabularies used	Publication-level subjects based on Library of Congress terms
Data quality: OCR; missing, incorrect, redundant data, noise	For example. OCR error rate, OCR process; different raw data available? Used software?	Specification required that OCR for prominent fields (title, author, abstract, etc) be cleaned up to at least 99.95%. Background article text OCR confidence rating not available, but due to variable age and condition of original source materials, it is unlikely that this would be higher than 80%.
Administration, cleanups,	e.g. handling of missing data, cutting, rescaling, NLP preprocessing, used software	Batch validation performed (on image and data quality.)
Scope /Size	size of data records	~10M pages, 11.49M article records
Metadata	Format/ Standards,	<ul style="list-style-type: none"> <li>All non-image components of each article described in an XML file (one per article), with page zones information (and zone location coordinates). XML tags in data indicate the page on which the article starts and the page on which the zone is located. The XML includes the OCR article text with word coordinates. Article elements (e.g. title, author, issue date) are specified and article type values are</li> </ul>

		included.
Rights	licenses for metadata, full texts (TDM), rights / use (e.g. on-site, groups, scientific use)	Standard terms and conditions for the licence and use of ProQuest research databases: <a href="https://about.proquest.com/about/terms-and-conditions.html">https://about.proquest.com/about/terms-and-conditions.html</a> (includes information about TDM)  Supplemental terms of use for TDM Studio: <a href="https://about.proquest.com/about/Supplemental-Terms-of-Use-TDM-Studio.html">https://about.proquest.com/about/Supplemental-Terms-of-Use-TDM-Studio.html</a>
Ethical Issues	Personal and / or Confidential Information; Bias / representation; offensive / insulting / sensitive content	Low risk of personal/confidential information in this content set. Material spanning 248 years, dating from the 18 <sup>th</sup> century (including satirical content) may sometimes contain opinions at variance with modern sensibilities, but generally low risk of offensive content.
Use	Recommendations for use/ not recommended use	See permitted uses sections at <a href="https://about.proquest.com/about/terms-and-conditions.html">https://about.proquest.com/about/terms-and-conditions.html</a> for ProQuest's research databases and <a href="https://about.proquest.com/about/Supplemental-Terms-of-Use-TDM-Studio.html">https://about.proquest.com/about/Supplemental-Terms-of-Use-TDM-Studio.html</a> for TDM Studio
Text and Data Mining	Additional costs? If so, how much? Trial possible?	Trial is possible for 30 days. PQ offers his own TDM solution called TDM Studio, which is available for annual subscription at USD 10K for 1 workbench shared between up to 5 researchers or unlimited number of workbenches for USD 30K. Trial for ProQuest TDM Studio is also possible for 30 days.  If customer prefers instead to use their own TDM solutions, there is a need to purchase the raw data of the product on a hard drive for an additional amount. We will provide you with a proposal, if you request it.

Stand 30.3.2022