

**Datenbank:** UK Parliamentary Papers

**Provider:** ProQuest

		UK Parliamentary Papers
Access	Web address, API, Dumps, offline back up copy	Online access via Proquest Platform: <a href="https://parlipapers.proquest.com">https://parlipapers.proquest.com</a>
Documentation	Web address	<a href="https://proquest.libguides.com/parliamentary">https://proquest.libguides.com/parliamentary</a>
Distribution		<ul style="list-style-type: none"> <li>Sold in the higher education market globally</li> </ul>
Scope	Content Purpose Field of use	<p>There are several modules included in the U.K. Parliamentary Papers database.</p> <p>Modules in the U.K. Parliamentary Papers</p> <ul style="list-style-type: none"> <li>House of Commons 18th Century *</li> <li>House of Commons 19<sup>th</sup> Century</li> <li>House of Commons 20<sup>th</sup> Century</li> </ul> <hr/> <p>* The 18th Century Collection includes the following materials:</p> <ul style="list-style-type: none"> <li>Harper Collection of Private Bills (1695-1814)</li> <li>House of Commons Sessional Papers (1715-1800)</li> <li>House of Lords Sessional Papers (1714-1805)</li> <li>Journals of the House of Commons (1688-1834)</li> <li>Journals of the House of Lords (1688-1834)</li> <li>Local and Personal Acts (1797-1834)</li> <li>Parliamentary Register (1780-1796)</li> </ul> <p><b>House of Commons Parliamentary Papers</b> are foremost among the richest and most detailed primary sources for the past three centuries, for Britain, its colonies and the wider world. As such, they constitute a major part of the world's historical record. Parliamentary Papers influenced public opinion and social and political philosophy and provided a forum for the ideas of thinkers of the day.</p> <p>House of Commons Parliamentary Papers (HCPP)</p>

		<p>includes the complete file of House of Commons Parliamentary Papers, also known as Sessional Papers or Blue Books, dating from 1715 through 2005.</p> <p>This includes:</p> <ul style="list-style-type: none"> <li>• Bills – drafts of legislation, to be reviewed through various parliamentary stages.</li> <li>• House of Commons Papers – documents resulting from the work of the House of Commons.</li> <li>• Command Papers – government papers conveying information or decisions the Government wishes to draw to attention of the House, presented ‘by Command of Her Majesty’.</li> </ul> <p>The latter two categories breakdown further into:</p> <ul style="list-style-type: none"> <li>• Reports of Committees – Select or the Whole House. Appointed to investigate issues of concern.</li> <li>• Reports of Commissioners – commissions appointed by the Crown to investigate social problems, to conduct inquiries into events, and as a preparation for legislation.</li> <li>• Accounts – statistical information, originating primarily from the Treasury, the Board of Trade and the War Office/Admiralty.</li> <li>• Papers – correspondence from ambassadors, governors, army officers abroad; commercial, trade and navigation accounts, statistical abstracts: judicial, taxation, etc; census data; slavery and slave trade documents; treaties.</li> </ul> <p>Plus, the 18<sup>th</sup> Century Collection includes pre-Hansard debates, House of Lords and House of Commons journals, rare private bills, other legislative materials, and reports and papers presented to both Houses.</p>
Time, Place, Language	temporal, local reference	<ul style="list-style-type: none"> <li>• Publications are predominantly published in English. However, there are a select number of treaties and foreign agreements that are also in other languages. However, in those cases, there is also an English translation.</li> </ul>
Data type	What are the basic data types?	<ul style="list-style-type: none"> <li>• XML files containing metadata and full text.</li> <li>• Document level PDF.</li> </ul>

Provenance, dependencies, accompanying material	original data source, manufacturer, data collection procedure, dependencies / links to other data sets / online resources, old versions	18 <sup>th</sup> C collection: Mainly the print collection at the University of Southampton, with gap filling from the collections of the British Library and Cambridge University. 19 <sup>th</sup> and 20 <sup>th</sup> C collections: Mainly from the microform collection compiled by Chadwyck Healey. This microform collection was compiled from the Parliamentary Papers previously housed at the Department of Trade and industry, with gap filling from the collections of the House of Commons Library, the Home Office, the British Library, and Cambridge University.
Description Structured text data	Text markup or data structure e.g. TXT, XML, ALTO, TEI, versions	XML is encoded as UTF-8 and created to ProQuest specification. Each collection (18 <sup>th</sup> , 19 <sup>th</sup> , and 20 <sup>th</sup> C) is delivered as one XML file containing all the metadata and full text for each record within that particular collection.
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)	Images are contained within document level coordinate text PDFs.
Standards, vocabularies	as precisely as possible: standards and vocabularies used	18 <sup>th</sup> C: no subject indexing 19 <sup>th</sup> C : subject indexing is the “Cockton Index”, a Chadwyck Healey gold standard of subject indexing created by Peter Cockton who compiled and harmonized >100 Parliamentary sessional indexes into the preeminent Parliamentary controlled vocabulary. 20 <sup>th</sup> C: Subject indexing is based on the “Cockton Index” with additions and modifications to suit changing circumstances of 20 <sup>th</sup> C content and events.
Data quality: OCR; missing, incorrect, redundant data, noise	For example. OCR error rate, OCR process; different raw data available? Used software?	OCR confidence ratings are not available for these Parliamentary Collections (digitization undertaken over 15+ years ago), but due to variable age and condition of original source materials, OCR confidence is dependent upon the quality of the original source item.
Administration, cleanups,	e.g. handling of missing data, cutting, rescaling, NLP preprocessing, used software	Post-scan image enhancement includes: Cropping: Pages are cropped to same dimensions within a volume, unless the page size or layout changes. Skew: Each image rotated to ensure that lines of printed text are close to the horizontal. Batch validation performed (on image and data quality.)
Scope /Size	size of data records	Total: 251,000 document records and 10.3M pages.

		<ul style="list-style-type: none"> <li>• 18th C: &gt;58,000 documents and nearly 600k pages</li> <li>• 19th C: &gt;87,000 documents and 4.1M pages</li> <li>• 20th C: nearly 105,000 documents and 5.6M pages</li> </ul>
Metadata	Format/ Standards,	<p>Each collection (18<sup>th</sup>, 19<sup>th</sup>, and 20<sup>th</sup> C) is delivered as a large xml file (contained within a compressed *.gz file) containing all the records for that particular century collection. Metadata includes the document's URL (containing the accession number), date, title, author, and other metadata elements as well as the fulltext. For 19<sup>th</sup> and 20<sup>th</sup> C collections, subject terms are also included.</p> <p>Individual PDFs are enclosed in compressed *.tar files. Each PDF is named with the same accession number ID that is reflected in the url segment in the corresponding xml record.</p> <p>MARC records for the Parliamentary Papers are available through the ProQuest Administrator Module (a.k.a. the PAM).</p> <p>To retrieve MARC records for your purchases, please go to the PAM and click the link in the left column for Title Lists/MARC records, and then on the link Request MARC records.</p> <p>MARC records for 19<sup>th</sup> and 20<sup>th</sup> C are on the individual document level. For the 18<sup>th</sup> C collection, there are 7 MARC records based on the subcollection level (noted above—Harper Collection, House of Commons Sessional Papers, etc)</p>
Rights	licenses for metadata, full texts (TDM), rights / use (e.g. on-site, groups, scientific use)	<p>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)</p>
Ethical Issues	Personal and / or Confidential Information; Bias / representation; offensive / insulting / sensitive content	<p>Low risk of personal/confidential information in this content set. Material spanning from the late-17<sup>th</sup> century through 2005 may sometimes contain descriptions of working and medical conditions or opinions at variance with modern sensibilities, but generally low risk of offensive content.</p>
Use	Recommendations for use/ not recommended use	<p>Research in social, economic, and political history, international relations, and other research disciplines. 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></p>

		<p><a href="#">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</p>
Text and Data Mining	Additional costs? If so, how much? Trial possible?	<p>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.</p> <p>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.</p>

Stand 30.3.2022