**Datenbank:** British Periodicals

**Provider:** ProQuest

| | | British Periodicals |
|---|---|---|
| | | |
| Access | Web address, API, Dumps, offline back up copy | Online access via ProQuest Platform at https://www.proquest.com/britishperiodicals/index |
| Documentation | Web address | Provider Website: https://about.proquest.com/products-services/british_periodicals.html#resources |
| Distribution | | • Sold in the higher education library market globally |
| Scope | Content Purpose Field of use | Facsimile page images and searchable full text for 491 British periodicals published from the seventeenth century through to the twentieth century.<br><br>British Periodicals is available in four separate collections, each of which can be purchased separately.<br><br>• British Periodicals Collection I consists of more than 160 journals that comprise the ProQuest microfilm collection Early British Periodicals, the equivalent of 5,238 printed volumes containing approximately 3.1 million pages. Topics covered include literature, philosophy, history, science, the fine arts, and the social sciences.<br>• British Periodicals Collection II consists of more than 300 journals from the ProQuest microfilm collections English Literary Periodicals and British Periodicals in the Creative Arts together with additional titles, amounting to almost 3 million pages. Topics covered include literature, music, art, drama, archaeology, and architecture.<br>• British Periodicals Collection III is the long-awaited follow-up to British Periodicals I-II and extends the scope of the series by focusing on the first half of the twentieth century. The titles are from the prestigious stable of illustrated periodicals known as the "Great Eight" in British publishing history. They are considered to be among the foremost popular periodicals of the period and were highly influential in their mix of news/politics, miscellany, art, photography, literature and comedy/satire, while launching the careers of many leading artists/illustrators of the age.<br>• British Periodicals Collection IV will build on |

| | | Collection III by continuing the expansion of the programme into the twentieth century, offering full archival coverage of several of the leading magazines of the period. Titles are: The Field (1853-2005), Tribune (1937-2005), The Highway (1903-1959), War Weekly (1939-1941), Answers to Correspondents (1888-1955), Humorist (1922-1940), Wide World Magazine (1898-1965), The Marvel (1893-1922), Top Spot (1958-1960). <br><br> By providing digital access to the key primary source serials from this period, British Periodicals opens up research to a much wider audience. This uniquely powerful multidisciplinary database gives students and researchers a clear pathway to an exhaustive body of content previously unavailable online. All of this material is available in page image format with fully searchable text. Users can filter results by article type and download articles as either PDFs or JPEG page images. Each facsimile page from Collections I-II is available in high-resolution grayscale, providing scholars with an authentic and nuanced record of the printed source. Collection III introduces full color images where color was present in the original print volume. <br><br> Institutions with access to ProQuest's edition of the Wellesley Index to Victorian Periodicals can instantly cross-search Wellesley indexing with the full text from the British Periodicals interface. Crucially, the addition of Wellesley also makes it possible to search for instances of a word or phrase in a given author's contributions to periodicals even where these originally appeared unsigned or under a pseudonym. Altogether, British Periodicals provides full text to all 43 important Wellesley titles. <br><br> In addition, institutions with access to other historical periodical resources from ProQuest such as *American Periodicals* and *Periodicals Archive Online,* may search across all of this content simultaneously, querying potentially thousands of historical periodicals with a single search. <br><br> The titles were selected by Professor Daniel Fader of the University of Michigan on the basis of their importance and difficulty of access in American libraries, and issues were sourced from over 100 libraries worldwide. <br><br> Rarity of source material means that there will be some gaps in the collection. |
|---|---|---|
| Time, Place, | temporal, local | Publications predominantly originally published in the United |

| Language | reference | Kingdom, with a small number of titles from Ireland and India. Original publications published between 1681 to 2007 |
|---|---|---|
| Data type | What are the basic data types? | JPEG page images (with article-level or page-level PDFs created on-the-fly also available to users). Article-level metadata records and metadata giving information about (and providing navigation for) the 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) | JPEG image files at 300dpi resolution. Collections I-II digitized in greyscale from microfilm originals. Collections III-IV digitized in greyscale or colour (wherever there was colour on the original printed page) from print originals. |
| 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, caption, etc) be cleaned up to at least 99.95%. Background article text OCR confidence rating requirement for British Periodicals Collections III & IV was minimum 80%. OCR confidence ratings not available for Collections I-II (digitization undertaken 15+ years ago), 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 | Post-scan image enhancement includes: Cropping: Pages will be 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 | 6.7M article records; ~7.5M page images |

| Metadata | Format/ Standards, | 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 and document feature values are 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: https://about.proquest.com/about/terms-and-conditions.html (includes information about TDM)<br><br>Supplemental terms of use for TDM Studio: https://about.proquest.com/about/Supplemental-Terms-of-Use-TDM-Studio.html |
| 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 three centuries, dating from the late-17$^{th}$ 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 https://about.proquest.com/about/terms-and-conditions.html for ProQuest's research databases and https://about.proquest.com/about/Supplemental-Terms-of-Use-TDM-Studio.html for TDM Studio |
| Text and Data Mining | Additional costs? If so, how much? Trial possible? | Trial is possible for 30 days.<br>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.<br>Trial for ProQuest TDM Studio is also possible for 30 days.<br><br>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