

Datenbank: Nineteenth Century Collections Online: Europe and Africa

Provider: Gale Cengage

| | | Nineteenth Century Collections Online: Europe and Africa |
|---------------|---|--|
| Access | Web address, API, Dumps, offline back up copy | <ul style="list-style-type: none"> text-mining drives (includes directories, title manifests, XML files and image files, containing metadata, article segmentation, and page facsimiles (fee, available only for content the UB subscribes to or has purchased)) User can create batches of specific issues or titles for bulk download through the Gale Digital Scholar Lab (subscription service) API access is not available |
| Documentation | Web address | https://link.gale.com/apps/NCCO?u=unibern |
| Distribution | | <ul style="list-style-type: none"> Only as an entire collection, closed |
| Scope | Content Purpose Field of use | <ul style="list-style-type: none"> Monographs, manuscripts, and newspaper accounts relating to European colonial conquest, policy, economy and missionary activity in Africa. Full list of collections included: <ul style="list-style-type: none"> African Mail African World Annales de l'Extreme Orient et de l'Afrique Bulletins du Comite de l'Afrique francaise Colonial Africa Newspapers from the British Library Despatches from U.S. Consuls in Bathurst, Gambia, British Africa, 1857-1889 Despatches from U.S. Consuls in Boma, Congo, 1888-1895 Despatches from U.S. Consuls in Cape Town, Cape Colony, 1800-1906 Despatches from U.S. Consuls in Gaboon, 1856-1888 Despatches from U.S. Consuls in Goree Dakar, French Africa, 1883-1906 Despatches from U.S. Consuls in Grand Bassa, Liberia, 1868-1882 Despatches from U.S. Consuls in Lourenco Marques, Mozambique, Portuguese Africa, 1854-1906 Despatches from U.S. Consuls in Pretoria, The Transvaal, 1898-1906 Despatches from U.S. Consuls in Santiago, Cape Verde Islands, 1818-1898 Despatches from U.S. Consuls in Sierra Leone, British Africa, 1858-1906 Despatches from U.S. Consuls in St. Helena, British West Africa, 1831-1906 Despatches from U.S. Consuls in St. Paul de Loanda, Portuguese Africa, 1854-1893 |

| | | |
|---|---|---|
| | | <p>Despatches from U.S. Consuls in Tamatave, Madagascar, 1853-1906</p> <p>Despatches from U.S. Consuls in Zanzibar, British Africa, 1836-1906</p> <p>Deutsche Kolonialzeitung</p> <p>Exploration and Colonization of Africa</p> <p>German Colonialism in Africa: Selections from the Kolonialbibliothek (Colonial Library) Collection</p> <p>International Population Census, Africa</p> <p>Joseph Chamberlain Papers Relating to Africa</p> <p>Methodist Episcopal Church Missionary Correspondence, 1846-1912: Africa</p> <p>Personal Narratives and Reminiscences of the Exploration of Africa</p> <p>Publications of the Cape Colony</p> <p>Publications of the Natal Colony</p> <p>Report of the Boundary Survey between British Bechuanaland and German Southwest Africa of 1898-1903</p> <p>Selected Africa Collections from the Bodleian Library</p> <p>Selected Cabinet Papers on Africa</p> <p>Selected Colonial Office Files on Africa</p> <p>Selected Dominions Office Files on Africa</p> <p>Selected Foreign Office Files on Africa</p> <p>Selected India Office Records on Colonial Africa</p> <p>Selected Maps and Plans Files on Africa</p> <p>Selected Narratives and Reminiscences on Colonial Africa</p> <p>Selected War Office Files on Africa</p> <p>Yale Divinity School Periodicals on Africa</p> |
| Time, Place, Language | temporal, local reference | <p>«Long» 19th Century i.e. 1789-1914</p> <p>Africa, Britain, Germany, France</p> <p>English, German, French</p> |
| Data type | What are the basic data types? | <ul style="list-style-type: none"> • Facsimiles: JPEG • Document and issue text files with structural mark up (pages, subdivided or zoned into articles): XML • bibliographic information: XML, partly within issue text files |
| Provenance, dependencies, accompanying material | original data source, manufacturer, data collection procedure, dependencies / links to other data sets / online resources, old versions | <p>A DTD file is provided on the text-mining drives (not online) and the fields are comparable to those found in Dublin Core, MARC and other standard bibliographic standards</p> <p>The definitive dataset is kept in a proprietary XML format, known as the Gale Interchange Format or GIFT, and from this its text-mining and online datasets are derived.</p> |
| Description Structured text data | Text markup or data structure e.g. TXT, XML, ALTO, TEI, versions | <p>Periodical/Newspaper content: The TDM files are three separate xml;</p> <ul style="list-style-type: none"> • a publication xml - The publication xml includes publication title metadata • an issue xml - includes the issue and article metadata • and a text xml - includes the full text OCR for each |

| | | |
|--|--|--|
| | | <p>article.</p> <p>Manuscript and Monograph content: The TDM files are two separate xml;</p> <ul style="list-style-type: none"> • a document xml - The document xml includes document level metadata • and a page xml - includes the full text OCR for each page. |
| 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"> • 400 PPI grayscale and colour jpeg |
| Standards, vocabularies | as precisely as possible: standards and vocabularies used | |
| Data quality: OCR; missing, incorrect, redundant data, noise | For example. OCR error rate, OCR process; different raw data available? Used software? | OCR confidence rating varies across the corpus. The corpus was digitised from a mixture of physical copies and microfilm. |
| Administration, cleanups, | e.g. handling of missing data, cutting, rescaling, NLP preprocessing, used software | ABBYY OCR engine used to create OCR text. |
| Scope /Size | size of data records | 820k pages 723 monographs 2,599 manuscripts 12,633 newspaper/periodical issues |
| Metadata | Format/ Standards, | <ul style="list-style-type: none"> • bespoke metadata schema developed by Gale • hand-keyed issue, page and article-level metadata • metadata fields: article title, article subheadings, attribution information, illustration captions, source page number, section headers • separate metadata files: 1. title or publication-level metadata (XML), 2. Issue or page-level metadata (XML) |
| 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; | Historical content from the 19 th century will contain views and material that may cause offense. |

| | | |
|----------------------|--|--|
| | offensive / insulting / sensitive content | |
| Use | Recommendations for use/ not recommended use | All purposes of TDM |
| Text and Data Mining | Additional costs? If so, how much? Trial possible? | Option 1: Small cost for delivering the data on hard drives Option 2: Annual subscription cost for access to the Gale Digital Scholar Lab |

Stand 30.3.2022