The Atlas of Digitised Newspapers

Citation for published version:
Beals, M, Bell, E, Cordell, R, Fyfe, P, Galina Russell, I, Hauswedell, T, Neudecker, C, Nyhan, J, Oiva, M, Pado, S, Peña Pimentel, M, Rose, L, Salmi, H, Terras, M & Viola, L, *The Atlas of Digitised Newspapers: Reports from Oceanic Exchanges*, 2020, Web publication/site.
https://doi.org/10.6084/m9.figshare.11560059

Digital Object Identifier (DOI):
10.6084/m9.figshare.11560059

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Loughborough University

The Atlas of Digitised Newspapers

Reports from Oceanic Exchanges

M. H. Beals and Emily Bell

with additional contributions by

Ryan Cordell, Paul Fyfe, Isabel Galina Russell, Tessa Hauswedell, Clemens Neudecker, Julianne Nyhan, Mila Oiva, Sebastian Padó, Miriam Peña Pimentel, Lara Rose, Hannu Salmi, Melissa Terras, and Lorella Viola

and special thanks to
Seth Cayley (Gale), Steven Claeyssens (KB), Huibert Crijns (KB), Nicola Frean (NLNZ),
Julia Hickie (NLA), Jussi-Pekka Hakkarainen (NLF), Chris Houghton (Gale), Melanie Lovell-Smith (NLNZ),
Minna Kaukonen (NLF), Luke McKernan (BL), Chris McPartland (NLA), Maaike Napolitano (KB),
Tim Sherratt (University of Canberra) and Emerson Vandy (NLNZ)

DOI: 10.6084/m9.figshare.11560059

Disclaimers: The wider project is funded by the 'Transatlantic Partnership for Social Sciences and Humanities 2016 Digging Into Data Challenge'. The research conducted by UK Institutions is funded by the Economic and Social Research Council (ESRC) and the Arts and Humanities Research Council (Grant Reference: ES/R004110/1). The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged.

Executive Summary

Between 2017 and 2019, *Oceanic Exchanges* (http://www.oceanicexchanges.org), funded through the Transatlantic Partnership for Social Sciences and Humanities 2016 Digging into Data Challenge (https://diggingintodata.org), brought together leading efforts in computational periodicals research from six countries—Finland, Germany, Mexico, the Netherlands, the United Kingdom, and the United States—to examine patterns of information flow across national and linguistic boundaries. Over the past thirty years, national libraries, universities and commercial publishers around the world have made available hundreds of millions of pages of historical newspapers through mass digitisation and currently release over one million new pages per month worldwide. These have become vital resources not only for academics but for journalists, politicians, schools, and the general public. However, these digitisation programmes share a critical weakness: the very creation of national newspaper collections obscures the fact that international news exchange was central to the nineteenth-century press.

The *Atlas of Digitised Newspapers and Metadata* is an open access guide to a selection of newspaper databases around the world. Its initial selection is limited in scope, being comprised of the ten databases (including the aggregator Europeana) for which we were able to secure access and licensing to the machine-readable data. Nonetheless, it aims to form the foundation of a wider mapping of collections beyond its current North Atlantic and Anglophone-Pacific focus. It brings together their histories and digitisation choices with a deeper look at the language of the digitised newspaper, the evolution of newspaper terminology and the variety of metadata available in these collections. It explores how machine-readable information about an issue, volume, page, and author is stored in the digital file alongside the raw content or text, and provides a controlled vocabulary designed to be used across disciplines, within academia and beyond.

This report draws upon multiple taxonomies: our own open access dataset (doi:10.6084/m9.figshare.11560110), which provides a full catalogue of metadata fields across the collections, academic and industry discussions of the newspaper as a journalistic form and historical artefact, digitisation guidelines and strategies, library websites, annual reports, interviews with librarians and digitisation providers and the data files themselves. The maps of this *Atlas* explore each of our overarching categories in detail, providing a selection of language variants, the technical definition we employed in the categorisation process, and notes on its usage across the collections and in the wider world of press history. This allows a greater understanding of how the term is currently being used in different ways by different groups, and allows researchers to navigate to the specific type of information they require and ascertain its availability across these collections. Each entry also includes technical information for obtaining this data across the collections, including data types, which often vary considerably, and XPaths for locating the information within that dataset. With this information, researchers should be able to understand the different structures of these collections and develop computational means for robustly comparing datasets to explore deeper and more meaningful research.

After using the *Atlas*, we hope that readers will begin to understand the great wealth of metadata available for digitised newspapers, much of which is comparable across collections, nations and languages. As we explored these collections, we found a sincere effort on the part of librarians, scholars and commercial providers to converge upon a knowledge system that allowed meaningful enquiry and reflected a consistent layer (if not a complete reproduction) of these historical artefacts. However, this seeming convergence does mask important outliers and divergent interpretations of key bibliographic and conceptual categories—divergences that we hope these maps will highlight, and encourage future digitisers to consider when building or expanding their databases. In sum, the rise of digitisation promises great opportunities for those who wish to engage with newspaper archives, but as with all historical archives, digital collections require researchers to be mindful of their shape, provenance and structure before any conclusion can be drawn. It is the responsibility of both digitiser and researcher to understand both the map and also the terrain.

Table of Contents

List of Abbreviations

ALTO	Analysed Layout and Text Object
ANDP	Australian Newspaper Digitisation Program
ANPLAN	The Australian Newspaper Plan
API	Application Programme Interface
ARROW	Australian Research Repositories Online to the World
BL	British Library
CA	Chronicling America
CCITT	International Telegraph and Telephone Consultative Committee
CIP	Competitiveness and Innovation Framework Programme (European Commission)
CLARIN	Common Language Resources and Technology Infrastructure
CMDI	Component MetaData Infrastructure
CCS	Content Conversion Specialists
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
DIDL	Digital Item Declaration Language
DNCJ	Dictionary of Nineteenth-Century Journalism
DTD	Document Type Definition
EDM	Europeana Data Model
ENMAP	Europeana Newspaper METS ALTO Profile
DDD	Databank Digitale Dagbladen
GIFT	Gale Interchange Format
GLAM	Galleries, Libraries, Archives and Museums
HNDM	Hemeroteca Nacional Digital de México
HTML	Hypertext Markup Language
ICC	International Colour Consortium
ID	Identifier
IE	Intellectual Entity
IIIF	International Image Interoperability Framework
IFLA	International Federation of Library Associations and Institutions
ISO	International Organisation for Standardisation
ISSN	International Standard Serial Number
JISC	Joint Information Systems Committee
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
KB	Koninklijke Bibliotheek (Royal Library of the Netherlands)
LOC	Library of Congress
LZW	Lempel–Ziv–Welch

MARC	Machine-Readable Cataloguing
METS	Metadata Encoding and Transmission Standard
MIX	Metadata for Images in XML Schema
MODS	Metadata Object Description Schema
MPEG	Moving Picture Experts Group
NDNP	National Digital Newspaper Program (America)
NEH	National Endowment for the Humanities
NISO	National Information Standards Organisation
NLA	National Library of Australia
NLNZ	National Library of New Zealand
NORDINFO	Nordic Council of Scientific Information
NSLA	National and State Libraries of Australasia
NTAS	Northern Territory Archives Service
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCR	Optical Character Recognition
OLR	Optical Layout Recognition
PANDORA	Preserving and Accessing Networked Documentary Resources of Australia
PICA	Project for Integrated Catalogue Automation
PP	Papers Past
PPI	Pixels Per Inch
PPN	PICA Production Number
PURL	Persistent Uniform Resource Locator
PREMIS	Preservation Metadata: Implementation Strategies
SBB	Staatsbibliothek zu Berlin (Berlin State Library)
SRGB	Standard Red Green Blue
SRU	Search/Retrieve via URL
TDA	Times Digital Archive
TEL	The European Library
UID	Unique Identifier
UNAM	Universidad Nacional Autónoma de México (National Autonomous University of Mexico)
URL	Uniform Resource Locator
XML	Extensible Markup Language
XPATH	XML Path Language
ZEFYS	Zeitungsinformationssystem der Staatsbibliothek zu Berlin (Newspaper Information System of the Berlin State Library)
ZDB	Zeitschriften Datenbank (Magazines Database)
ZVDD	Zentrales Verzeichnis Digitalisierter Drücke (Central Directory of Digitised Prints)

Introduction

The Project

This *Atlas* is a product of *Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories*, 1840–1914 (http://www.oceanicexchanges.org). The project was funded through the Transatlantic Partnership for Social Sciences and Humanities 2016 Digging into Data Challenge (https://diggingintodata.org) and was undertaken by researchers from Universität Stuttgart, Staatsbibliothek zu Berlin, Universidad Nacional Autónoma de México, Universiteit Utrecht, Turun Yliopisto, Loughborough University, University College London, University of Edinburgh, Northeastern University, North Carolina State University, and University of Nebraska-Lincoln between 2017 and 2019.

The dramatic expansion of newspapers over the nineteenth century created a global culture of abundant, rapidly circulating information. The significance of the newspaper has largely been defined in metropolitan and national terms in scholarship, while digitisation by local institutions further situates newspapers in national contexts. *Oceanic Exchanges* brought together leading efforts in computational periodicals research from six countries—Finland, Germany, Mexico, the Netherlands, the United Kingdom, and the United States—to examine patterns of information flow across national and linguistic boundaries. Through computational analysis, the project crosses the boundaries that separate digitised newspaper corpora to illustrate the global connectedness of nineteenth-century newspapers, uncovering how the international was refracted through the local as news, advice, vignettes, popular science, poetry, fiction, and more. By linking research across large-scale digital newspaper collections, *Oceanic Exchanges* offers a model for data custodians that host large-scale humanities data. Recent research from the *Always Already Computational: Collections as Data* project (2016–2018) has highlighted the need to reshape our understanding of digital collections and find ways to better support computational use of data, not as an afterthought but as part of the design of those data collections. Our *Atlas* is among the first major projects to align with the principles outlined by the *Collections as Data* project, including lowering barriers to use, sharing documentation, fostering interoperability, and doing so in an open access format.

The *Atlas of Digitised Newspapers and Metadata* arose out of the need for the data from different datasets to be transformed into a single unitary standard that could be inputted into project workflows across the project. At the most basic level, what was needed was a bespoke JSON format compatible with the text-matching software *passim* (https://github.com/dasmiq/passim), the primary tool used by the project to identify textual reappearances across the collections, as well as discrete plain-text files of the newspaper content. Although the basic bibliographical and content fields of a database could be quickly identified to allow for unique identification of specific texts, a deeper understanding of the meaning of the metadata—and therefore its full potential for digital research—was difficult to obtain. The different vocabularies (such as Dublin Core, METS, ALTO, PREMIS, MIX) were used inconsistently and combined differently in different instantiations. In order to meaningfully connect these collections, the researchers needed to bring these collections together—their data, metadata and paradata—and then examine them as research objects in and of themselves.

In 2017–18, led by Paul Fyfe of North Carolina State University, *Oceanic Exchanges* gathered together fourteen instantiations of ten distinct digitised newspaper databases, detailed below, alongside histories of their creation, composition and licensing. In 2018–19, a team led by M. H. Beals of Loughborough University worked to catalogue the data and metadata available across these collections, to undertake detailed interviews with data providers and libraries, and to develop a robust taxonomy for discussing the digitised newspaper not only as a facsimile but as a research object in its own right. This *Atlas* represents our current conception of this often-misunderstood research object, an ontology that describes the relationships between a database's internal components, between the data and metadata in different collections, and between the digital object and its physical predecessors.

Aims and Objectives

The *Atlas* aims to facilitate more historically informed understandings of digitised newspapers for researchers across disciplines. The nineteenth-century newspaper was a messy object, filled with an ever-changing mix of material in an innumerable number of amorphous layouts; digitised newspapers are no different. Each database contains a theoretically standardised collection of data, metadata, and images; however, the precise nature and nuance of the data is often occluded by the automatic processes that encoded it. Moreover, no

true universal standard has been implemented to facilitate cross-database analysis, encouraging digital research to remain within existing institutional or commercial silos. To overcome this, and to promote a remixing of discrete repositories, researchers must solve several technical and philosophical challenges.

At the start of the project, *Oceanic Exchanges* aimed to explore six digitised newspapers datasets from the North Atlantic, covering the nations of Mexico, the United States, the United Kingdom, the Netherlands, Germany, and Finland. These collections were chosen on two primary criteria. First, they were collections with which the Principal Investigators from each national team were intimately familiar, having previously investigated internal reprinting patterns with them. Second, the full machine-readable data and metadata sets was made available to the team as either public domain data or under widely available licensing agreements for text-mining. Once the project began, two additional English-language collections—Papers Past from the National Library of New Zealand and the (London) Times Digital Archive by commercial publisher Gale—were made available to the team. Because these datasets, alongside the Trove newspaper collection of the National Library of Australia, were well known to team members and were quickly accessible to researchers under clear licensing schema, they have been included in the *Atlas* despite being beyond its initial North Atlantic remit. Although we recognise that many other digitised newspaper databases currently exist, and that large sections of the globe are not represented in this guide—namely South America, Africa and Asia—we feel that the *Atlas* nonetheless represents the first international, multi-lingual collaboration of its kind, providing an exemplar of physically reconnecting siloed datasets and working with collections as data, their content, structure and metadata intimately bound together. Moreover, by opening the digital *Atlas* (http://www.digitisednewspapers.net) to public collaboration, we hope that additional collections will be chronicled and mapped by the wider community in the future.

Each database instantiation that we have surveyed uses different language to describe physical objects (for example a newspaper issue, edition or volume), layout terminology (article or advertisement, title or heading), and more abstract concepts (genre, document type). Moreover, each database organises these terms into different hierarchies of classification. Layers of nested items, containers, and technical metadata unique to different standards (and often unique to specific repositories) raise challenging questions about what data matters and what data can be dismissed as too technical to be of interest to the digital humanities researcher. Finally, although there is some truth to the claim that "everyone uses METS/ALTO", or something very similar, when encoding digitised newspapers, this surface level consistency lulls us into a false sense of security. We are rarely comparing apples to apples—sometimes we are not even comparing fruit to fruit.

However, it is not the aim of this *Atlas* to provide a single "better" standard for digitised newspaper data, a catalogue of what *should be* across all collections. Instead, our aim is to provide a specific type of map for this rough and often perilous terrain; to allow everyone, regardless of their previous experience, to explore these collections in relative safety. It is hoped that the electronic version of this *Atlas* will continue to be updated with the most recent surveys of these collections, as well as with additional databases. To this end, we hereby release the collection, open access, on GitHub, so that the community of periodical researchers and digitisers may strive not to homogenise our knowledge systems, but to make them mutually intelligible and navigable, now and in the future.

Methodology

The *Oceanic Exchanges* project aimed to bring together experienced researchers and rich data sources from around the world in order to better understand nineteenth-century patterns of news dissemination. Each of the partner institutions hosted researchers who had previously worked with at least one of the collections, and these scholars worked with data providers to obtain and, where needed, licence static versions of these databases for computational research across the project. The collections were hosted on a secure server by Northeastern University, which could be consulted remotely by project partners around the world, and samples from each of these collections were examined by the team at Loughborough University to catalogue their contents, working with the providing libraries and project members to ensure this sample reflected the wider database from which it was derived.

Our first step was to create a master elements list, which provided a single entry for every possible XPath or JSON path in every collection. This included information about type of data stored, the technical definition given in the documentation (if there was one), as well as interpretive definitions, based on the research experience of project members or close examination of multiple examples within the

database. The aim was to create a multi-rooted hierarchy, one that showed the position of every type of data within a given collection, by file and hierarchical placement within that file, as well as to assign to each a category, grouping metadata fields across collections. This work was not straightforward; many collections used similar schema (Dublin Core, METS, ALTO) but obtained and populated their data fields in different and often undocumented ways. Simple concepts like "title", "publisher" and "article" became very complicated, very quickly—for example, was it a standardised title conceived by the original cataloguer, or the title as it appeared on this specific physical object, or a variant title from a particular edition? Was the information labelled "article" the media type or the editorial genre?

In the end, working with these collections required a creative and flexible interpretation of these terms and a clear understanding of the history and character of specific digital files to inform these interpretations. We began by retro-engineering the implementation of these metadata standards, starting with document type definitions (DTDs) and schema specifications and then complementing these with documentation on the cataloguing standards used. Although official documentation was often publicly available, or directly packaged with the data, we also consulted grey literature such as promotional materials or online discussions by users on the nature of the data available. Other queries were approached through experimentation—testing hypotheses on the meaning of certain fields through an examination of a large sample of records, particularly the METS/ALTO variants in multiple countries.

To clarify apparent inconsistencies within the data and with our conceptions of these terms, the project team further explored the specific histories of cataloguing, microfilming and digitising historical newspapers in these nations. By building upon previous research by the *Oceanic Exchanges* team, we were able to develop a robust longitudinal understanding of how the data has been augmented or repackaged by different institutions in response to technical innovations and user feedback over the past twenty years. Finally, researchers from Loughborough University, University College London, the University of Edinburgh and la Universidad Nacional Autónoma de México conducted semi-structured interviews and engaged in correspondence with digitisation providers and libraries in Mexico, the United States, the United Kingdom, the Netherlands, Germany, Finland, Australia and New Zealand. The earnest engagement of these providers allowed us to better understand the histories of these collections, their selection and processing of historical newspapers, and their development or adaptation of metadata for describing them.

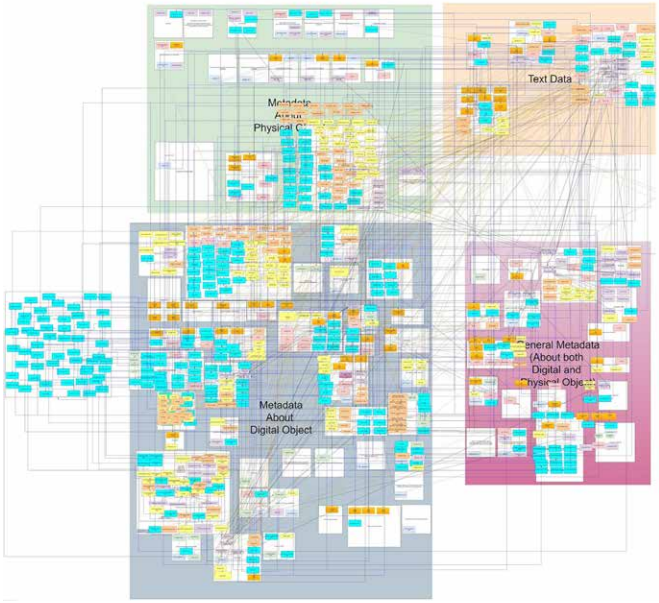

Fig. 1. Map of all metadata fields from our samples (each one represented by a different colour), with connecting lines showing the internal hierarchy of each, broken down by metadata of physical object, digital object, metadata pertaining to both, and text data. Unmapped blue boxes represent an overflow of repetitive administrative technical metadata.

As we finalised our master list of data and metadata fields, we attempted to visually group, or map, all possible elements across all collections, using the visualisation tool draw.io; we anticipated that the majority of fields would correspond directly to similar fields in other databases and thus a visual representation would be the clearest means of conveying the information overlap between collections. However, attempts to create a single map of all possible elements and attributes, to provide provenance of internal structures while grouping object by type and subtype, raised ontological issues not only in developing hierarchies and links between the fields of the various databases but in understanding the vocabulary employed to describe elements and characteristics of the physical newspaper as well.

As a result, we decided instead to create this textual ontology of the metadata fields based upon the structure of the newspaper as a physical object—a format that we hope will provide a deeper, more nuanced understanding of this ubiquitous and ambiguous medium. In the spirit of Ptolemy, this *Atlas* represents our attempt to provide future researchers with an initial field guide to exploring this heretofore unmappable terrain.

Database Histories

The following report abbreviates the histories of the ten newspaper databases used in the *Oceanic Exchanges* project, providing information on fourteen instantiations of the data across web interfaces, APIs and text mining drives. This includes a chronology of the development of each, a discussion of its aims and offerings, and a report on its current status and availability. These histories also include a discussion of the data within the collection, its overall structure and technical detail. The following repositories are discussed:

Database	Provider	Project Date	Version(s) Described	Metadata Standard
British Library 19th Century Newspapers	British Library; JISC; Gale: A Cengage Company	Pilot 2001; Main 2004	September 2013 December 2018	Bespoke XML & GIFT
Chronicling America	Library of Congress and Partner Collection Holders	2005	September 2017	METS/ALTO
Delpher	National Library of the Netherlands	Pilot 1999; Main 2007	November 2018	MPEG21/ALTO
Europeana	Europeana Foundation	2012	September 2018	ENMAP METS/ALTO
Hemeroteca Nacional Digital de México	Hemeroteca Nacional Digital de México	2002	October 2018	Bespoke JSON & METS/ALTO
Papers Past	National Library of New Zealand	2000	April 2018	METS/ALTO & Bespoke XML
Suomen Kansal-liskirjaston digi-toidut sanomalehdet	National Library of Finland	2001	October 2019	METS/ALTO
Times Digital Archive	Gale: A Cengage Company	2002	April 2018	GIFT
Trove	National Library of Australia	2007	February 2018	METS/ALTO & Bespoke XML
ZEFYS	State Library of Berlin	2009	December 2019	METS/ALTO

Metadata Maps

This report contains a series of maps to better align and compare data from different digitised newspaper collections. Each page is devoted to a low-resolution category of the metadata, such as *title, publisher* or *text*. Within that category, we have attempted to subdivide the relevant elements and fields into categories of metadata that are the most comparable across databases, for example *normalised title* or *individual word*.

Each section provides a technical definition of that category as well as an exploration of how the term, or variant terms, have been used by modern researchers in periodical studies, literary studies, library science and computer science (where appropriate) as well as in a nineteenth-century context. This is followed by a discussion of any exceptions or eccentricities regarding this category within the data collections. It concludes with a list of relevant XPaths, or other identifiers, and key information regarding the nature of the data in each element. With this information, the reader should be able to understand the different structures of these collections and develop computational means for robustly comparing datasets.

These maps are accurate to October 2019 for the specific the collection dataset listed above; however, it has been our experience that data providers frequently update, tweak or otherwise modify their metadata schema, both for new collections and in order to retrofit previous collections based on end-user feedback. We are also aware of specific forthcoming updates to several of these collections, details of which have not yet been made publicly available. It is therefore advisable that you consult with the data provider on their current schema before undertaking any data mining project.

Suggested Citation

Beals, M. H. and Emily Bell, with contributions by Ryan Cordell, Paul Fyfe, Isabel Galina Russell, Tessa Hauswedell, Clemens Neudecker, Julianne Nyhan, Mila Oiva, Sebastian Padó, Miriam Peña Pimentel, Lara Rose, Hannu Salmi, Melissa Terras, and Lorella Viola. *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges*. Loughborough University, 2020. doi: 10.6084/m9.figshare.11560059.

British Library 19th Century Newspapers

British Library, United Kingdom and Gale: A Cengage Company

History of the Collection

Parts I and II of the British Library's 19th Century Newspapers collection (https://www.gale.com/intl/c/british-library-newspapers-part-i and https://www.gale.com/intl/c/british-library-newspapers-part-ii), now part of the British Library Newspapers collection, were created as part of a public-private partnership between the British Library and Gale, a Cengage company. The British Library began developing a prototype system for newspaper digitisation in 2001, focusing on nineteenth-century newspapers. The project aimed for efficiency in digitising its newspapers through the development of automatic indexing, andsought to make the newspapers open to advanced searching. The initial focus was on newspapers outside of copyright, andeighteen microfilm reels of varying quality were selected as part of this test. Around 20,000 pages were processed in the first two months of the project.

The British Library's main efforts to digitise its newspaper collections, beginning in 2004, were funded by a £2 million grant from the United Kingdom's Joint Information Systems Committee (JISC). The project had an initial target of making 2 million pages available and broadly useful to scholars, researchers, and the public. The British Library partnered with commercial vendors to process the scanned images, including Gale and Brightsolid; the latter partner continues to expand the collection as the British Newspaper Archive (https://www.britishnewspaperarchive.co.uk), and the library itself is currently undertaking a large-scale digitising effort entitled *Heritage Made Digital*, neither of which are discussed here.

Consulted Libraries

The British Library's newspaper collection is based upon material obtained through legal deposit legislation. By law, a copy of every UK print publication must be given to the British Library by its publishers and to five other major libraries that request it. Since 1869 newspapers have been included within the legislation, and between 1820 and 1869, publishers were obliged to provide copies to the Stamp Office for the purposes of taxation; the latter were passed on to the British Museum and now form part of the British Library's collection. The original digitisation programme was exclusively derived from this collection.

Microfilming Projects

From the early 1940s to 2010, the usual method of preserving newspapers at the British Library was through the creation of access surrogates by microfilming. Approximately 30% of the newspaper collection was microfilmed during this time and, upon examination in the early 2000s, it was deemed that only 2% of the historical microfilm collection was unfit for digitisation by the library's Zeutschelmicrofilm cameras. Microfilming continued at the British Library alongside the original digitisation programme and was seen as an intermediary stage of the digitisation process rather than a replaced technology. Microfilming has been funded in different ways. Primarily it was funded by the BL, but external microfilm providers have also been used, notably MicroFormat (now a part of Stor-a-File), under contract to the BL. Microfilming of newspapers from other libraries was undertaken as part of a number of co-operative projects, most significantly Newsplan 2000, in which at-risk newspapers titles from libraries across the UK were microfilmed and distributed to the partner libraries, with master copies held by the BL and the National Library of Wales. The project was funded by the Heritage Lottery Fund, with additional financial support from the newspaper industry, and ran from 2000 to 2005 (though Newsplan 2000 as a body still exists). 1,325 newspaper titles (or 12,800,000 pages were microfilmed) producing 30,476 reels of microfilm.

Digitisation Projects

The British Library's initial nineteenth-century newspaper digitisation project took place in two phases. The first took place between 2004 and 2007, while the second ran from 2008 to 2009. The second phase was specifically aimed at expanding he digital collection's coverage of regional and local news, as well as including the eighteenth-century issues of existing titles. Owing to budget constraints and available technology, newspapers were not directly scanned to digital files in either phase of digitisation. Instead, new microfilms were made of newspapers, where needed, and these films were subsequently scanned. The exception to this was *The Standard*, which was scanned directly from paper copies at the Boston Spa repository. These in-house scanned images and microfilm reels were sent to external vendors, first Apex CoVantage (JISC I), followed by Content Conversion Specialists (JISC II), for processing, providing the library and Gale with an archival master for each page, as well as bitonal and greyscale images, and processed OCR text. Although this is a static collection, the BL has continued to

expand its newspaper digitisation: over 30 million pages have been produced since 2010 through the BL's relationship with Findmypast, augmented recently by the BL-funded Heritage Made Digital programme.

Selection

For the first newspaper digitisation project, the British Library opened an online consultation with academics forming an advisory group of library staff and scholars to develop a framework of titles that provided a representative image of the country on a given date; forty-eight titles were selected to provide a broad yet detailed view of British life in the nineteenth century. Focus groups and user panel meetings were not held for the second digitisation project because it was decided that all titles could be of interest to some users. Although the British Library's physical collections of historical newspapers are far more extensive, newspapers were selected for digitisation to provide a representative sample of the wider collection, covering the metropolitan and provincial press, ranging in political and geographical coverage, and representing both English- and Welsh-language titles.

Preservation and Access

The British Library digitises newspapers as part of its remit to preserve its collections. Its policy is to provide access through surrogates rather than the originals, where possible. Traditionally this has been done through microfilm, but the policy was updated in the early 2000s to create additional access copies through the digitisation of existing or new microfilm reels. Therefore, for the original nineteenth-century newspaper collections, digital copies do not act as the sole preservation copy of the newspapers but rather an additional form of access.

Composition of the Collection

Selection Available

The Gale 19th Century Newspapers collection contains sixty-nine distinct publications. Of these, twenty-one were published in London, thirty-three in England outside London, five in Scotland, five in Wales, and two in Ireland. Many of the titles published in Scotland, Wales, and Ireland are primarily held by their respective national libraries, which have pursued separate digitisation projects. The library aimed to provide the full date range of each selected title to the extent allowable by the physical collection and within the project criteria (1800–1900). Thus, titles such as the *Glasgow Herald*, which began publication in 1783 and continues today, was only digitised from 1820–1900, or from the first issue held by the British Library until the project cut-off date. Although the entire collection covers the period from 1800 to 1900, the number of titles increases substantially as the century progresses, with 10% of Part I being published before 1833, 10% of Part I being published before 1840, and 50% of both collections appearing after 1874.

Data Quality

Text

Each page was processed into machine-readable text by Prime OCR. Part I was processed by Apex, with the hand-keying or OCR correction of article titles. Part II was captured by Gale, who re-keyed articles titles for both parts and later commissioned the keying of article subheadings from an external contractor. Independent studies have suggested that the overall OCR quality of Part I and Part II is approximately 60-85%, but this varies widely within and between titles.

Image

For Part I of the collection, Apex provided an archival master file, in TIFF format, at 300 PPI and 8-bit greyscale, as well as lower resolution images for use on the web interface, including bitonal images of text blocks and greyscale images of illustrations or photographs, to facilitate use over dial-up modem technology. For Part II of the collection, the resolution was raised to 400 PPI and the package of images included an unedited archival TIFF, a slightly cropped, lossless JP2 or JPEG2000 master image, and a compressed JPEG for use on the web interface. Images of earlier and local newspapers were generally scanned to produce a lighter image to improve OCR word accuracy.

Metadata Schema

The data contained within the British Library Newspapers is available under three distinct metadata schemes: two provided by Gale and one held for project work by the British Library.

Gale Legacy Text Mining Drives

Before 2018, the Gale Text Mining Drive contained metadata and text content in a single XML file. Although similar in coverage to the METS/ALTO schema used by many public institutions, Gale established a bespoke metadata schema to label information consistently across its different newspapers and collections. A DTD file is provided on the text-mining drives and the fields appear to be adapted from Dublin Core, MARC and other standard bibliographical standards, to which they have been successfully mapped when working with external content partners.

Each XML file contains bibliographic information for a single issue, automatically zoned during the OCR process. The metadata for the issue is followed by the machine-readable text, in which each individual word is encoded with spatial coordinates of its location on the corresponding image, as well as marker elements indicating new pages or columns. The metadata was created partly through automatic processes and partly by direct input by contract workers.

Gale Current Text Mining Drives

After 2016, the Gale Text Mining Drive separated metadata and text content into three XML files: title or publication-level metadata, issue-level metadata, and issue-level content data. As with the previous schema, the data is encoded using Gale's standardised metadata schema and a DTD file is provided on the text-mining drives. Although distinct from the METS/ALTO schema, this system is similar to a combination of library MARC records and METS/ALTO XMLs.

British Library Project Drives

A pre-processed version of the data is held by British Library Labs and has been used by BL Labs Competition winners and award recipients in supported projects. This version of the XML is encoded at page rather than issue or article level. As with the Gale version, each word is encoded with the spatial coordinates. As it is encoded at page level, it does not contain the marker elements for page or column breaks. This provides a possibly more researcher-friendly variant of the XML, with human-readable element names and an intuitive nesting of elements, but lacks any form of delimitation between articles, which can be found in the Gale version.

Backend Structure

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. In addition to the metadata provided on text-mining drives or online, this database stores image metadata on resolution, file format, bit depth, colour map, file size and image dimensions. The image database stores image metadata, including image resolution, file format, bit depth, colour map, file size, width and height.

User Interface Structure

Web Interface

British Library Interface

Users of the collection through the British Library interface can perform basic or advanced searches of the collection, or browse by publication or location. The basic search can be filtered to a specific metadata field or the full text, a date range, a specific title or a specific digitised collection. The advanced search allows for standard Boolean operators and fuzzy searching as well as filtering by place of publication, issue section, title publication frequency, language and whether an image is included. Results can be sorted by publication date, article title, publication title or relevance, and further filtered by publication section or article type. Individual results can be viewed at article, page, or issue level. At article level, the searched terms will be highlighted; at page level, the article will be outlined in red. The image viewer allows users to navigate the issue and enlarge the image. The selection can be printed, emailed, downloaded or bookmarked using the interface at the top of the screen but there is no access to the underlying OCR text. A suggested citation, including a word count, is provided for each result.

Gale Primary Sources Interface

Users of the collection through the Gale Primary Sources interface can search using the same features as the British Library interface, with additional filters and simple analysis tools, such as topic finder and term frequency. Once selected, users are presented with the chosen article and options to navigate the issue or other search results. Users can adjust the image contrast and brightness to improve legibility and download it using standard browser context menus. The image may also be downloaded as a PDF or the plain text of the OCR content. Bibliographical information, images and text content can also be saved to cloud storage on Gale's servers or through integrations with Google Drive and OneDrive.

API

API access is not currently available through the British Library or Gale.

Direct Download or Drives

Gale Legacy Text Mining Drives

The previous version of Gale's Text Mining Drive for Part I had separated data into directories containing either scanned images or machine-readable texts. Images were contained in numerically labelled batch directories (for example, BLC_Images_Archive_01) in which there were individual directories labelled by the four-letter title abbreviation used in the XML metadata. Within these were directories labelled by individual issue dates in ISO format (for example, 18840102) inside which were the page images. Machine-readable text files were stored in a separate file structure, using numerically labelled batch numbers (for example, BLC_XML_Archive_14). Within these are all the XML files created as part of that processing batch.

The previous version of Gale's Text Mining Drive for Part II packaged image and text data together within a single data structure. Within numerically labelled batch directories, data was separated into directories using the four-letter abbreviation associated with a particular title. Within these were individual directories for specific dates, each of which contain one XML file, containing metadata and content data, and the individual page images, stored as JPGs.

Gale Current Text Mining Drives

The current version of Gale's Text Mining Drives separates data into image and XML data. Within these directories, the data is separated into processing batches, each with an individual alphanumeric code. Within these are directories for each individual title that has been digitised within that batch; a single title code might exist in multiple processing batch directories. Within this directory are either the metadata and content XML or image files for each issue. These are not separated by specific date ranges, but a full index of all issues, and their location within the structure, are provided on a spreadsheet file on the drive.

British Library Project Drives

XML data from the British Library Project Drives is divided into 524 ZIP files. These are not indexed or separated by title or date, and therefore complete decompression is required to ensure a full title or date range is extracted. Each file represents a single page and is named with the project code, the title abbreviation, the year, the month, the day, and the page of data.

Rights and Usage

Web Interface

In the United Kingdom, the collections can be freely accessed within the British Library reading rooms as well as remotely through Gale and British Library interfaces provided to all UK Higher Education Institutions (and some others) via JISC agreements. Outside the UK, collections are accessible by institutional purchase, including many public or national libraries; there is currently no individual purchase model available.

API

API access is not currently available through the British Library. However, users can create batches of specific issues or titles for bulk download through the Gale Digital Scholar Lab (https://www.gale.com/intl/primary-sources/digital-scholar-lab), a separate subscription service.

Direct Download or Drives

Gale Cengage makes content from its collections available to academic researchers for data mining and analysis through physical hard drives for a nominal "cost recovery" charge. This includes directories, title manifests, XMLs and image files. In

the United Kingdom, as part of the original agreement with JISC, the underlying data can be accessed by request and a cost recovery fee by all Higher Education Institutions. Elsewhere, the data is only accessible to those with institutional purchase to the relevant Gale products. Material obtained on text mining drives may be used to examine individual text or for large-scale analysis for purposes of performing personal or non-commercial research purposes, but cannot be duplicated or shared without express permission.

Re-Publication

As part of the user agreement, XML, OCR and image data cannot be re-published in any form, physical or digital, without the express permission and licensing of the British Library (web interface) or Gale (web interface and drives). Small quotations, using standard citation practices, may be reproduced in accordance with local fair use provisions and should be accompanied by a DOI link that points back to the individual full text article or book chapter and a proprietary notice in the following form: "Some rights reserved. This work permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited."

Chronicling America

Library of Congress, United States of America

History of the Collection

Chronicling America (https://chroniclingamerica.loc.gov) is a free, online repository of newspapers printed in the United States, primarily from 1836 to 1922. It is managed as part of the National Digital Newspaper Program, a collaboration between the National Endowment for the Humanities and the Library of Congress. The collection began its initial digitisation phase in 2005 and, as of December 2018, continues to support digitisation of new material. In addition to this digitised collection, the partnership also manages the national newspaper directory, which lists all US newspapers from 1690 to the present day. This catalogue was originally produced by the United States Newspaper Program, which ceased operation in 2011.

Consulted Libraries

Digitisation began in 2005 with grants to: University of California, Riverside; University of Florida Libraries, Gainesville; University of Kentucky Libraries, Lexington; the New York Public Library, New York City; the University of Utah, Salt Lake City; and the Library of Virginia, Richmond. This initial round of funding, covering newspapers from 1900 to 1910, concluded in 2007. The programme has since continued to allocate grants to institutions to digitise state collections. As of December 2018, this has included the University of Alabama, Tuscaloosa; Alaska State Library Historical Collections; Arizona Department of Libraries, Archives, and Public Records; Arkansas State Archive; University of California, Riverside; History Colorado; Connecticut State Library; University of Delaware; University of Florida, Gainesville; Digital Library of Georgia (University of Georgia Libraries/GALILEO); University of Hawaii at Manoa; Idaho State Historical Society; University of Illinois, Urbana; Indiana State Library; State Historical Society of Iowa; Kansas State Historical Society; University of Kentucky, Lexington; Louisiana State University; Maine State Library; University of Maryland, College Park; Central Michigan University; Minnesota Historical Society; Mississippi Department of Archives and History; State Historical Society of Missouri; Montana Historical Society; University of Nebraska-Lincoln Libraries; University of Nevada, Las Vegas; Rutgers University Libraries; New Jersey State Archives and New Jersey State Library; University of New Mexico; New York Public Library, Astor; Lenox and Tilden Foundation; University of North Carolina, Chapel Hill; State Historical Society of North Dakota; Ohio History Connection; Oklahoma Historical Society; University of Oregon; Penn State University Libraries, University Park; University of Puerto Rico, Rio Piedras; University of South Carolina; South Dakota Department of Education; University of Tennessee; University of North Texas; University of Utah, Marriott Libraries; University of Vermont; Library of Virginia; Washington State Library; West Virginia University Libraries; and Wisconsin Historical Society.

Microfilming Projects

The digitisation work of the National Digital Newspaper Program is built upon earlier preservation programmes managed by the United States Newspaper Program, which worked from 1982–2011 to identify, describe, and preserve historical newspaper collections. These programmes, funded by the National Endowment for the Humanities and given technical support by the Library of Congress, supported the preservation of historical newspapers through microfilming rather than the retention and conservation of loose or bound copies. Funding guidelines encouraged the removal of newspapers from bound volumes to facilitate a speedier microfilming process; however, this method largely prevented the rebinding and conservation of the original newspapers. This preference for microfilm as the "object of record" has continued under the National Digital Newspaper Program, as its technical and funding guidelines instruct awardees to scan existing microfilm copies, with only a brief mention made of scanning of original copies in order to complete a microfilm collection.

Digitisation Projects

The National Digital Newspaper Program builds on the work of the United States Newspaper Programme by running a biannual, competitive grant programme for institutions to digitise approximately 100,000 newspaper pages representing their state, with the option of applying for a second or third grant in subsequent rounds. The programme provides awardees with technical guidelines on selection, digitisation, encoding and delivery to ensure consistency across institutions and grant cycles but allows institutions to employ local expertise in fulfilling these guidelines, particularly regarding selection and populating bibliographic metadata. The programme originally limited the date range of submission to 1836–1922, but since July 2016 has allowed digitisation of newspapers from 1690–1963.

Selection

The guidelines for the National Digital Newspaper Program highlight four primary intellectual considerations for selecting titles to include in the Chronicling America database. First, the title should reflect the political, economic and cultural history of the state or territory, with preference given to titles recognised as "papers of record". Second, they should provide state, or multi-county, coverage of the majority of the state or territory's population. Third, titles with longer chronological runs are preferred over those with short or sporadic runs. Finally, particular consideration is given to titles that have ceased publication and therefore are less likely to be digitised by other providers. Technical guidelines state that a title should have a complete, or largely complete, run available on microfilm, and that use of the microfilm should not be restricted in any way that would affect the newspaper's inclusion in the Chronicling America database. Additional guidelines for microfilm quality and reduction ratio are also provided, noting that titles with higher-quality microfilm should be given preference.

Preservation and Access

Chronicling America's primary aim is to enhance free, public access to historical newspapers. Although archival-quality TIFF images are created for all NEH-funded digitisation projects, microfilm remains the preferred method of permanent preservation.

Composition of the Collection

Selection Available

As of December 2018, the collection included 2,689 distinct historical newspaper titles comprising 14,181,901 historical newspaper pages. A full list of included titles can be found at http://chroniclingamerica.loc.gov/ newspapers.txt. The collection contains issues from the years 1789–1963, but the bulk of the collection is from 1850 to 1922. This latter date is commonly understood as the US copyright boundary. Of the 2,689 tiles available, only eight include issues before 1800 and only forty-two include issues after 1922.

Data Quality

Text

Because Chronicling America is a distributed digital content creation programme, individual awardees are responsible for selecting content, evaluating microfilm, assigning metadata and writing descriptive newspaper histories for each title, the latter hosted directly by the Chronicling America website. Despite sharing common technical guidelines, this had led to the quality and character of the METS/ALTO data files varying considerably by state awardee, date of digitisation, quality of the source material, and the title itself. In some cases, OCR quality varies widely even within a given title. This may be the result of shifts in the original quality of the source material, which changed rapidly over the course of the nineteenth century, or the paper's conservation status. In addition, variation in OCR quality can represent changes made to OCR software over the course of a long-term digitisation project. The quality of the OCR data is so variable that any summary would be largely inaccurate, even for a specific title.

Images

For each page in their submission, contributors to Chronicling America are required to include an 8-bit, grayscale, 300–400 PPI TIFF preservation-quality image, as well as a derivate JPEG2000 (lossless 8:1 compression) and PDF file. These derivate images are made available through the Chronicling America web interface, while the TIFF can be provided to researchers by request. As JPEG2000 is a proprietary format, the National Digital Newspaper Program does not currently consider it a suitable archival substitute for TIFF, despite its higher compression allowing for more efficient online distribution.

Metadata Schema

The data contained within Chronicling America adheres to four specific metadata encoding schema. The METS XML schema is used for structural metadata, ALTO XML for the OCR content, PREMIS10 and MIX11 for technical metadata, and MARC and MODS12 for descriptive and bibliographic metadata. The descriptive and bibliographic metadata is largely based on title-level MARC records, many of which were created by the United State Newspaper Program. The technical guidelines for awardees direct them to map specific MARC fields to the required and optional metadata components. Other descriptive and administrative metadata is populated by awardees when collating or evaluating the microfilm and may include page numbers, section and edition labels, and preservation metadata. Since 2011–2013, many of the technical metadata fields relating the digitisation process have been reclassified as recommended, rather than required, components of the data files.

Backend Structure

The data for each issue is stored in multiple image and text files within a alphanumerical directory structure (https://chroniclingamerica.loc.gov/data/batches/), organised by unique batch identifiers. Each batch contains several unique titles, listed by their canonical Library of Congress Call Number. Within each title subdirectory, files are separated into subdirectories by issue date. Within these, there are two METS-encoded files providing technical, administrative and descriptive metadata for the issue. There is also a separate image (.jp2/.pdf) and ALTO-encoded XML files for each page within that issue. The METS files are named by the date for the issue and the ALTO files are numbered sequentially across the title within that batch. Archival TIFF files are stored offline. Thus, data for each issue can be obtained using the following standardised URL:

https://chroniclingamerica.loc.gov/data/batches/[Batch Name]/data/[Library of Congress Title ID/[Processing ID]/[YYYMMDDEE]/[File ID].xml

The National Digital Newspaper Program decided to serve image and OCR data at the page rather than article level for the sake of efficiency, but allows users to zoom, pan and crop images—in essence, manually zoning the text. Many of the collections are hosted by individual awardees as well as the Library of Congress, and some awardees have undertaken and provide a higher resolution of zoning and descriptive metadata through their own hosting venues, going beyond the minimum zoning of individual columns with appropriate coordinate information to facilitate text highlighting within the images.

User Interface Structure

Web Interface

The online user interface is an open-source Django installation called chronam that allows user to perform a simple or advanced search of the underlying descriptive and OCR data, or to browse images by title and date. The advanced search allows for filtering by state, title, years, page number, and language, and employs checkbox Boolean operators such as "any", "all", "exact" and "near". Tiled images with highlighted search results are displayed in an image viewer with an attached citation. The viewer allows users to pan and zoom as well as navigate through the issue. The underlying data (plain text, PDF and JP2) and manually selected snippets can be downloaded using icons at the top of the viewer. The underlying code for chronam is available on GitHub.

API

Materials from Chronicling America can be obtained through the site's API which, as of December 2018, does not require a unique access key. Datasets filtered by metadata or content and can be retrieved in HTML, Atom/XML or JSON formats through URL queries. More information and examples can be found at https://chroniclingamerica.loc.gov/about/api.

Direct Download or Drives

Bulk data from the collection can be obtained through web crawling tools, using the standardised file structure. This is aided using Atom and JSON feeds to detail the structure of the data and indicate when it is updated.

Rights and Usage

Web Interface

All material obtained from the Chronicling America web interface may be used freely for personal research. When browsed or searched through the user interface, users are presented with a full citation for the digitised image and text.

API and Direct Download

Users are allowed to access archive (.tar.bz2) files of all texts, images and metadata on Chronicling America for text and data mining. Those using data via the API are requested to use the URL and a website citation, such as "from the Library of Congress, Chronicling America: Historic American Newspapers site".

Re-Publication

The Library of Congress believes that the newspapers in Chronicling America are in the public domain or have no known copyright restrictions. Newspapers published in the United States prior to 1923 are in the public domain in their entirety. Any newspapers in Chronicling America that were published after 1922 are also believed to be in the public domain but may contain some copyrighted third-party materials and should be independently cleared for derivative use.

Delpher

Koninklijke Bibliotheek, Netherlands

History of the Collection

Delpher (http://www.delpher.nl) is the free, online repository of digitised printed material from the Netherlands. It was created and is maintained by the Koninklijke Bibliotheek, the national library of the Netherlands. It was officially launched in 2013, bringing together several previous digitisation projects, and now includes more than 15 million newspaper pages, 7.3 million magazine pages and approximately 900,000 books from the fifteenth to the twenty-first century. The digitised newspaper collection, originally funded by a subsidy of €12m from the National Programme for Investments in Large-Scale Research Facilities, has the explicit aim of being a resource for humanities researchers. The library has worked with academics and journalists throughout the process of selecting and obtaining the most representative and relevant collection possible.

Consulted Libraries

Delpher is the product of a collaboration between the university libraries of Amsterdam, Groningen, Leiden, and Utrecht, the Meertens Institute, and the Koninklijke Bibliotheek (The Hague). Newspapers are included from the collections of various organisations from the Netherlands and abroad, including: the Eemland archive; Arnhem Library; Calvin College Archives, Michigan; Gelders Archive; the Municipal Archive of Hulst; Municipal Sluis; Groningen Archives; Herzog August Bibliothek, Wolfenbüttel; Historical Center Overijssel; Joint Archives of Holland, Michigan; Royal Tropical Institute; Royal Institute for Language, Land and Ethnology; Royal Library; Kungliga Biblioteket, Stockholm; L'Archivio Segreto Vaticano; Museum Enschede; Meermanno Museum; National Archives Suriname, Paramaribo; Niedersächsiches Landesarchiv – Staatsarchiv, Oldenburg; NIOD Institute for War, Holocaust and Genocide Studies; Northwestern College, Orange City; North Holland Archives; Press Museum (now part of Sound and Vision); the private collection of André de Rijck; Radboud University Nijmegen; Regional Archive Alkmaar; Regional Archives Leiden; Roosevelt Institute for American Studies; Russian State Archive of Ancient Acts, Moscow; Rutgers University Library, New Brunswick; the Reformed Political Party office; the Social Historic Center for Limburg; City Archives of 's-Hertogenbosch; City Archives of Rotterdam; City Archive of Vlaardingen; Maastricht City Library; the National Archives Kew, Richmond; Tresoar – Frisian Historical and Literary Center; Trinity Christian College, Palos Heights; Ghent University Library; University Library of Groningen; Leiden University Library; Tilburg University Library; University Library Amsterdam; VU Amsterdam University Library; Waterlands Archive; West Frisian Archive; Wisconsin Historical Society, Madison; Zeeland Archives; ZB Planning Bureau and Library of Zeeland; Central Library, Zurich.

For those newspapers still in copyright, the following rights holders have contributed to the newspaper collection: AD News Media; Audax Publishing; Erven A. J. Morpurgo; Erven D. G. A. Findlay; Erven J. A. Pengel; Erven Varekamp; Erven Wormser; FD Mediagroep; Friesch Dagblad; HDC Media; Media Group Limburg; NDC Mediagroep; Nederlands Dagblad; NRC Media; Omroepvereniging VPRO; De Persgroep Netherland; Ried fan de Fryske Bewegig; Reformed Political Party; Foundation for the Management of the CPN Archives; Nieuw Israelitisch Weekblad Foudation; Utjouwerij Frysk en Frij Foundation; Trouw; Amigoe publishing house; De Telegraaf publishing house; Dhr. W. Lionarons; Vereniging Algemene Omroepvereniging AVRO; Wegener Nieuwsmedia.

Microfilming Projects

Since the 1970s, the Koninklijke Bibliotheek and other institutions have engaged in microfilm preservation of their newspaper collections. These efforts, however, were undertaken with differing specifications and not all the collections have been found suitable for digitisation, usually owing to the high contrast employed in some filming processes. Therefore, while it was generally considered preferable to digitise from microfilm collections as doing so was more efficient and cost-effective than scanning originals, the microfilm status of a title was not a primary selection criterion for digitisation.

Digitisation Projects

Digitisation of newspapers began at the Koninklijke Bibliotheek in 1999 as part of the Roaring Twenties and War & Revolution projects, in which three national newspapers from the 1920s and 300,000 pages from the 1910s were digitised and underwent OCR. Building upon these pilots, as well as a number of digitisation projects at other Dutch archives and libraries, full-scale digitisation began in 2007 with the Database of Digital Daily Newspapers or Databank Digitale Dagbladen (DDD) project. In 2013, this database was combined with the library's other digital collections to be made available through the Delpher interface. Digitisation of newspapers, alongside other heritage materials, now continues as part of the general operations of the library and is overseen by Metamorfoze, the Dutch national programme for

the preservation of paper heritage. All digitisation companies working for Metamorfoze were selected by means of a European tender procedure and all files are checked for completeness and correctness at the Koninklijke Bibliotheek.

Selection

At the start of the DDD project, there was no complete catalogue of the more than 5,000 national, colonial, regional and local newspapers that had been published in the Netherlands since the seventeenth century. Although the Koninklijke Bibliotheek owns the largest collection of Dutch newspapers, several titles exist only in libraries and archives elsewhere in the Netherlands or in other European nations. The selection process of the DDD project started with the installation of an advisory board, consisting of press historians and journalists, charged with selecting a number of major national newspapers as well as influential and long-standing colonial, regional and local publications. As the digitisation of these newspapers was primarily seen as serving academic researchers, the advisory commission was comprised largely of Dutch academics and journalists who were explicitly asked to consider the scientific significance or press history relevance of the proposed selection, as well as geographical, political and religious significance. The committee first defined what publications qualify as a newspaper: a product of a printing press, thus having been made into multiple identical copies, published at a set periodicity and day, high amount of content relating to current affairs of all types, and available for purchase by the general public. Then chronological periods were defined, based on important developments in the history of the press: 1618–1800, 1800–1814, 1814–1869, 1869–1914, 1914–1965 and 1965–1995. For each period a set of criteria was developed that was used to select a set of important, trend-setting and representative newspapers for each period. The aim of this was to have a representative set of newspapers digitised from the beginning of the project. As more and more newspapers were digitised over the following years, this initial selection became less relevant.

Preservation and Access

The Koninklijke Bibliotheek considers the digitisation of both microfilm and original newspapers as part of its preservation strategy. It has discontinued its microfilming programmes and is instead creating a collection of high-resolution JPEG files as its archival objects of record.

Composition of the Collection

Selection Available

As of December 2018, the database contains over 1.4 million newspaper issues representing over 15 million newspaper pages. A full list of included titles can be found at https://www.kb.nl/sites/default/files/ docs/Beschikbare_kranten_alfabetisch.pdf. The collection includes newspapers from 1618 to 1995, as allowed by copyright restrictions, and efforts have been made to represent the entire chronology fully. There is, however, a disproportionately large number of issues relating to the Second World War owing to targeted digitisation of this period through exceptional government funding.

Data Quality

Text

Newspapers within Delpher have been scanned for OCR using ABBYY versions 7.0 to 10.0. According to a 2018 study by the library, the newspaper collections, excluding those from the seventeenth century, have an average word-error rate of 11.3% (standard deviation: 9.96). This represents a material improvement over the databank's earlier OCR transcriptions, with newspapers from the Roaring Twenties and War & Revolution projects having error rates of roughly 30%. The study indicated that updating the current OCR by re-scanning archival masters with ABBYY 11 will improve this to 9.01%. The library is currently investigating this option. Owing to the Gothic font employed by seventeenth-century newspapers, the OCR on these items is particularly error-prone. To correct this, the library worked in a close collaboration with the Meertens Institute and with a group of skilled volunteers to manually re-key these transcriptions.

Images

For each page in the collection, Delpher maintains a JPEG2000 (lossless 8:1 compression) and PDF file, the latter of which is made available through a structured URL based on an item's unique identifier. Lower resolution images (96 PPI) are available through the general web interface and API.

Metadata Schema

Newspaper data is divided into three file types: an MPEG21 XML file describing the issue, an ALTO XML file for each page, and an OCR XML file for each article. This data adheres to two primary metadata schemas: structural metadata using MPEG21-DIDL and descriptive metadata using Dublin Core and

derivatives thereof. The MPEG21 file describes the structural hierarchy of the newspaper issue and provides descriptive metadata for the issue. Its constituent pages and individual articles include segment coordinates and stable URLs. Because a page may contain multiple articles and an article may appear on multiple pages, these are listed on the same hierarchical level with sub-elements describing their relationships. The choice was made to segment issues at article level in order to facilitate user searches, improve relevance rankings, and allow for the removal of an article for copyright reasons without having to remove an entire page or issue.

Backend Structure

Each issue within Delpher's newspaper collection is comprised of several items: the issue itself, each individual page and each unique article. Each item is, in turn, comprised of multiple components, including metadata, text, and image data, which are stored in individual or nested resource files. The issue MPEG21 file contains the IDs for all other items, components, and resources within that issue. Of these, only the archival master images and complementary technical metadata are not accessible through API or direct download.

User Interface Structures

Web Interface

The current user interface, Delpher, was built by an in-house team at the Koninklijke Bibliotheek. It replaced an earlier interface designed specifically to serve the DDD newspaper project. It is currently only available in Dutch. The interface allows users to perform a simple or advanced search of the underlying descriptive metadata and OCR text, or to browse images by date and title. The advanced search allows for filtering by article type, newspaper type, title, years, and place of publication. The full-text search can be filtered using standard Boolean operators. By default, search results are ranked by relevance but can also be ordered by date, article title or newspaper title. Users can also visualise their results as a date-segmented histogram. Once a result is selected, a full-page image, centred on the relevant article with highlighted search results, is displayed in an image viewer. The viewer allows users to pan and zoom as well as navigate through the issue. The underlying data (plain text, PDF and JPEG) and manually selected snippets can be downloaded using icons at the top of the viewer, and the metadata and OCR text can be viewed in retractable widgets to either side of the image.

API

Delpher supports a range of API queries. It can harvest complete sets of metadata through the OAI-PMH protocol or sub-sets through a Java implementation of the SRU search protocol. Search queries are made via a structured URL at either issue or article level.

Direct Download

The Delpher Open Newspaper Archive contains the texts (OCR, ALTO, XML) of all newspapers from the period 1618 to 1876. The archive is 111 GB in size and divided into 22 ZIP files, available at http://www.delpher.nl/data. For copyright reasons, the archive does not include newspapers after 1876 but, under certain research conditions, a licence may be granted for bulk use of post-1876 dates.

Rights and Usage

Web Interface

All material obtained from the Delpher web interface may be used freely for personal research. When browsed or searched through the user interface, users are presented with a full citation to the digitised image and text. Any materials that remains under third-party copyright are clearly labelled, including the specific conditions of use for that item.

API and Direct Download

Users are allowed to access archive (ZIP) files of all out-of-copyright texts and metadata on Delpher for text and data mining. Advance permission is required to access datasets that contain copyright-protected materials. Although individual texts have been released into the public domain, the dataset as a single object has been released under a CC-BY license and must be properly attributed in derivative works.

Re-Publication

Out-of-copyright text and images have been released into the public domain and may be used or republished for both personal and commercial purposes. Items that remain under third-party copyright may not be redistributed or republished, in print or digitally. They may, however, be linked to directly.

Europeana Newspapers

Europeana Foundation, Netherlands

History of the Collection

Europeana (https://www.europeana.eu) is a collection of collections, drawing from more than 2,000 cultural heritage partners all throughout Europe. The full Europeana platform contains over 50 million artworks, artefacts, books, films and music from European museums, galleries, libraries and archives.

A thematic newspapers collection within Europeana was established in July 2019. It mainly builds on the 2012–2015 Europeana Newspapers project (http://www.europeana-newspapers.eu) that was funded by the European Commission's ICT Policy Support Programme, as part of the Competitiveness and Innovation framework Programme (CIP). The project received an initial €5.2 million with a maximum European Commission contribution of €4.1 million. Europeana Newspapers aggregated metadata for more than 20 million digitised newspaper pages from a variety of European libraries, collected around 12 million page images of digitised newspapers and processed them with OCR and OLR, and developed a prototype portal for full-text search in digitised newspapers at The European Library (TEL). Following the closing of the TEL service at the end of 2016, the newspaper collection was migrated to a "thematic collection" subsite of the main Europeana platform.

Consulted Libraries

The following full Europeana Newspapers project partners have typically contributed a public domain subset of their digitised newspaper collections to the project: National Library of Austria; National Library of France; National Library of Netherlands; Dr. Friedrich Teßmann State Library; National Library of Estonia; National Library of Latvia; National Library of Finland; State and University Library of Hamburg; Berlin State Library; University Library of Belgrade; National Library of Poland; and National Library of Turkey. Furthermore, these associate partners have shared some data, but generally only metadata: National and University Library of Slovenia; National and University Library of Iceland; Royal Library of Belgium; National and University Library of Zagreb; St. Cyril and Methodius National Library of Bulgaria; National Library of Romania; National Library of Luxembourg; National Library of the Czech Republic; National Library of Spain; National Library of Portugal; National Library of Wales; National Library of Serbia; and National and University Library of Slovakia.

Microfilming Projects

Europeana draws from existing databases, so it does not undertake any microfilming or material preservation but relies solely on the capture processes of its project partners, some of which have individual entries within this *Atlas*.

Digitisation Projects

Europeana draws from existing databases, so it does not undertake any digitisation but relies on contributions of digitised materials from gallery, library, archive and museum (GLAM) institutions.

Selection

The main selection criteria for the aggregation of digitised newspapers to Europeana is the conformance with the Europeana Publishing Framework (https://pro.europeana.eu/post/publishing-framework), which details several tiers of quality requirements and recommendations for metadata, text and image content. For inclusion within the thematic collection, the most general requirement is that the data provider populates the dc:type field with one of the following terms: Newspaper Title, Newspaper Issue, Analytic Serial, Newspaper, Journal, or Printed Serial.

Preservation and Access

Since Europeana is an aggregator platform for digitised cultural heritage, it does not undertake any preservation activities itself, but does act as an additional access site for included digitised newspapers.

Composition of the Collection

Selection Available

Europeana Newspapers is a collection of collections. It assembles digitised historical newspaper content from more than twenty distinct data providers, each with their own data parameters and histories. In total, the Europeana Newspapers collection includes 28,816,750 pages in forty different languages. Because of the multiple origins of the data and the migration process, it now holds data in varying degrees of refinement–while roughly 12 million pages of the originally aggregated 20 million pages have machine-readable full text from OCR, another approximately 8 million pages are only available in Europeana through their metadata. Furthermore, a wide variety of digitised newspaper data that already existed on Europeana prior to the migration of Europeana Newspapers from TEL to the main

Europeana platform have since also been merged into the thematic collection.

Data Quality

Text

The overall OCR quality of the Europeana Newspapers collection is approximately 70-85% but it varies widely; a full report from 2015 is available via the Europeana Newspapers project website. The majority of the Europeana newspapers were scanned from microfilm. A performance evaluation has shown that this only slightly impacts OCR quality, with less than 1% loss of accuracy, while it enables an efficient digitisation of large-scale volumes.

Images

Images from project partners were provided to Europeana Newspapers in various formats. Since Europeana Newspapers conducted the OCR and OLR for these images, images were only used if they were considered to be of archival quality. For the purposes of OCR/OLR processing, all images were converted to TIFF. In addition, JPEG2000 versions were created for all images to enable display and zooming on the web.

Metadata Schema

All Europeana Newspapers project data and metadata follows the ENMAP profile, a dedicated metadata profile for digital newspapers based on established community standards METS and ALTO. Full details of the encoding schema (v 1.0, 2015) are available on the Europeana Newspapers project website. For aggregation into Europeana, a special profile for representing full text alongside the Europeana Data Model was created (https://docs.google.com/document/d/1t5yGEzQ0KV2rqU0sFDoKnI2bIDBGrmj0f1gSOCRUgJ4/). It is also aligned with the IIIF standard (https://docs.google.com/document/d/1vhQstotXm4b-t8FHCzStHNCoz1dVzGFsaXLrn2vCPVI/). Europeana metadata has also been released in EDM in XML, EDM in JSON-LD, Dublin Core in XML, and CLARIN CMDI in XML.

User Interface Structure

Web Interface

In Europeana, users can search for terms across the full text or the newspaper title. It is also possible to filter by collection, media, rights availability, providing country, language, aggregator and institution. Users can browse content through date, title, or a map. The search results display the issue-level metadata of the content-providing libraries, with varying degrees of information: full text and newspaper page image are displayed, with links to view the original in the source library, if available.

API

Europeana provides several APIs to use with the newspaper collection:

- Newspapers Search API, which allows searching of the full text and metadata of the newspapers
- IIIF Manifest API, which allows access to IIIF manifests
- IIIF Image API, which allows access to newspaper images
- IIIF Fulltext API, which allows access to the full text of the newspapers

Documentation and examples of use are available at https://pro.europeana.eu/resources/apis.

Direct Download or Drives

Newspaper metadata and full text sets per data provider are available for direct download from https://pro.europeana.eu/resources/apis/iiif#download. Users can also directly download individual digitised newspaper pages via the Europeana user interface.

Rights and Usage

Web Interface

All titles are freely searchable through Europeana and each individual object contains information about its terms of use, following https://rightsstatements.org/en. However, full access to recent twentieth-century content is problematic owing to copyright. Europeana Newspapers has developed a "Roadmap for Improving Access to Newspapers" to urge policy makers to consider more open licences for digitised newspapers.

API

Item-specific information about terms of use is also available via the API.

Re-Publication

Most of the newspapers assembled by Europeana Newspapers are dedicated to the public domain with all metadata being released under a CC0 license and may therefore be re-published freely. See individual terms of use for further details.

Hemeroteca Nacional Digital de México

Universidad Nacional Autónoma de México

Universidad Nacional Autónoma de México

HEMEROTECA NACIONAL DIGITAL DE MÉXICO

History of the Collection

The Hemeroteca Nacional Digital de México is the online repository of digitised printed material from Mexico. The digital collections were derived from materials belonging to the Fondo Reservado (special collections) of the Hemeroteca Nacional (National Newspaper Library). The Hemeroteca Nacional and the Biblioteca Nacional (Mexico's National Library) are under the stewardship of the Universidad Nacional Autónoma de México (UNAM), Mexico's national university, and both the physical and digital collections are housed and maintained on campus. Planning for the HNDM began in 2000. The project focused on digitising the National Newspaper Library's microfilm collection, which was created in 1960. Work began in earnest in 2002 and the first version was made available that same year, with restricted access. Since then, the HNDM has undergone several system migrations and new designs, the latest one being in 2015.

Consulted Libraries

The HNDM includes the newspaper collections belonging to the Fondo Reservado (special collections) of the Hemeroteca Nacional.

Microfilming Projects

Microfilming began at the Hemeroteca Nacional in the 1960s, when the National Newspaper Library was moved from downtown Mexico City to the new, purpose-built facility on the UNAM campus in the south of the city. The microfilm collection has two copies for each newspaper, one in silver halide for preservation purposes and the other in diazo film for public lending. Currently there are approximately 28 million microfilmed pages.

Digitisation Projects

Although plans to digitise the library's microfilm collection began in 2000, finding a company in Mexico City that had the equipment and the experience to do the work proved more complicated than expected. Eventually, a company based in another Mexican city began work in 2002. The same company undertook the OCR as well as the digital platform and hosting. In later years, the system was updated by UNAM and the system migrated to UNAM servers.

Selection

The project was designed to include digitisation of the complete microfilm collection. This was founded on the idea that the materials on microfilm had already undergone a selection process which was based on the following criteria: the newspaper's degree of conservation, demand, and historical, heritage, journalistic, and documentary importance. One of the main problems in the early stages was that there was not a reliable catalogue of the microfilm collection that included the technical micrographic characteristics for the physical artifacts as well as their state of conservation and the metadata about the newspapers themselves. Therefore, the quality of the microfilm varies leading to variation in the quality of the digitised images and OCR.

Preservation and Access

The library continues to use microfilm as its main preservation strategy for the newspaper collection. In recent years the library has turned its attention to the digital preservation of the approximately 8 million original digitised TIFF images from which the HNDM collection is created.

Composition of the Collection

Selection Available

Currently, the HNDM has almost 8 million digitised pages from 941 newspaper titles between 1722 and 2006. Out of these, 516 titles (1722–1889) are open access, and 425 (1890–2006) are restricted and can only be accessed from inside the newspaper library's facilities and on campus. The microfilmed and digital archives can be found in the National Newspaper Library's Fondo Reservado (special collections), which includes Mexican periodical publications between 1722 and 1917, foreign periodical publications from 1665 to 1920, the Garcia Valseca collection, and the newspaper miscellanea and important newspapers microfilms collections. Newspapers from the long nineteenth century, 1800 to 1914, contain 771 titles and 2,069,247 pages.

Data Quality

Text

According to an analysis by the Dirección General de Cómputo y de Tecnologías de Información y Comunicación at UNAM (the National University's Computing and Communication and Information Technologies General Office), the OCR average success rate is 54%. It was found that 72% of the images have good readability, about 15% regular, and 13% poor. The image quality was 56% good, 34% regular, and less than 10% poor.

Image

The HNDM's digitisation was obtained from 35mm microfilm rolls that the National Newspaper Library created. Publication stills have been reduced relative to their original size and contain between 140- and 180-line pairs per millimeter (pl/mm). This reduces the impact of the deterioration of paper and ink tones from the original documents over time and produces a better contrast. To create the HNDM collection, every page of every item from the newspaper was digitised as a bitonal TIFF image.

Backend Structure

The HDNM backend structure has been modified several times since its creation in 2002. The current platform was developed in-house in a collaboration between the UNAM's central computing services and the library's computing service department. This collaboration was launched in 2011. It is held in a MongoDB v.3.6 database and Solr for indexing. This database is in JSON format, with metadata descriptions of the collection, publication, item, page, and publication; it does not include the actual OCR content. Work on the platform is ongoing and new modules are being developed, particularly ones that allow the library to administer the digital collection, such as uploading new images, and adding or modifying new metadata.

The dataset's XML archive, containing the OCR of the digitised images, works with three metadata schemes (Dublin Core, custom OCR/XML data, and METS). In the OCRed documents, the METS metadata schema includes, at the beginning of the file, a publication description in Dublin Core with date, title, city, state, country, category, collection and language. It is followed by an OCR structure that divides page content into columns, regions, paragraphs, lines, and words with an associated coordinate, allowing the search terms to be highlighted by users.

User Interface Structure

The current user interface was designed in-house and released in 2015.

Search

The interface allows users to perform simple searches by title, year or place, or advanced searches which combine the previous options together with language, type of access and frequency of publication. The interface also includes indexes which allow users to restrict their searches by country, state, city, type of access, language, frequency of publication, title and dates. Users can also search by word, which uses the system's underlying OCR files, and the results highlight the search terms in the PDF. The interface also allows users to access detailed bibliographic descriptions of the newspaper titles.

API

There is currently no API available for the HNDM.

Direct Download or Drives

The HNDM does not allow bulk downloads or access to any of its files, neither the images nor the OCR. Newspaper titles can be printed or downloaded one at a time as PDFs. However, in both cases the print version is either extremely large, showing only one part of the image, or extremely small, with the newspaper page occupying just a fraction of the PDF page. This renders the print option unviable. Users who would like printouts must go to the HNDM office located in the library.

Rights and Usage

Materials in the HNDM which are considered under Mexican law to be in the public domain are freely available online. The library worked in close collaboration with the UNAM's legal team, but it was difficult to determine what is uncontestably within the public domain. The UNAM made the decision to be cautious and so the cut off year was set as 1889. This decision was made a few years ago and should, logically, be updated each year, but to our knowledge this has not been done yet. Newspapers from 1890 onwards are restricted access and can only be consulted from within the library or the university campus. These are clearly indicated in the interface with a padlock icon. The HNDM uses the term "open access" for the newspapers that are freely available, although there are restrictions. Materials can only be used for personal or educational purposes and due credit must be given to the HNDM.

Web Interface

The complete collection can be searched and, in the results, metatextual information is presented with a thumbnail image of the newspaper. Closed access newspapers are indicated with an icon and the PDFs cannot be accessed.

Re-Publication

Direct access to the OCR and image files is not available. Open access PDFs cannot can be printed. Digital content cannot be incorporated into any digital platform or application without prior authorisation by the HNDM.

Papers Past

National Library of New Zealand

History of the Collection

Papers Past (https://paperspast.natlib.govt.nz) began in 2000 with a project to provide access to nineteenth-century New Zealand newspapers and periodicals. The initial aim was to make available 250,000 pages within a year. The website was launched in 2001 with 300,000 digitised pages, which were viewable and printable but not searchable, being page images only. In 2005, the library ran a pilot project using OCR to generate full text and to make the newspapers searchable. The website was then relaunched in 2007 with a new interface to reflect this new search functionality. In 2016, Papers Past was relaunched for a second time, with the addition of three more sections—magazines and journals, letters and diaries; and parliamentary papers–bringing together full-text content from several standalone websites, such as the magazine Te Ao Hou. This redevelopment also allowed for a mobile view.

Consulted Libraries

The Papers Past digital archive draws from the National Newspaper Collection, which contains newspapers from New Zealand, Australia, and the Pacific, part of the Alexander Turnbull Library collections in Wellington. The National Newspaper Collection maintains these newspapers in various formats, including paper, microfilm and digital media; their microfilm collections have thus far been the primary source medium for digitisation.

Several smaller projects have used other libraries' collections in conjunction with those of the Turnbull. The early Auckland newspapers combined the holdings of the Auckland War Memorial Museum, the Auckland City Libraries and the Hocken Library, University of Otago with those of the Turnbull Library in order to get the most comprehensive coverage of three titles from the 1840s: *the Auckland Chronicle and New Zealand Colonist, Auckland Times*, and *New Zealand Herald* and *Auckland Gazette*. Likewise, the digitisation of Niupepa: the Māori Newspapers used digital files from the University of Waikato, which had been originally created from microfiche made from the Turnbull Library holdings. In some cases, if the Turnbull Library does not hold the master or intermediate microfilms, they will borrow another institution's intermediate reels and get a duplicate set made for digitisation, as was the case for the Samoan newspapers, which were digitised from film held by the State Library of New South Wales (Australia).

Microfilming Projects

The library has been microfilming newspapers since 1953. For the years 1953–1977, this was focused on regional and local papers, and the filming was to save space rather than for preservation; physical copies of the major metropolitan dailies were kept. Between 1977 and 1984, efforts shifted to microfilming older newspapers rather than the current titles. The programme was reviewed in 1984 and the revised goal was to microfilm all New Zealand newspapers (current and historic). Initially, this retrospective filming only covered issues up to 1940 but was this extended to 1977 in 1992. In 1990, the microfilming of community newspapers was stopped; filming continues today, with a focus on the major contemporary newspapers.

Currently, microfilming is undertaken by an external vendor and the master reels are maintained for the long-term preservation of the collections and as an intermediary step prior to digitisation. However, microfilming for a number of current newspapers is still carried out by regional libraries, who also carry out some retrospective newspaper microfilming and contribute hard-copy holdings to microfilming projects organised by the National Library. The National Library also borrows physical newspapers from other institutions for filming if it does not hold a complete run.

Digitisation Projects

Newspaper digitisation at the National Library has three strands; the National Library programme itself, the collaborative programme and a partnership programme. Digitisation scanning was initially contracted to one vendor for the initial scanning and a second vendor to undertake the OCR conversion. Now, both scanning and conversion are done by a single vendor.

The Collaborative Digitisation Programme adds additional newspapers to Papers Past using the combined resources of community groups around New Zealand. Every year, the library invites applications for new additions, which are listed at https://natlib.govt.nz/librarians/national-library-services/collaborative-digitisation. The programme started in 2010–2011 and is focused on small, local newspapers. Local libraries and community groups can apply to have microfilmed newspapers digitised or to have them microfilmed and then digitised in a subsequent year. The library handles all the microfilming, digitisation and uploading to Papers Past. There is also the opportunity for larger

institutions to partner with the library, normally over a three-year period. This follows a similar model to the Collaborative Digitisation Programme but is over multiple years and is generally used for the digitisation of the regional dailies. With both programmes, the applicants contribute towards the direct costs of digitisation, generally 50% of the cost.

Selection

As New Zealand has a relatively short print history, the first newspaper being published in 1839, it was thought plausible to digitise an extensive range of the nation's historical newspaper collections for public access alongside continued microfilming for preservation purposes. In order to avoid copyright, the library initially only digitised nineteenth-century New Zealand newspapers, though this date range was later extended to titles published before 1920. Moreover, the initial choice of titles was largely based upon the recency of the microfilm processing, to ensure a high-quality transfer to digital media; as the library is now focused on filling geographical and temporal gaps in the digitised newspaper collection, the date of microfilming has become less significant as a criterion for digitisation.

Currently, the scope of the newspaper programme for Papers Past is newspapers published in New Zealand or the Pacific up to the end of 1950, using the Library of Congress definition of a newspaper. In 2015, as part of this expanded remit, the library added four titles from Samoa and the library aims to add more Pacific titles in the future, with a focus on the areas of the Pacific covered by the library's comprehensive collecting policy: American and Western Samoa, the Cook Islands, Fiji, Nauru, Niue, Pitcairn Island, Tokelau and Tonga.

The ongoing selection process tracks suggested titles and is focused on expanding geographic and temporal coverage and addressing user demand, considering also copyright status and the quality and availability of microfilm. It is managed by the library's Digitisation Team, with selections approved by a small, in-house committee consisting of the Digitisation Team, the Service Manager of Papers Past, the Curator of Newspapers and Serials, and the Microfilm Librarian. Previously, digitisation has been prioritised for particular events; for example, the centenary of the First World War led to the digitisation of materials from 1914–1918. Current priorities for selection are: geographical gap filling; additional material in te reo Māori; extending existing titles up to 1950; and additional Pacific material. A record of user requests is kept, and those requests are used to help prioritise titles for digitisation.

Preservation and Access

A separate record for the digitised title is created in the library catalogue, which contains a link to the title page on Papers Past. This page includes an essay about the history of the publication, using material from the library catalogue and Te Puna National Bibliographic Database, as well as copyright information and acknowledgements. The digitised copy acts as the primary access copy for the object with the microfilm acting as the long-term preservation medium.

Composition of the Collection

Selection Available

The National Library of New Zealand has now made available 5,789,376 newspaper pages from 847,719 separate issues. Of the 147 titles, eighty have issues digitised before 1880, with 118 having digitised issues between 1880 and 1920 and ninety-three having issues published after 1920. The majority of individual articles (roughly 80%) fall between 1880 and 1920; however, depending on one's definition of an article, the percentage of pre-1880s content may be slightly larger than this number reflects; automatic zoning of articles cannot always separate individual articles without headlines, a common practice in the mid-nineteenth century, and may therefore combine several items into a single digital object.

In 2015, the National Library added a collection of historical newspapers published primarily for a Māori audience between 1842 and 1935. This is based on the digital Niupepa Collection developed and made available in 2000 by the New Zealand Digital Library Project, at the Department of Computer Science, University of Waikato. The source material for this digital collection is Niupepa 1842–1933, a collection of newspapers published in Māori or for a Māori readership, microfilmed by the Alexander Turnbull Library (1988). That same year, four titles from Samoa were also added, namely the *Samoa Times* and *South Sea Advertiser* (1888–1896), the *Samoa Times* and *South Sea Gazette* (1877–1881), the *Samoa Weekly Herald* (1892–1900) and *Samoanische Zeitung* (1903–1930).

Data Quality

Text

There is currently no independent study of the OCR accuracy of the collection.

Image

The Papers Past newspapers were digitised from microfilm as 400 PPI bitonal images. Images were originally delivered to users as archival-quality TIFF files, to eliminate the need to reformat or host duplicate images. These were served through a proprietary viewer called Daeja ViewONE. This changed with the 2007 relaunch and the move to Greenstone (now Veridian). Images are now provided to users as embedded GIF files, derived from the Modified Master files, but are still stored as 400 PPI TIFF files for preservation reasons. Titles are also now scanned as bitonal or greyscale, depending on the quality of the original newspaper and the filming.

Users can request a copy of the 400 PPI Preservation Master via email, and this will be supplied as long as there is no conflict with third-party copyright. Users can also download a copy of a page as an image file or as a PDF. The image files are large GIF files, while the PDF pages are 200 PPI. GIFs appear, when downloaded, as having a resolution of 96 PPI, but they are physically large files, which would print at approximately 23x29 inches at 300 PPI.

Metadata Schema

The data contains several different metadata types: METS XML schema is used for structural metadata, ALTO XML for the OCR content, and MIX for technical metadata. The library supplies basic metadata to the vendor, including title, date range, and bibliographic ID, with the rest captured as part of the scanning and OCR process. The technical metadata embedded in the image files, such as make and model of scanner and software used, is harvested as part of the scanning process. Manual headline correction and manual illustration caption correction is also done as part of the digitisation process.

Backend Structure

The data for each issue is stored in multiple files within a directory structure that provides date and title information, as follows: TitleAcronym/YYYY/TitleAcronym_YYYYMMDD. Within this directory, there are four sub-directories: one containing preservation TIFFs; one containing modified master TIFFs, the METS file, and text data in the form of ALTO-encoded files for each page of content, numbered using a four-digit standard such as 0001.xml; one containing the IE METS file, needed for integration into the National Digital Heritage Archive; and one for page-level and issue-level PDFs.

User Interface Structure

Web Interface

The newspaper web interface for Papers Past allows users to perform a simple search of the underlying descriptive metadata and OCR text or to browse images by date, region and title. Users can filter their searches by date, title, region, or content type—as of December 2019 these include *article, advertisement* or *illustration*. The full-text search can be filtered using standard Boolean operators. Once a result is selected, an image of the article with highlighted search results is displayed in an image viewer. The viewer allows users to view the unhighlighted image or the OCR text. A breadcrumb navigation menu, such as:

Newspapers > Auckland Star > 9 November 1907 > Page 13 > This article

allows users to navigate the issues or browse the title by date. Bibliographic and copyright information is also provided.

API

With the DigitalNZ API, it is possible to access and use a sub-section of the Papers Past metadata, including titles, dates, and URL. Currently API data is limited to material digitised before 2013. An API key can be obtained by registering through the DigitalNZ website.

Direct Download or Drives

The National Library of New Zealand is currently developing processes for making out-of-copyright data available to users in bulk format.

Rights and Usage

Web Interface

Newspaper material on the Papers Past website has been provided in good faith for users by the National Library on the basis that the newspaper publications provided from the nineteenth and early twentieth century are out of copyright, that in most cases digitised copies replace microfilm versions previously provided to the public, and that for more recent newspaper publications, permission has been sought from the publisher to reproduce the material on this website. In many cases, the publisher has made the material available under a Creative Commons licence, most commonly CC-BY-NC-SA. Usage and copyright information for a title is available under "Using this item" next to the article.

API and Direct Download

Metadata available through the DigitalNZ API has been licensed for use by its owners, and API access has some restrictions, such as not sharing API keys and ensuring you identify the source. Papers Past data accessed through the DigitalNZ API is for non-commercial use only.

Re-Publication

When re-publishing material from the website that is out of copyright or for which you have gained permission from the copyright holder, the library requests that you acknowledge the National Library of New Zealand as the source of the information. If the material is republished online, they further request a link to where you found the information on this website.

Suomen Kansalliskirjaston digitoidut sanomalehdet

Suomen Kansalliskirjasto, Finland

History of the Collection

The Suomen Kansalliskirjaston digitoidut sanomalehdet has its origins in the Helsinki University Library Centre for Microfilming and Conservation, established in Mikkeli in 1990. Now known as the Centre for Preservation and Digitisation, part of the Suomen Kansalliskirjasto (National Library of Finland), the centre joined three other libraries in 1998 to form the Nordic digitisation project TIDEN. In 2001, the National Library launched its initial digital newspaper collection with 36,000 of an intended 90,000 pages of forty-four different Finnish titles published between 1771 and 1860. In 2005, the collection received 1.9 million page-requests and 160,000 unique visits. By 2018, the collection included all newspapers and journals published in Finland between 1771 and 1929 and comprised over 880,000 newspaper issues, containing 6.2 million pages of content.

Consulted Libraries

The digitisation of Finnish newspapers was undertaken as part of the Nordic Project TIDEN, comprising the Royal Library of Sweden; the National Library of Norway; the University Library of Aarhus (Denmark); and the Helsinki University Library. The digitised collections are generally based on microfilms held by the Suomen Kansalliskirjasto, with some digitised from physical objects and new issues received electronically.

Microfilming Projects

Newspapers had been the primary focus of reformatting programmes throughout Scandinavia since the 1950s and microfilming has been used as an access and storage format at the Suomen Kansalliskirjasto since 1951. Through the late 1980s, newspapers were stored on 35mm cellulose acetate film, first by a private service provider, Rekolid, and subsequently by the Helsinki University Photographic Institute. Since 1997, all Finnish newspapers have been microfilmed at the National Centre for Preservation and Digitisation. Until the 1970s, microfilm reels were stored in the same accommodation as print collections, after which they were transported to the Viikki bomb shelter. Since 1990, they have been held in air-conditioned vaults at Mikkeli.

As part of the Rescue Project, newspapers between 1771 and 1945 were re-filmed (either copied onto a more stable medium or re-filmed from originals), while newspapers between 1945 and the 1970s are currently under consideration for re-filming, owing to the difference in filming technology before and after 1980. All versions of the newspapers are currently retained, and this has allowed the creation of composite collections, wherein the original has been lost but a microfilm version remains, for digitisation. The completeness of the microfilm collection has allowed digitisation from microfilm, rather than originals, where quality is sufficiently high. In cases where quality was insufficient, originals were first microfilmed before being selected for digitisation. Currently, a greater number of filming projects include supplements as well as borrowed materials (in order to complete runs) than was previously the case.

Digitisation Projects

The initial work undertaken by the University of Helsinki as part of TIDEN was funded by the Nordic Council of Scientific Information (NORDINFO) and the Ministry of Education in Finland. These initial digitisation tests were important in defining best practice for future microfilm digitisation projects, and the findings were published alongside other recommendations by the International Federation of Library Associations and Institutions in its 2002 supplement on *Microfilming for Digitisation and Optical Character Recognition*. In particular, the project developed test criteria for the digitisation of microfilm and experimented with best practice in developing automated production workflows. The project received €10,000 to €40,000 in funding from the Nordic Council for Scientific Information, with additional funding from the Ministry of Education allocation to the Finnish National Library.

Selection

The digitised historical newspaper collection of the National Library of Finland is based upon the newspapers acquired through free deposit laws since the eighteenth century. At the time of TIDEN, the Finnish legislation defined a newspaper as "a printed product published at least once a week". Historical newspapers had been microfilmed systematically from the 1950s onwards, and the aim was to digitise the whole older collection step-by-step, using microfilm as an intermediary. The first Finnish newspaper was published in 1771, and the first collection to be digitised was from this year forward until about 1860. After the TIDEN project, the next stages covered the newspapers from 1861–1890, 1891–1900 and 1901–1910 according to the allocated funding. Digitisation work followed the alphabetical order of newspapers within the chosen timeframe.

Composition of the Collection

Selection Available

The digitised collection contains all Finnish newspapers held by the Suomen Kansalliskirjasto for the years 1771–1929, with later years digitised and made available through special agreements with copyright holders where possible; newer digitised newspapers are available at the six national deposit libraries. As of October 2019, the full collection included over 998 distinct newspaper titles, comprising 6,259,133 historical newspaper pages. The majority of these pages (4,031,018 pages, representing 64% of the collection) are currently available for public use, with a further 2,228,115 pages, (post-1929), held in restricted use. A full list of publicly accessible titles can be found using the filter on the newspaper web interface.

Following the general trend, the volume of newspaper publishing in Finland increased towards the turn of the century: when all issues from 1771 to 1910 are counted, 82.7% of the data is from 1890–1910, and 92.3% is from the last four decades, 1870–1910. The majority of the newspapers are in Finnish and Swedish, but there are some pages in Russian and German, and other languages. Different languages dominated the Finnish public sphere in different periods: more than 50% of the publications before the late 1880s were in Swedish, after which the share of Finnish language publications increased to over 75% by 1910. The Russian language publications emerged after 1900, while there were already German language publications during the 1820s and 1830s. Out of the total number of newspaper pages in the collection, 1,063,648 are in Finnish, 892,191 in Swedish, 8,997 in Russian, and 2,551 in German.

Data Quality

Text

The majority of nineteenth-century newspapers digitised by the Suomen Kansalliskirjasto were printed using Gothic (Fraktur, blackletter) typeface, with a minority of printed using Antiqua; the difficulty standard OCR software has recognising the former typeface is well known. By 2006, the Suomen Kansalliskirjasto had implemented automated encoding of word coordinates and grayscale scanning, utilising the digitising software DocWorks (CCS), with OCR by ABBYY Finereader, and structuring metadata in-house by combining OCR data with catalogue information. The next phase of development focused on automating OCR for both Fraktur and Roman on the same page, and conforming to international METS encoding standards. Analysis of parallel samples and word error rates showed that about 69% of all word tokens can be recognised with the modern Finnish morphological analyser, Omorfi. If orthographical variation is considered and the number of out-of-vocabulary words is estimated, the recognition rate increases to 74-75%. Overall the collection has a relatively good quality rating of about 69-75%; around 25-30% of the collections needs further processing in order to improve the overall quality of the data.

Images

The publicly available images from the collection are available as a PDF or JPEG file with a resolution of 300 PPI, the latter of which is made available through a structured URL based on an item's unique identifier. High resolution images are stored at the server of the National Library of Finland and released as part of METS packages in TIFF format at https://digi.kansalliskirjasto.fi/opendata.

Metadata Schema

The data hosted by the National Library of Finland uses the METS XML schema for structural metadata, ALTO XML for the OCR content, MIX11 for technical metadata, and MODS12 for descriptive and bibliographic metadata.

Backend Structure

The main database contains metadata, page data, and file data containing the archive directory information. The database offers page images of the content and access to the content of the pages in ALTO XML format. However, the URL structure is not easily translatable from bibliographic data; it places the text files within numerical directories representing individual bindings. For example:

https://digi.kansalliskirjasto.fi/sanomalehti/binding/566631/page-4.xml

In the data packages, pages are located in two separate directories: one based on ISSN and the other on publication year. Below the publication year in the data structure is the language of the publication, below which are the actual ALTO XML files. They are named descriptively by ISSN, year, date, issue and page. For example:

0355-8347_1881-07-01_13_012.xml

User Interface Structures

Web Interface

The web interface allows users to perform a simple or advanced search of the underlying descriptive metadata and OCR text. The advanced search allows for filtering by material type, title, collection, years, place of publication, author, keyword, publisher and language. The full-text search can be filtered using standard Boolean operators, a fuzzy search option, or by limiting to content or metadata fields. By default, search results are ranked by relevance but can also be ordered by date, title, author or date of inclusion in the collection. Once a result is selected, a full-page image with highlighted search results is displayed in an image viewer. The underlying data (plain text, PDF and JPEG) and manually selected snippets can be downloaded using icons at the left of the viewer, and the metadata and OCR text can be viewed in retractable widgets.

API

There is currently no API system in place for accessing the newspaper data, though bulk data from the collection can be obtained through web crawling tools, using the aforementioned standardised file structure.

Direct Download or Drives

There are currently several options for obtaining the newspaper data in bulk format. The Digital Collections maintain an Open Data website (https://digi.kansalliskirjasto.fi/opendata) in Finnish and English, allowing the download of both METS and OCR data as date- or language-delimited ZIP files. The years available vary and are not yet comprehensive of the entire collection. The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771–1874) was released in 2011. The data package is in the METS/ALTO format and downloadable via the Language Bank of Finland (http://urn.fi/urn:nbn:fi:lb-2015051201). The Newspaper and Periodical OCR Corpus of the National Library of Finland (1875–1920) was released in November 2017. The data package is downloadable via the Language Bank of Finland. The dataset includes all those newspapers and journals that had been digitised by the end of the year 2013. This includes all published newspapers 1875–1920.

Rights and Usage

Web Interface

All out-of-copyright material obtained from the Suomen Kansalliskirjaston digitoidut sanomalehdet web interface may be used freely but it is requested that they be cited using standard citation conventions. Any materials that remain under third-party copyright are clearly labelled and provide the specific conditions of use for that item; users may not redistribute in-copyright digitised material without permission from the rights holder.

API and Direct Download

Users are allowed access to archive (ZIP) files of all out-of-copyright texts, images and metadata. Although individual texts have been released into the public domain, the dataset as a single object should be properly attributed in derivative works. Users cannot deliver in-copyright digitised material onwards without the permission of the rights holder.

Times Digital Archive

Gale: A Cengage Company

THE TIMES
Digital Archive 1785-2010

History of the Collection

The Times Digital Archive (https://www.gale.com/intl/c/the-times-digital-archive) was the first online digitised newspaper collection of British newspapers. Produced by Gale (then Thomson Gale Publishers), the collection debuted in 2002 with an initial remit to make available the entirety of *The Times*, including its previous incarnations, from 1785–1985, after which date digital text files were already available. This early adoption of digitisation, and its building upon the popularity of the Palmer's index of *The Times*, ensured its prominence in historical and journalistic research, including its use by the House of Lords in researching past legal debates. As of 2019, it was the most searched digitised newspaper database among Cengage's news media collections. The archive is refreshed annually, adding new issues in one-year sets on a rolling basis.

Consulted Libraries

As *The Times* Digital Archive is a single newspaper archive of a continuing title, the entire database is derived from *The Times*'s own archive, which included a complete backfile on microfilm.

Microfilming Projects

Up through 1985, *The Times* backfile is preserved on 35mm microfilm. After that date, issues have been archived using both microfilm and born-digital files. The majority of the backfile microfilming was undertaken in the 1990s, though some reels were processed by Gale's predecessors as early as 1974. The microfilm archives preserve the final London edition of each issue and excluded regional editions, such as the Irish or Scottish editions, or editions which came out earlier in the day. *The Times* was microfilmed from a variety of sources, including bound volumes, and conditions, including defects such folding, tearing, the use of adhesive tape and bad trimming.

Digitisation Projects

Digitisation of *The Times* began in 2002 through the scanning of existing microfilm. Because the initial 1785–1985 material was digitised over just two years, there was a consistency of staff, equipment, method and product, both in terms of image and OCR quality. The reels were scanned using the Mekel M500 greyscale scanner and converted into 300 PPI bitonal TIFFs at a rate of six frames a minute, allowing the digitisation of between two and five reels a day, depending on the quality of the source material and the degree of intervention required by the operator. After filming, the images were subdivided by month and then cropped and de-skewed in preparation for OCR processing. Digital restoration was undertaken to reduce the appearance or impact of damaged pages, including manually cropping and cleaning and the insertion of digital titles or page number where needed. Zoning of articles was done partially through automated processes, leading to the amalgamation of smaller individual units such as classified advertisements, and partially by manual clipping, as was the case for birth, death and marriage notices in order to aid family historians and professional genealogists using the service.

Selection

The aim of the initial project was to digitise the microfilm collection, in its entirety, from 1785 until 1985. The content was released in several batches: the first was 1936–1946, growing monthly to include 1880–1985 by the end of 2002. The entire microfilm run was completed by the close of 2003. Since its acquisition by Cengage in 2007, Gale has continued to expand the collection, which currently offers the complete run of the publication from 1785 to 2013.

Preservation and Access

While the full historical archive of *The Times* has now been digitised, and new issues meet legal deposit legislation digitally, Gale continues to microfilm *The Times* on a monthly basis as a commercial product and as a preservation archive for *The Times* itself. Access is via purchase or subscription only.

Composition of the Collection

Selection Available

The Times Digital Archive currently contains material from 1785–2013. This includes over 1.6 million pages from 70,000 issues, sub-divided or zoned into 11.8 million articles. These are catalogued by category, including advertising, editorial and commentary, news, business, news, people and photojournalism. Although the modern *Times* began publication in 1788, the collection includes digital issues of its precursors, *The Daily Universal Register* (1785–1787) and *The Times*, or, *Daily Universal Register* (1788). The collection continues to expand with additional content added on an annual basis.

Data Quality

Text

As OCR software provides only a digital confidence rating, which cannot be meaningfully translated into a quality metric, and independent quality studies have not yet been undertaken, the overall OCR quality of *The Times* Digital Archive is currently unknown.

The machine-readable text appears within a single XML file per issue, surrounded by layered metadata that describes the features of the issue, pages and articles. Issues between 1785–1985 were created during a single project, undertaken by the same staff, using the same equipment and processes, and working with microfilm that had been filmed over a short period. Therefore, the data for these years has a relatively consistent level of metadata and OCR quality, depending on the age or preservation status of the source material at time of microfilm creation. All issue-level metadata is hand-keyed, alongside the article title, article subheadings, attribution information and illustration captions. Subsequent additions have been included on a rolling basis and their data is contained in a separate but similar substructure within the collection, using the same metadata schema, capture requirements and level of detail but with new metadata fields to provide additional image metadata including the height, width, file format and colour map.

Image

Images in the collection before 2007 were captured as 300 PPI bitonal TIFF files; since then, they have been captured at 400 PPI. These are not compressed or reformatted before display through the web interface.

Metadata Schema

Gale Legacy Text Mining Drives

Before 2016, *The Times* Digital Archive data contains metadata and text content in a single XML file at issue level. Although similar in coverage to the METS/ALTO schema used by many public institutions, Gale established a bespoke metadata schema to label information consistently across its different newspapers and collections. A DTD file is provided on the text-mining drives and the fields are comparable to those found in Dublin Core, MARC and other standard bibliographical standards, to which they have been successfully mapped when working with external content partners.

Each XML file contains bibliographic information for the entire issue, automatically zoned during the OCR process, with individual pages and articles are represented as child elements. At the article-level, each individual word is encoded with spatial coordinates of its location on the corresponding image, as well as marker elements indicating new pages or columns. Metadata fields including publication name, year, date, issue number, page number, article title, article subheading, attribution and illustration capture were manually entered or verified by those processing the data.

Gale Current Text Mining Drives

After 2018, the Gale Text Mining Drive separated metadata and text content into three XML files: title or publication-level metadata, issue-level metadata, and issue-level content data. As with the previous schema, the data is encoded using Gale's standardised metadata schema and a DTD file is provided on the text-mining drives. Although distinct from the METS/ALTO schema, this system is similar to a combination of library MARC records and METS/ALTO XMLs.

Backend Structure

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. In addition to the metadata provided on text-mining drives or online, this database stores image metadata on resolution, file format, bit depth, colour map, file size and image dimensions. Our image database stores image metadata, including image resolution, file format, bit depth, colour map, file size, width and height.

User Interface Structure

Web Interface

The Times Digital Archive can be searched using the Gale Primary Sources interface. The basic search can be filtered to a specific metadata field or the full text, a date range, a specific title or a specific digitised collection. The advanced search allows for standard Boolean operators and fuzzy searching as well as filtering by publication date, publication section, document type and whether an image is included. Results can be sorted by publication date, article title, publication title or relevance and page number. On the search results page, users are presented with additional filters and simple analysis tools, such as term clusters and frequency. Individual results can be viewed at article, page, or issue level. At article

level, the searched terms are highlighted, and users can navigate the issue by moving to other pages or articles within it or refining their search terms. Then users can also adjust the image contrast and brightness to improve legibility, and download it using standard browser context menus. The image may also be downloaded as a PDF, as can the plain text of the OCR content. Bibliographic information and a suggested citation are provided at the bottom of the page.

API

Direct Download or Drives

Gale Cengage offers to make available content from its collections to academic researchers for data mining and textual analysis through physical hard drives containing source data for a nominal cost recovery charge. This includes directories, title manifests, XML files and image files, containing metadata, article segmentation, and page facsimiles.

Rights and Usage

Web Interface

The Times Digital Archive is accessible by institutional subscription or purchase and is currently held by many public or national libraries worldwide; there is currently no individual subscription model available.

API

API access is not currently available. However, users can create batches of specific issues or titles for bulk download through the Gale Digital Scholar Lab (https://www.gale.com/intl/primary-sources/digital-scholar-lab), a separate subscription service.

Direct Download or Drives

Gale Cengage makes content from its collections available to academic researchers for data mining and analysis through physical hard drives for a nominal cost recovery charge. This includes directories, title manifests, XMLs and image files. This data is only accessible to those with institutional subscriptions or purchases of the relevant Gale products. Material obtained on text mining drives may be used to examine individual text for large-scale analysis for purposes of performing personal or non-commercial research but cannot be duplicated or shared without express permission.

Re-Publication

As part of the user agreement, XML, OCR and image data cannot be re-published in any form, physical or digital, without the express permission and licensing of News UK (https://newslicensing.co.uk/en/page/show_home_page.html). Small quotations, using standard citation practices, may be reproduced in accordance with local fair use provisions and should be accompanied by a DOI link that points back to the individual full text article or book chapter and a proprietary notice in the following form: "Some rights reserved. This work permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited."

Trove

National Library of Australia

History of the Collection

Trove (http://trove.nla.gov.au) was launched in 2009 in order to create a single point of entry for online discovery services developed by the National Library of Australia since 1997, including the Register of Australian Archives and Manuscripts, Picture Australia, Libraries Australia, Music Australia, Australia Dancing, the Preserving and Accessing Networked Documentary Resources of Australia (PANDORA) search service, the Australian Research Repositories Online to the World (ARROW) discovery service and the Australian Newspapers Beta service. The digitised newspaper collection (http://trove.nla.gov.au/newspapers) was included alongside these aggregated resources in 2010. It was an extension of the Australian Newspaper Plan (ANPlan) founded in 1992, and aims to make freely available as many Australian newspapers as possible and make sure that they remain available in perpetuity, regardless of future technological change.

The pilot by the Australian Newspaper Digitisation Program (ANDP) aimed to digitise 50,000 pages from twelve (later eleven) titles, moving towards three million pages over four years. These were curated from existing microfilm copies by the State and Territory Libraries of Australia with the aim of providing a sample of historical newspapers evenly distributed across the country. Funding for the digitisation of the newspaper collection has come from various sources. As of 2015, the National and State Libraries Australia, other cultural heritage and research organisations, and community groups had directly funded the digitisation of about half of the newspaper pages available on Trove. This includes the State Library of New South Wales who, alongside the National Library, have been the most significant contributor to digitisation of newspapers and journals. The infrastructure costs are borne by the National Library without an additional appropriation from the government and digitisation not funded by contributors is funded from the library's collections budget.

The Australian Newspapers Beta service was launched in July 2008 as a standalone website, and a year later became a fully integrated part of Trove. Shortly after launch, the system incorporated a platform for crowdsourced text-correction, allowing the public to improve the searchable text, as well as the ability to apply social tags to materials, create curated lists, and leave comments. These features have allowed a high degree of community engagement and enriched the collection. By 2009, users were able to access 720,000 pages of digitised content. By mid-2014, the newspaper collection had grown to 13.5 million pages, claiming the title of the largest digitised newspaper collection in the world, and as of December 2019, Trove contains over 25 million newspaper pages. Since 2018, the National Library of Australia has invested heavily in expanding and improving access to its collections through updated digitisation, API and web interface protocols.

The digitisation, delivery and crowd enhancement services originally associated with newspapers are now also available for other published material within Trove, including government gazettes, journals (magazines and newsletters) and books as well as special collections.

Consulted Libraries

Most of the newspapers within Trove were scanned from microfilm collections held by members of the National and State Libraries Australasia: The National Library of Australia; the State Library of Western Australia; State Library of New South Wales; State Library of Victoria; Libraries ACT; Library and Archives NT; State Library of Queensland; State Library of South Australia; and Libraries Tasmania. Additional collections, held by private organisations, have also been digitised through the digitisation partnership programme.

Over forty Australian public libraries have also selected titles to digitise based on significance to their local communities and have directly funded the digitisation. They have also coordinated local community groups and groups of libraries within their regions to raise the funds required. In some cases, public libraries have also provided previously uncatalogued physical copies for microfilming, as part of the digitisation process. Organisations beyond the library sector have also nominated and funded newspapers for digitisation. This includes local, state and federal governments, historical societies, archives, universities, community groups, foreign embassies and businesses.

Microfilming Projects

Established in 1992 as the National Plan for Australian Newspapers, ANPlan brought together independent programmes of preservation by the National and State Libraries Australia (NSLA), alongside the National Library of New Zealand, which holds observer status. It continued the devolution of the responsibility for collecting, preserving and

providing access to newspaper titles to respective jurisdictions but initiated a coordinating role for the State Library of South Australia. Part of this strategy was to ensure, as far as possible, that at least one hardcopy instance of every newspaper was retained alongside a surrogate copy, such as microfilm, to ensure long-term public access. In 2001, the coordinating responsibility was taken up by the National Library in Canberra. Although microfilming had been the primary means of long-term preservation for fifty years, by the mid-2000s, ANPlan partners had begun to express concerns about the long-term viability of microform preservation, citing concerns about film stock, microfilming services and the cost of suitable storage facilities. These difficulties have been compounded by decreased industry support, including manufacture and repair of microform readers and duplicators as well as user preference for digital delivery.

Digitisation Projects

Selection

It is the aim of the ANDP to make freely available all Australian newspapers published prior to 1955. During the initial phase of the programme, newspaper issues were selected under the National Library of Australia's Australian Newspaper Digitisation Program (ANDP) by the National, State and Territory libraries. When digitisation began in 2007, the library deliberately chose one title from each state and territory in order to be geographically representative. Afterwards, particular attention was paid to the oldest or biggest newspapers from each state, though there was a general desire among the partners not to focus exclusively on newspapers with the largest circulation but also to represent smaller or more remote communities with less reliable physical access to research libraries. Under this devolved model, State and Territory libraries nominated the newspaper titles or issues for digitisation and provided the microfilm from their collections, while the National Library also selected titles for digitisation, its focus shifting year on year to address important themes, such as the First World War, and to represent non-geographical communities. This selection process involved the consultation of newspaper historians and a microfilm supplier. This has led to some unevenness in periodisation across the collections, but this will diminish as digitisation continues. In general, this devolved decision-making process considered user demand, historical significance, geographical and regional coverage, and microfilm status. The initial selection process, therefore, largely depended upon librarian expertise of significance and user demand, as well as the availability of microfilm copies of sufficient quality for digitisation.

Up until the mid-2010s, selection occurred within the framework of the library's Collection Digitisation Policy (https://www.nla.gov.au/policy-and-planning/collection-digitisation-policy), which considered a newspaper's cultural and historical significance, utility to a broad range of audiences, uniqueness or rarity, perceived public demand, conservation status, rights conditions, planned digitisation by other providers, and other practical and technical considerations regarding its digitisation.

Since 2010, individual users and groups have been encouraged to take part in a contributor-funded model, wherein they nominate and fund the digitisation of a title, so long as it falls within their general selection guidelines. Over one hundred and eighty organisations and groups have participated in the programme. In 2020, the National Library implemented a new fundraising strategy under the Treasured Voices initiative to significantly increase its digitisation output. Australian newspapers pre-1955 are a finite set, and digitisation of the entire corpus into Trove is a long-term goal.

As of December 2019, the library maintains online lists of current titles (https://trove.nla.gov.au/newspaper/about), forthcoming titles (https://help.nla.gov.au/trove/for-digitisation-partners/newspapers-coming-soon) and new additions (https://trove.nla.gov.au/newspaper/rss/titles).

Preservation

Although their 2010 five-year plan included discussion of ongoing microfilming for preservation, the most recent strategy document for ANPlan focuses almost exclusively on the digitisation of historical newspapers for preservation and the retention of born-digital newspapers files for legal deposit. A key aim of the 2015–2018 strategy was to implement agreed minimum scanning standards for newspapers across all member libraries and detailed guidance on digitisation from both microfilm and hardcopy is available on the Trove Digitisation Partners webpage (https://help.nla.gov.au/trove/for-digitisation-partners).

Access

The Trove newspapers collection provides users access to the most comprehensive selection of historical Australian newspapers in a single location; it is available free-of-charge, worldwide. Existing

microfilm collections remain accessible at individual State and Territory Libraries as well as the National Library of Australia and, at the discretion of individual libraries and where conservation status allows, users may still consult original hardcopies of historical newspapers that have been digitised.

Composition of the Collection

Selection Available

As for December 2019, Trove Digitised Newspapers provides access to over 25 million pages across almost 1,500 Australian newspapers, including government gazettes, from each state and territory, from the earliest published newspaper in 1803 to 1955, when copyright is assumed to have expired. There are also fifty titles with digitised issues after 1954, and nine after 2000, which have been made available with the agreement of the publisher. This includes the *Canberra Times*, the *Australian Women's Weekly*, *Woroni* and the *Chaser*. In addition to the English-language press, the collections also include Australian publications in community languages such as Bahasa Indonesia, Chinese, Danish, Estonian, French, Gaelic, German, Greek, Italian, Japanese, Macedonian, Polish and Swedish. A list of newspaper titles already digitised is available on Trove, as well as a list of newspaper titles selected for digitisation for the current year (http://www.nla.gov.au/content/new-titles-coming).

Data Quality

Text

A single contractor was responsible for OCR and content analysis in the initial phase, while a panel of OCR and content analysis providers have been used since 2010 to cater for the expanding programme. OCR contractors process page-level image files provided by National Library of Australia according to publicly available guidelines (https://help.nla.gov.au/trove/for-digitisation-partners/digitisation-workflow-process-overview) and provide hand-keyed metadata for key fields of each issue. Afterwards, the Digitisation and Photography Branch engage in a quality control process by which they check a sample of articles from each batch.

The overall OCR quality of the Trove newspapers collection varies owing first to variations in the source materials, and second to the non-systematic inclusion of end-user corrections. During the initial processing, titles, sub-titles, authors and the first four lines of each zoned article are re-keyed, resulting in 99% percent accuracy for these components. Once online, Trove users are encouraged to help improve the accuracy of the OCR text by allowing line-by-line correction, and some users have self-organised into volunteer groups to undertake systematic corrections of certain parts of the collections. As of December 2019, over 333 million column-lines of the OCR text had been manually updated by Trove users, with one user having worked on almost 6 million lines. However, this represents only a small percentage of the growing collection and is not evenly distributed, with a disproportionate number of changes being made to family notices and other material useful to genealogical research. A history of these changes is recorded, allowing staff to roll-back vandalism, and the web interface searches both the original OCR and corrections to it. Articles that have or have not undergone manual corrections can be filtered using the web interface facet "has:corrections", while the API will return the number of corrections and the last date the article was modified.

Thus, any given article within Trove may have had a small or significant manual correction to the original OCR transcription, which itself varies considerably depending the condition and typography of the original item. As these corrections are updated hourly, the OCR quality of the collection should be specifically tested for any sub-corpus used at the time of analysis. Moreover, periodicals in Chinese, Estonian, French, German, and Italian currently have a greater variance in OCR quality than the English-level titles, owing to software limitations. These provisos acknowledged, independent research undertaken in 2013 showed a general OCR accuracy of 80-90%, with the late 1840s rising to 94% and the early 1920s dipping to just under 80%, and the library has undertaken research into how to evaluate the improvement of crowdsourced corrections in order to improve the reliability of their machine-readable text as the number of digitised pages increases.

Images

During its digitisation programme, the majority of Trove newspapers were scanned from 35mm master negative silver gelatin microfilm reels or second-generation silver gelatin microfilm reels into a pair of digital images, consisting of a 400 PPI raw

greyscale TIFF and an Image Optimised Bitonal TIFF. The National Library currently requires hard copy newspapers to be digitised in colour with a bitonal image for each page for OCR purposes. They are required to be formatted as a TIFF 6.0 at 400 PPI, compressed to LZW for the colour master image and CCITT Group 4 for the bitonal image.

Metadata Schema

The OCR metadata contained within Trove utilises the METS XML schema for structural metadata and ALTO XML for the OCR content. The descriptive and bibliographic metadata is largely based on human-inputted records, either by library staff or by human operators at OCR processing facilities. Additional metadata regarding user annotations and corrections is held in a separate metadata schema accessible via the API.

Backend Structure

The data for each issue is stored in multiple image and text files, with two digital image files, including a raw greyscale TIFF image and a bitonal TIFF image, for each newspaper page. File "pairs" will have identical names apart from the character that distinguishes bitonal and greyscale files (g for greyscale file, b for bitonal files; c is used for hardcopy-derived colour images). One XML file contains most of the human-supplied metadata for the issue, conforming to the METS schema. There is then an XML file for each page containing the OCR results using the ALTO schema. Each file has a name consisting of the base, generally "nla.news-issn", followed by the ISSN for the publication (eight numeric characters, sometimes "x" as the final character, with no hyphen), followed by a unique sequence number for that page starting with "-s", then by "-g" for the Greyscale image or "-b" for the Bitonal image, and ending with an extension for the file type. Sequence numbering is continued across scan jobs or microfilm reels for each individual newspaper title so that all file names are unique for a title. Image files are named sequentially based on the order in which they appear in the microfilm.

User Interface Structure

Web Interface

The current user interface allows users to perform a simple or advanced search of the underlying descriptive metadata and OCR text, or to browse images by date, place, category, tag and title. Facets, as well as the advanced search, allow for filtering by article type, article length, illustration inclusion, title, date, and place of publication. The full-text search can be filtered using standard Boolean operators. By default, search results are ranked by relevance but can also be ordered by date. Once a result is selected, a full-page image, centred on the relevant article with highlighted search results, is displayed in an image viewer. The viewer allows users to pan and zoom as well as navigate through the issue. The underlying data (plain text, PDF and JPEG) as well as user-inputted categories and comments can be downloaded using icons at the left of the viewer and the metadata, preferred citation and OCR text can viewed in retractable widgets. When downloading an image, the article is segmented and then embedded into an HTML to facilitate printing onto A4 paper. A new web interface is currently in development.

API

The Trove API provides users with bulk access to the underlying data of the Trove collections, including user-generated data, in a machine-readable form. The API currently allows for the display of Trove results on external websites, the harvesting of data for offline analysis, the retrieval of user annotations, and the creation of new tools and visualisations. A personal API key can be obtained automatically via a Trove user account; a commercial key is also available but requires explicit permission from the National Library. Materials can be accessed using a URL-based request, which assists users in formatting their requests. A full technical description of the API is available through the Trove Help pages.

Direct Download or Drives

The Australian government gazettes and the *Australian Aborigines Advocate* are available for bulk download through the Trove Help pages.

Rights and Usage

Web Interface

The Trove web interface is freely accessible to all users, worldwide. All material obtained from the web interface may be used freely for personal research. When browsed or searched through the user interface, users are presented with a full citation to the digitised image and text.

API

The API is free and open, with a key that is automatically obtainable for personal use. Commercial use requires explicit approval by Trove. Material derived from the Trove API may be used under the same conditions as that derived from the web interface.

Re-Publication

Digitised newspapers up to 1954, whether delivered through the web interface or the API, are available to users as greyscale or colour images and machine-readable texts. Copyright in Australian newspapers is complex. Neither Trove nor the National Library of Australia can grant special permission to use copyrighted items; only the copyright holder can do this. Before reproducing any newspaper articles, the user is asked to confirm whether they are out of copyright. If the article is out of copyright, it is free to use; however, proper attribution and citation should be applied when using all newspaper content.

ZEFYS

Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Germany (Berlin State Library)

Staatsbibliothek zu Berlin
Preußischer Kulturbesitz

History of the Collection

ZEFYS (http://zefys.staatsbibliothek-berlin.de) is a "creative acronym" for the Newspaper Information System, which was expanded to serve as the newspaper portal for the Berlin State Library (SBB). The SBB has the largest and most extensive newspaper collection in Germany. In 1993 the State Library of Berlin established its dedicated newspaper department, the first such department in Germany. In addition to German newspapers, the collection benefitted from funding from the Deutsche Forschungsgemeinschaft (DFG) to facilitate the acquisition of newspapers from abroad that were held in card catalogue form at the library. In 2009, the SBB launched its digitised newspaper portal ZEFYS. ZEFYS currently gives access to 193 newspaper titles with roughly 7 million pages overall.

Consulted Libraries

Digitised newspapers published in ZEFYS typically stem from the collection of the Staatsbibliothek zu Berlin. However, in some cases there have also been collaborations with other libraries or archives to digitise historical German newspapers from holdings that are not part of the SBB collection, and to make them available through ZEFYS, including some newspapers from former German colonies. The *Deutsches Nachrichtenbüro*, for example, was digitised in collaboration with the University of California, Los Angeles. The British Library contributed around 13,000 volumes, and the Federal Press and Information Office provided 40,000 volumes and packages.

Preservation Projects

Newspapers have been included in the Prussian legal deposit regulations since 1699, though it was not until 1824 that an amendment added a specific provision for periodicals. During the mid-nineteenth century, the library (then the Königliche Bibliothek, under the direction of Friedrick Wilcken) expanded its holdings, making the newspaper collection the most comprehensive in Germany. As a result of the evacuations of the holdings during the Second World War, many items of the newspaper collection suffered severe damage and loss. After the war, the library (then the Deutsche Staatsbibliothek), received every newspaper that was published in the German Democratic Republic via legal deposit legislation. The holdings of the Preußische Staatsbibliothek, which were evacuated to western parts of Germany during the war, were the foundation of the Westdeutsche Bibliothek. These collections form the basis of the modern newspaper department, formed after reunification.

The library aims to make available microfilms of newspapers that are in high demand or exposed to frequent use. In 1999 a reprographic company was placed in a studio in the library's Newspaper Department with microfilming supported by the DFG and the conservation funds of the Staatsbibliothek. The Berlin State Library generally digitises from microfilm masters rather than from hard copy original. In addition to the increased efficiency of digitisation, this also preserves the paper originals, and internal evaluations of its OCR have shown that there is next to no quality loss when the microfilm master is used for image capture. Preservation microfilming is still supported on a small scale and the German Research Foundation has mandated that the newly created microfilm be listed in the ZDB catalogue.

Digitisation Projects

In addition to the continuous digitisation of newspapers based on user demand, the Berlin State Library carried out a number of specific newspaper digitisation projects such as the Amtspresse Preußens, which digitised the Prussian political press collection, DDR Presse, which digitised three popular newspapers from the German Democratic Republic (for reasons of copyright, these are only available after log-in), and most recently the Europeana Rise of Literacy project, which digitised the *Vossische Zeitung*.

Selection

The primary driver behind digitisation at the Berlin State Library has been user demand. In addition, the aim of the Digitisation of the Amtspresse Preußens project was to create a collection of the most influential political press of the last third of the nineteenth century, together with commentary and a search function.

Preservation and Access

Microfilm is used as the main means of preservation of the Berlin State Library newspaper collection. Hard copies that have been microfilmed are not typically available for manual inspection, and visitors are instead directed to the microfilm. Digitisation is done mainly for reasons of access, and digitised copies of newspapers do not constitute a replacement of the microfilm masters for archiving purposes.

Composition of the Collection

Selection Available

ZEFYS provides access to 193 newspaper titles, or 311,234 issues, between 1857 and 1939. The main focus of digitisation is on newspapers from Berlin and its surroundings that were of national influence. Highlights of the collection include early papers such as *Berlinische Nachrichten* (founded 1740), foreign German-language papers such as *Deutsche La Plata-Zeitung*, and Russian-language papers published in Berlin. A full list of titles can be found on the ZEFYS website: http://zefys.staatsbibliothek-berlin.de/index.php?id=list.

Data Quality

Text

Only about 50% of the digitised newspapers available in ZEFYS have been processed with OCR. Currently, the OCR text is neither used for full text search nor displayed in ZEFYS. There are plans to integrate the digitised newspapers in ZEFYS with the main presentation platform for digitised collections at the Berlin State Library and, as part of this, implement full text search and display.

The quality of the OCR that does exist varies between 70-80% word accuracy, with newspapers in Fraktur typically having somewhat lower accuracy than those printed in Antiqua. In the past, OCR processing was done using ABBYY FineReader 10 and 11 because of its layout segmentation performance, which is much better than that delivered by open source OCR engines for historic newspaper layouts.

Images

Images are captured with 300 PPI and 8-bit colour (greyscale) in TIFF format with 90% JPEG compression. This slightly decreases file size without negative effects for OCR/OLR processing or viewing images on the web.

Metadata Schema

For the digitised newspapers, the Berlin State Library follows the METS/MODS profile for digitised newspapers, established in collaboration with the German Digital Library (https://wiki.deutsche-digitale-bibliothek.de/display/DFD/Anwendungsprofile+und+Best+Practice+Guides#AnwendungsprofileundBestPracticeGuides-GesamtaufnahmeeinerZeitung). ALTO is used to store the OCR full text.

Backend Structure

Digitised newspapers are organised by title, with subdirectories for year and date of issue. In the case of multiple editions on a single day, issues are separated by unique identifiers.

User Interface Structure

Web Interface

The ZEFYS portal allows users to search and access the digitised newspapers by title and date of the issue. Titles can be browsed by year and country of publication.

API

The Berlin State Library provides an API for the digitised newspapers, which allows retrieval of images, full text and metadata. It is documented at https://lab.sbb.berlin/5393/?lang=en.

Direct Download or Drives

Currently, the API is the only way to bulk-download digitised newspapers from ZEFYS. On request, full exports of images, full text and metadata can be provided via external hard disk.

Rights and Usage

Web Interface

The title-level metadata displayed in ZEFYS provides information about the rights and usage for each newspaper.

Re-Publication

All digitised newspapers assembled in ZEFYS are dedicated to the public domain. Specific terms of use apply for the sub-collections Amtspresse Preußens and DDR Presse.

Metadata Mappings

The maps below are representations of the data and metadata that is comparable, to varying degrees, across the collection. They have been grouped teleologically by their most likely use—content, citation, bibliographical, holdings, description, user interaction and technical information—and given an inclusive but specific definition. We have primarily focused on fields that contain data, rather than mapping all container elements (i.e. nested layers that make up the issue structure). Within each category, examples of academic, technical and industry variants have been included and indexed to better define the broad curvatures of these sometimes-ambiguous concepts and to provide multiple points of access. Finally, each map provides a table with the following standardised information:

- A locator, a four-letter ID indicating the collection and filetype in which the data can be found
- The XPath or JSON path to the data within that collection file
- A data type indicator, comprised of a standardised three-letter ID
- An example of the content of that field, with long strings truncated by a [...]

The data types and collection IDs are as follows:

Data Types

DATA TYPE INDICATOR	DESCRIPTION
BOO	A Boolean char such as 0/1 or Y/N
COO	A set of numeric coordinates to delineate a segment of an image
DAT	A single date
DAR	A range of dates
FIN	A filename
STR	An open-ended string of content; alphanumeric
MCH	Multiple pre-defined choices
NUL	Holds no content; used as a container element for other fields
NUM	Numeric value; may include the symbols . , -
UID	Any form of unique ID or acronym
URL	A URL

Collection IDs			
LOCATOR	**COLLECTION NAME**	**STANDARD**	**DESCRIPTION**
B1GI	British Library 19th Century Newspapers, Part I, Gale's Current Text-Mining Drives	GIFT	Issue Metadata XML File
B1GP	British Library 19th Century Newspapers, Part I, Gale's Current Text-Mining Drives	GIFT	Publication Metadata XML File
B1GT	British Library 19th Century Newspapers, Part I, Gale's Current Text-Mining Drives	GIFT	Text Content XML File
B2GI	British Library 19th Century Newspapers, Part II, Gale's Current Text-Mining Drives	GIFT	Issue Metadata XML File
B2GP	British Library 19th Century Newspapers, Part II, Gale's Current Text-Mining Drives	GIFT	Publication Metadata XML File
B2GT	British Library 19th Century Newspapers, Part II, Gale's Current Text-Mining Drives	GIFT	Text Content XML File
B1JI	British Library 19th Century Newspapers, Part I, British Library's Text-Mining Drives	Bespoke	Content and Metadata XML File
B1GL	British Library 19th Century Newspapers, Part I, Gale's Legacy Text-Mining Drives	GIFT	Content and Metadata XML File
B2GL	British Library 19th Century Newspapers, Part II, Gale's Legacy Text-Mining Drives	GIFT	Content and Metadata XML File
CAAL	Chronicling America	ALTO	Content and Layout XML File
CADI	Chronicling America		Directory Structure
CAME	Chronicling America	METS	Issue Metadata XML File
DEAL	Delpher	ALTO	Content and Layout XML File
DEMP	Delpher	MPEG	Issue Metadata XML File
DEOC	Delpher	Bespoke	OCR Text XML File
EUAL	Europeana	ALTO	Content and Layout XML File
EUME	Europeana	METS	Issue Metadata XML File
F1AL	Finnish National Library 1771–1910	ALTO	Content and Layout XML File
F2AL	Finnish National Library 1771–1910	ALTO+	Content, Layout and Metadata XML File
F1ME	Finnish National Library 1771–1910	METS	Issue Metadata XML File
HNME	Hemeroteca Nacional Digital de México	METS+	Content, Layout and Metadata XML File
HNDM	Hemeroteca Nacional Digital de México	Bespoke	Content and Metadata JSON File
PPAL	Papers Past	ALTO	Content and Layout XML File
PPDI	Papers Past		Directory Structure

PPME	Papers Past	METS	Issue Metadata XML File
SBAL	State Library of Berlin	ALTO	Content and Layout XML File
SBME	State Library of Berlin	METS	Issue Metadata XML File
SBMA	State Library of Berlin	METS	Publication Metadata XML File
SBMY	State Library of Berlin	METS	Publication-Issue Metadata XML File
TDAG	Times Digital Archive	GIFT	Content and Metadata XML File
TRAL	Trove	ALTO	Content and Layout XML File
TRAP	Trove	Bespoke	API XML Return
TRME	Trove	METS	Issue Metadata XML File

Content Data

Data in this section may be considered the content or text data of the file. It includes any text as it appears on the physical object, such as the **title** of the article, **section headings** and **text**. For the sake of consistency, text components, such as **typographical hyphens**, have also been included in this section.

Section Heading

Language Variants

Abschnittsüberschrift; Sectietitel; Osaston Otsikko; Sección de Título.

Technical Definition

Specifies the printed title for a section; distinct from an article headline or title.

Usage Notes

Section is used in modern journalism to refer to different departments of the newspaper, such as travel or sport. The nineteenth-century newspaper similarly had sections for elements such as **letters to the editor**, **advertisements** and **reviews**, with the term most commonly used for news from a specific location. The term "section" was also used to refer to specific areas of the page layout that had been broken up by an image, as well as the division of one article into separate pieces that could be published in different issues. Section headings would be printed in the folio line and, in modern journalism, often have their own mastheads. With the exception of Trove, differentiated sections are generally zoned algorithmically and assigned a standardised **article category** rather than the section heading as printed in that instance.

Examples:

Referring to a divided article

- "Scott also sent his first 'vision' to Croker for publication in the *Guardian*, where it duly appeared–arbitrarily cut into two **sections**–on 19 December 1819 and 9 January." [Garside, 508]

Referring to discrete departments of the newspaper

- "For single-**section** newspapers, the pages or parts of pages only require a modest labelling in scale with the layout, a panel or reverse block, and preferably in the same place on every page." [Evans, 52]
- "The **sectionalisation** of British newspapers is best exemplified by the juggernaut that is *The Sunday Times*." [Reeves and Keeble, 23]
- "For unlike its unstamped predecessors, it did include a **section** of leaded editorials (generally placed 6B–C) that like most newspaper leaders commented directly on current politics." [Jacobs, 624-25]
- "In contrast to newspapers in other countries where opinions are expressed throughout the paper, in the United States, opinion is relegated to the editorial or 'op-ed' **section** of the newspaper." [*Encyclopedia of Journalism*, Kimberley Meltzer, 3.1033]
- "By 1830 its four pages were more structured into **departments** than in 1800 [...]. Departments included Foreign Intelligence, Ship News, Domestic Intelligence, Edinburgh News, Births etc., Police Court, Aberdeen Shipping, Correspondence, Miscellaneous Articles, Postscripts and Markets." [DNCJ, WHF, 1]
- "If we look at historical newspapers we can easily see that their main principle of structuring the content was to indicate from where the news were coming from and therefore to list the news according to their place of origin." [Europeana Newspapers 2015, 15]
- "Nineteenth-century publishers were more open about the conventions of publishing, reminding contributors of press times in their 'Notices to Correspondents' **sections**..." [Hobbs 2018, 113]
- "*The Manchester Guardian* devoted large **sections** to 'commercial intelligence', detailed information on stocks and shares, commercial markets (e.g. Bank of England figures), shipping news, and market prices (both local, such as the Liverpool corn market, and more distant such as Birmingham cattle markets and the London Produce Market)." [Cronin, 580]

Category Notes

This field appears in the Trove API (TRAP), SBME and SBMA only. The content is elsewhere usually included as an article's text or headline.

Individual Collection Notes

TRAP: The newspaper or periodical section this article appeared in. In the Trove web interface, this is shown on the **page** view. A section is part of an **issue**.

Instantiations

SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:relatedItem\mods:titleInfo\mods:partName	STR	Wirtschaftsausgabe
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo\mods:partName	STR	Wirtschaftsausgabe
TRAP	article\section	STR	1, Special Home Feature

Article Title or Headline

Language Variants

Überschrift; Titel; Otsikko; Titular.

Technical Definition

Provides the headline or title of the item. This may be hand-keyed or the result of OCR. Distinct from a section heading.

Usage Notes

For most of the nineteenth century, the front page of major daily newspapers would include classified adverts rather than headlines in a larger font size; in the UK, *The Times* continued this practice until the 1960s. The British New Journalism of the 1870s and 80s saw the introduction of headlines to catch readers' attention. Type-revolving presses, used in the 1840s and 1850s, made it impossible to have headlines across more than one **column** until the rotary presses made it possible to design a layout horizontally rather than simply vertically. Even when the technology allowed, Australian newspapers were similarly reluctant to adopt them until the end of the century. The term itself was more commonly used in the nineteenth century for book layouts, referring to the running title, **pagination** and other information at the top of the page. Within the pages of the newspaper, article titles were often fixed from **issue** to issue (for example "Advertisements & Notices", "Provincial News"), comparable to the idea of a regular column today. Because titles were not commonly used (they were more common in periodicals), many databases do not zone by article but by page. Gale segments by article, but with older newspapers they do not try to figure out where one article starts and ends. Many providers hand-key and check titles and headlines.

Examples:

Front-page headlines

- "Newspapers at the beginning of the nineteenth century made no effort, in their vertical display, to put a **headline** over the main story or to make sure that the main story began at the top of a page." [Evans, 23]
- "New Journalism pioneered the use of **headlines** to give an eye-catching typographical emphasis to the latest piece of important news." [DNCJ, AK/MaT, 450]
- "They serve, first, as a signpost for readers, telling them what the most important stories are; those with the bigger **headline** size." [Franklin, 226]
- "Laajan lukijakunnan saavuttamiseksi alettiin käyttää uudenlaista iskevää **otsikkotyyppiä** unohtamatta sensaatiojuttuja ja human interest -aineistoa."
- "Beneath a **headline**, 'Return of the North End team. Brilliant ovation from twenty-seven thousand people', the *Herald* reported that the Public Hall 'was packed as it has never been packed before by an audience representative of all classes of the community', emphasising the unifying nature of the occasion." [Hobbs 2018, 343]

Individual article titles

- "The *Chronicle* again demonstrated its support for Nield with a report (29 Sept. 1894) **entitled** 'Labour Meeting at Crewe: Messers Compton and the Factory Girls'." [Bunting, 141]
- "Wilkie Collins, writing in Dickens's weekly miscellany *Household Words* in 1858 was more pessimistic, despairing of the quality of the 'penny fiction weeklies' that were catering to a vastly increased readership, the 'Unknown Public' of the **title** of his now famous article." [Shattock 2017b, 3]
- "Nevertheless there are several ways to detect the **headline** in an indirect way, by utilizing the background knowledge which is set out here. E.g. repeated section headings, the mentioning of the place and date of the news and the copyright statement may be utilized for this purpose." [Europeana Newspapers 2015, 38]

Recurring column titles

- "This penny satirical journal was started during the Reform agitation, and originated those periodical comments on magisterial decisions which subsequently became general in the newspapers, under the **title** 'Justices' Justice'." [Vizetelly, 1.218]
- "I supplied dramatic criticism, and a column of **'literary and artistic gossip**,' my first attempt at anything of the kind." [Yates, 1.265-66]
- "The subjects I wrote upon in the leading **column** of the *Daily Telegraph* were, comparatively speaking, innumerable…" [Sala 1895, 1.330]

Category Notes

This field is common but does not refer specifically to headlines in the sense of front-page news items, nor does it make a distinction between headline and article title.

Individual Collection Notes

B1GL: Provides the headline as both individual words and a unified string.

B2GL: Provides the headline as both individual words and a unified string.

TDAG: Provides the headline as both individual words and a unified string.

Instantiations as individual words			
B1GL	issue\article\text\text.title\p\wd	STR	PROVINCIAL
B2GL	issue\article\text\text.title\pg\wd	STR	PROVINCIAL
TDAG	issue\page\article\text\text.title\p\wd	STR	To

Instantiations as full headline or title			
B1GL	issue\article\ti	STR	Advertisements & Notices
B1GI	issue\page\article\ti	STR	Advertisements & Notices
B2GL	issue\article\ti	STR	PROVINCIAL NEWS
B2GI	issue\page\article\ti	STR	Multiple Classified ads
DEOC	text\title	STR	Het jaar 1869.
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\ didl:Re-source\srw_dc:dcx\dc:title	STR	EXTRAKTEN UIT DE TWEEDE EDITIE VAN ONS Vorig Nommer. AMSTERDAM, Zandag 29 Junij. Belgische Dagbladen.
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\ srw_dc:dcx\ dc:title	STR	Zevende verstrekking van Roode Kruisgoederen
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div\ mets:div\mets:div\mets:div\mets:div@LABEL	STR	Aus der Goersch'schen Chronik
TDAG	issue\page\article\ti	STR	LISBON.-In HENRY ASTON BARKER'S PANORAMA, Leicester-square, a view of that beautiful City
TRAP	article\heading	STR	Agriculture around Kelvin Grove and Kedron Brook. [BY OUR AGRICULTURAL REPORTER.]
TRME	mets:mets\mods:dmdSec\mods:mdWrap\mods:xmlData\mods:-mods\ mods:titleInfo\mods:title	STR	Local and General.

Attribution

Language Variants

Zuschreibung; Auteur; Lisänimi; Atribución.

Technical Definition

The name of the author of the newspaper article, as printed.

Usage Notes

The term **by-line** did not come into use until the twentieth century, though writers' names could be included in the headnote, and the term is occasionally anachronistically applied. Debate about anonymous contributions versus signed articles was rife throughout the nineteenth century: journals such as *The Athenaeum* and *The Saturday Review* implemented anonymity as policy until the 1890s, seeing it as "conducive to the public interest" [Tener, 63]. Similar debates were happening elsewhere; for example, when the national daily newspaper *El Universal* was established in Mexico in 1888, signatures were removed in favour of having the newspaper present itself as a unified institution. This might explain why so few of the collections include author information. Journalism was only recognised as a profession in the mid-nineteenth century, and with recognition came subdivisions such as **reporter** and **columnist**. Occasional contributors, generally known as "special correspondents" in the Anglophone press, also played an important role.

Examples:

Writers' names within the newspaper layout

- "The **headnote** introduces the name pointedly though neatly, and no one else among the Morning Post's contributors of verse at this period seems to have used it, in relation to 'Sappho' or otherwise, although pen names of the same kind were much in evidence." [Landon, 395-96]
- "Under her usual **byline**, Greville's article for *The Graphic* describing this event competed with more substantial articles in *London's Morning Post, Times*, and *Daily News*; her article nonetheless reveals her importance as a writer, at least in terms of what she called fashionable Society (*Gentlewoman* 108)." [Cogdill, 184-85]
- "In the 20th century **contributors**, such as authors/journalists, photographers, illustrators or cartoonists, are explicitly marked in the article whereas historical newspapers do very rarely mention the actual writer." [Europeana Newspapers, 21]

Writer as regular contributor

- "It is true that newspaper **writers** were compelled to devote their time, energy, and interest to the loud and ephemeral demands of the city room; and while, at first glance, this atmosphere may seem stifling, it was actually a richly rewarding and fertile one for the men who were to lead the revolt against the ideals of the genteel tradition." [Kwiat, 99]
- "These sentiments, espoused by the regular *Daily Mail* **columnist**, known only as Lady Charlotte in 1896 as part of her 'Woman's Realm' feature, epitomize the flagrant contradictions afforded to professional women towards the close of the nineteenth century." [Hunt, 130]
- "He was an owner, editor, and **featured contributor** of the Buffalo Express." [Branch, 584]

Journalist as profession

- "The word '**journalist**' retained negative connotations from the previous century as a partisan hack employed by a patron or political party." [DNCJ, MR, 326]
- "For Stead, however, often considered the father of investigative journalism, this must have been truly what a **journalist** was supposed to do: he not only investigated the injustice he wanted to uncover, he also became involved." [Demoor 2013, para. 13]
- "Slender of means but sanguine of temperament, Wilfrid worked as a **journalist** and editor, and the couple began a hectic period of production, both literary and familial." [Gray, 148]
- "The **reporter** should be seen as a subset within the developing profession of journalism, and a result of the division of labour within this new occupation." [DNCJ, MaT, 536]

- "The nineteenth-century **reporter** Emily Crawford had a formidable journalistic reputation." [Pusapati, para. 1]

Non-professional or occasional writers

- "One of the most prolific **contributors** was Thomas L. Masson. Whilst working as a travelling salesman in New York, Masson submitted jokes and humorous essays to *Life*, one of the period's most successful comic periodicals." [Nicholson 2012, 280]
- "Although her articles hark back to much earlier ideas of woman's place, there is an unmistakable air of change on the horizon in the very possibility that woman can stand equal to man, as Lady Charlotte does to some extent in offering her columns, frivolous though they may be, alongside those of male **contributors**." [Hunt, 130]
- "In the course of time, however, various other poems by Wordsworth have been found in the *Morning Post* during the period when Coleridge was a recognized **contributor**, from late 1797 to 1802." [Landon, 392]

Category Notes

This field only appears in TRME and takes the name from the printed newspaper. This means the name listed may be a pen name.

Instantiations			
TRME	mets:mets\mods:dmdSec\mods:mdWrap\mods:xmlData\mods:-mods \mods:name@mods:namePart	STR	Conducted by "Penelope."

Article Subheading

Language Variants

Untertitel des Artikels; Ondertitel van Artikel; Alaotsikko.

Technical Definition

Smaller titles used to break up the article.

Usage Notes

Subheadings were not common in nineteenth-century journalism anywhere until the later decades of the century. For example, in the UK, the advent of New Journalism in the 1870s and 80s saw publishers beginning to experiment with layout to attract a wider readership. Prior to this, **headlines** were not common and even individual article titles were relatively rare. Any reference to a subtitle would usually be referring to a **subtitle for the newspaper**. The term "subheading" generally refers to any small headline inserted into a run of text. There are two main styles: the **crosshead**, centred on the text below it, and the **sidehead**, which is aligned to the left edge of the text. Some Dutch and Spanish newspapers set subheadings aligned to the right edge of the text. Europeana also allows for a top heading above the main title, though this would usually not appear until the twentieth century.

Examples:

- "Columns are basic units of design in newspapers and periodicals that distinguish serials from most printed books and predate later additions of page design such as tiered headlines, **subheads**, font variation and imaginative layout." [DNCJ, LRB/AK, 134]
- "It may seem a simple matter to recommend the **sub-heading** as a relief to the reader in a column of text, yet this humble device has a curious way of arousing passion. There are newspaper designers today who regard it as a gimmick, a confession of design failure, just as the Victorian critics of the 'New Journalism' regarded it as a frivolous intrusion in an eight-column Budget speech." [Evans, 107]
- "**Sub-headings** appear rather late in the first half of the 20th century." [Europeana Newspapers 2015, 40]

Category Notes

This field only appears in the Trove METS file.

Individual Collection Notes

TRME: Subheading as printed on the article.

Instantiations			
TRME	mets:mets\mods:dmdSec\mods:mdWrap\mods:xmlData\mods:-mods\ mods:titleInfo\mods:subTitle	STR	There and Back for a Dollar

Text

Technical Definition

Article text content. For text content in article titles, See title of the article. For text content in subheadings, see **article subheading**.

Usage Notes

In literary studies, the text has a more abstract meaning somewhat removed from the physical conditions of publishing: the same text can be serialised, published in book form, and made available online. In periodical studies, article text is not always referred to as such: articles can be referred to as **items, pieces, text, columns**, and so on. In the collections, text can be broken down by word, string, article or page.

Examples:

- "My first task was to tabulate the number of emotions expressed and observed in each serial installment of The Woman in White in order to reveal the emotional fluctuations occurring throughout the **text**'s serialization…" [Beekman, 11]
- "The article's tone, especially in its opening passages, is not unlike that of many early-to-mid-century newspaper articles that detail accidental deaths, for its **text** lingers not only on the deep sadness of the event (its melancholic, lamentable, mournful, grief-filled, and awful qualities) but also on its potential to teach readers a lesson." [Fieldberg, 14]
- "In the absence of anything like a working **text**, or even a rudimentary bibliography, it is essential to return to the original pages of the paper." [Garside, 505]
- "The ActivePaper Archive, created by Olive Software, used an image processing technique called 'segmentation,' which breaks each page of newspaper **text** down into its smaller information units (articles, pictures, advertisements)." [Edmund King, 175]
- "For example, some were interested in researching past crimes, or others interested in public transport history would undertake very heavy **text** correcting in stories that involved these topics" [Alam, 2012]
- "Currently there are mainly two ways how digitised newspapers are treated: Either newspapers are scanned, ordered on issue level and enriched with full-**text** on page level. Or the structuring is done on 'article level' which means that all articles are separated and structured." [Europeana Newspapers 2015, 11]
- "Consider the following newspaper **items** dating from the eight years between the summer of 1853 and the summer of 1861." [Branch, 576]
- "Though these **items** are probably the source of the references in the novel…" [Rosengarten, 593]
- "In March 1819, the *Alexandria Herald* published the following anonymous **piece**…" [Gelmi, 151]
- "Between 1851 and 1861 he contributed several dozen **pieces** of varying kinds to newspapers." [Branch, 583]

Category Notes

This field is common across the collections.

Individual Collection Notes

B1GT: Full text of article, run through OCR.

B2GL: issue\article\text\text.title\pg\wd refers to words in the title, while issue\article\text\text.cr\p\wd refers to text.

B2GT: Full text of article, run through OCR.

CAAL: Can indicate half of a hyphenated word.

TDAG: issue\article\text\text.title\pg\wd refers to words in the title, while issue\article\text\text.cr\p\wd refers to text.

TRAP: The full text of the article, including all corrections. Article paragraphs are enclosed in a <p> element. Lines are enclosed by . Article text is not available for some articles, such as those from the Australian Women's Weekly and articles with a status of "coming soon". For these articles, there will be no articleText element present.

Instantiations			
At word level			
B1JI	BL_newspaper\BL_page\pageText\pageWord	STR	The
B2GL	issue\article\text\text.cr\p\wd	STR	PROVINCIAL
HNME	METS:mets\METS:dmdSec\METS:amdSec\METS:mdWrap\ METS:xmlData\hiddentext\pagecolumn\region\paragraph\line\ word	STR	BALTAS
SBAT	PcGts\Page\TextRegion\Textline\Word\TextEquiv\Unicode	STR	futeſte
TDAG	issue\page\article\text\text.cr\p\wd	STR	PA.
At string level			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ CONTENT	STR	.vet
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ CONTENT	STR	WOENSDAG
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ CONTENT	STR	Moren-Ausgabe
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ CONTENT	STR	I
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\ TextBlock\ TextLine\String@CONTENT	STR	204
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMar-gin\ PrintSpace\TextBlock\TextLine\String@CONTENT	STR	omat
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ CONTENT	STR	Hawke's
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\ TextBlock\TextLine\String@CONTENT	STR	y
At text line level			
SBAT	PcGts\Page\TextRegion\Textline\TextEquiv	STR	futeſte Feimatzeitung für die Stadt Garuth und für die Amtsbezirke papſitz und Radeland (gegr. 1865)
At paragraph level			
SBAT	PcGts\Page\TextRegion\TextEquiv	STR	— futeſte Feimatzeitung für die Stadt Garuth und für die Amtsbezirke papſitz und Radeland (gegr. 1865)
At article level			
B1GT	articles\artInfo\ocrText	STR	AUgck aul tirn}fr,[…]
B2GT	articles\artInfo\ocrText	STR	Lt-vtUJu. […]
DEOC	text\p	STR	fcdi.' hl […]
TRAP	article\articleText	STR	<p><span> […]

Hyphenation

Technical Definition

Provides information about words that have been typographically hyphenated.

Category Notes

This field is common in ALTO files.

Individual Collection Notes

CAAL: String@SUBS_CONTENT indicates the full word that is hyphenated. HYP@CONTENT always contains "-".

DEAL: String@SUBS_CONTENT indicates the full word that is hyphenated. SUBS_TYPE indicates that a word is hyphenated, and which part of the hyphenation is contained. HYP@CONTENT always contains "-".

EUAL: String@SUBS_CONTENT indicates the full word that is hyphenated. SUBS_TYPE indicates that a word is hyphenated, and which part of the hyphenation is contained. HYP@CONTENT always contains "-".

F2AL: String@SUBS_CONTENT indicates the full word that is hyphenated. SUBS_TYPE indicates that a word is hyphenated, and which part of the hyphenation is contained. HYP@CONTENT always contains "-".

PPAL: String@SUBS_CONTENT indicates the full word that is hyphenated. SUBS_TYPE indicates that a word is hyphenated, and which part of the hyphenation is contained. HYP@CONTENT always contains "-".

Instantiations

CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String\SUBS_TYPE	STR	HypPart2
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String @SUBS_CONTENT	STR	reserves
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@CONTENT	STR	-
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP	NUL	
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String @SUBS_CONTENT	STR	Kruisgoederen
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@SUBS_TYPE	STR	HypPart2
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@CONTENT	STR	-
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String @SUBS_CONTENT	STR	Hamburgrr
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@SUBS_TYPE	STR	HypPart1
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@CONTENT	STR	-
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String@SUBS_CONTENT	STR	Vanhurskaat
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String@SUBS_TYPE	STR	HypPart1
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\HYP@CONTENT	STR	-
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\ PrintSpace\TextBlock\TextLine\String@SUBS_CONTENT	STR	Vanhurskatt
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\ PrintSpace\TextBlock\TextLine\String@SUBS_TYPE	STR	HypPart1
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\ PrintSpace\TextBlock\TextLine\String\HYP@CONTENT	STR	-
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String @SUBS_CONTENT	STR	October
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@SUBS_TYPE	STR	HypPart1
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@CONTENT	STR	-

Supplement Title

Language Variants

Titel aanvullen; Título Suplementario; Täydennä Otsikko; Ergänzungstitel.

Technical Definition

The physical newspaper or periodical section this article appeared in if not the **issue** itself; i.e. if it appeared in a **supplement**.

Usage Notes

This is different to sections within the **layout**.

Examples:

- "**Supplements** sustained their parent titles, contributed to their formal organization and enhanced their editorial contents." [DNCJ, LRB, 610]
- "The *Weekly Dispatch* not only emphasized its connection to Cook by distributing her portrait as a newspaper **supplement** but also capitalized on her fame by frequently mentioning her name in advertisements aimed at potential subscribers." [Easley 2016a, 698]
- "He was an associate editor of the Sunday **supplement**." [Kwiat, 102]

Category Notes

This field appears in Trove API (TRAP) only.

Individual Collection Notes

TRAP: In the Trove web interface, this is shown on the page view. A supplement is part of an issue.

Instantiations			
TRAP	article\supplement	STR	Something to do

Citation Metadata

The metadata in this section represents the locator information used in identifying specific **articles** or **pages** within the **run** of a **publication**. It includes information about the **volume**, **issue**, **edition**, **date** and **publisher**.

Volume Number

Language Variants

Volumen; Jaargang; Vuosikerta; Volumen.

Technical Definition

Provides the volume information, either a numerical volume number relative to the **newspaper title** or a **unique identifier**. One volume comprises many issues.

Usage Notes

The term "volume" refers to either: collections of several newspaper **issues**, published at the end of the year, by a certain writer (or on a certain theme), or bound together from the issues as they were published; or, a numbering decision made by the **publisher**, often to indicate a new publisher or editor. As newspaper issues were intended to be printed and consumed quickly, rather than kept and revisited, publishing volumes offered readers a more expensive, lasting edition of the articles. Sometimes, the new volumes would include additional **illustrations**.

The term is often used to refer to the bound collections held by libraries. Many digitisers rely on these volumes as the bound format better preserves them; however, this binding does not guarantee completeness or correct sequencing, as these volumes can be dedicated to one specific newspaper, or issues from a specific year taken from a range of newspapers. Most digitised collections began with microfilmed collections, which in turn were largely derived from bound volumes. This affected the suitability of many microfilm reels for digitisation and OCR processing. With regard to the Trove collection, the State Library of South Australia generally disbound its copies when **microfilming**, as did the Library of Victoria, or they loosened the bindings during the microfilming process in order to deal with **margins** and so on. Other digitisers indicate that they do not usually disbind, as it is not usually possible to re-bind them, and this threatens the physical object; many collections in the US were damaged during twentieth-century microfilming owing to the destruction of this binding. More recent digitisation at the British Library does typically involve the disbinding of volumes, followed by the re-wrapping issues in plastic for preservation.

Examples:

- "I was at that time writing in the *Morning Chronicle* newspaper (of which one series had lately been collected and published in two **volumes**, illustrated by my esteemed friend Mr. George Cruikshank)..." [Dickens, 11.76]
- "They were not appended to any of the material in the original folio sheets, and although the newspaper continued publication to 1752 there were no further reprint **volumes**." [Lockwood, 91]
- "Lehden **vuosikerta**, numero, osa, sivu, päiväys/ vuodenaika (painettu lehti, journaali)"
- "I keep on my knees a **volume** of the *Illustrated News* of many years ago ... or a **volume** of *Punch* published between the 'forties and the 'fifties" [Sala 1894, 1.xiii]
- "I saw it the other day, to my infinite amusement, reproduced in a **volume** compiled from the 'agony column'." [Yates, 1.328]
- "A newspaper may restart its numbering with **volume** 1, number 1, or an alternative designation that indicates a renumbering, without a change in title. This happens frequently with newspapers, often with a new publisher or editor." [Sagendorf and Moore, 26]

Category Notes

This field is relatively uncommon. In F1ME and PPME, volume information is located within a string including **newspaper title** and **date** information.

Individual Collection Notes

DEMP: Refers to the volume number as printed on the original.

F1ME: The volume information provided includes the **newspaper title** and the **year** of publication.

PPME: The volume information provided includes the **newspaper title** and the **date** of publication as well as a numerical volume number.

Instantiations			
Relative			
B1GL	issue\volNum	NUM	1
BIJI	BL_newspaper\BL_page\issue_metadata\volumeNumber	NUM	1
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:Identifier\mods:part\mods:detail[@type="volume"]\mods:number	NUM	18
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx:volume	NUM	10
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:part\mods:detail[@type="volume"]\mods:number	STR	Vol. 3
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:part\mods:detail@mods:number	NUM	41
Unique identifier			
F1ME	mets\structMap\div\div[@TYPE="VOLUME"]@LABEL	STR	Suomi no. 1 1841
PPME	mets\structMap\div\div[@TYPE="VOLUME"]@LABEL	STR	Daily Southern Cross no. I 1843-04-22

Issue Number

Language Variants

Zeitungsausgabe; Nummer; Numero; Tema.

Technical Definition

Gives the **issue number** for the item. This is sometimes a string, as printed, and other times a numerical value. It can also take the form of a unique identifier for the newspaper **issue**.

Usage Notes

The term "issue" is applied in academic and contemporary accounts to the newspaper as a physical item, as a conceptual unit, and for additional issues (i.e. the special issue). In contemporary Victorian accounts, the distinction between the verb and the noun is less distinct, and as such we have references to "the first week of [the newspaper's] issue" [Vizetelly, 1.10] or the discontinuation of "the issue of the paper" [Yates, 1.324]. The term "**number**" was more commonly used in the nineteenth century and, as such, "**special issue**" does not appear commonly in the literature.

Examples:

As a physical object

- "The last **issue** in the Mitchell Library, Sydney, is marked, 'Vol. xl., No. 4503,' and dated 'October 20, 1842.' It was:--'Printed and Published by Richard Sanderson, Proprietor, at *The Sydney Gazette* Office, Lower George-street, New South Wales.' There is no intimation in this number that it is the last nor that it was about to cease publication." [Ferguson, Foster and Green, 69-70]
- "It may well have been at her father's suggestion that Charlotte sent for the files of the Mercury to aid in her research; certainly Mr. Brontë would have recalled those old **issues** of Baines's paper, both on the grounds of his own contributions, and because of its detailed reporting of the Luddite disturbances near Hartshead, where he was curate from 1811 to 1815." [Rosengarten, 591]
- "The paper's liberalism showed itself even in the second half of each **issue**, that is, in the part devoted to reviews of literature and the fine arts." [Jump, 42]
- "he had an inveterate propensity for starting newspapers, magazines, and weekly periodicals, usually without the requisite capital for carrying out those publications to a successful **issue**." [Sala 1895, 1.198]
- "he had great pleasure in accepting the poem, which would appear in an early **number** of the magazine." [Yates, 1.222]

A conceptual unit

- "However, other jokes printed in the same **issue** appeared in the US as early as the previous September..." [Nicholson 2012, 281]
- "Speaking of the 'Familiar Epistle from a Parent to a Child' at the start of the March **issue**, in which Dickens facetiously addresses *Bentley's* as a toddler once, but no longer, in his care, the *Examiner* comments that 'every sorrow has a "sunny side",' for 'Mr Dickens is succeeded in his late office by Mr Ainsworth—whose opening of the 'romantic legend' of Jack Sheppard has already, under circumstances ominous of a certain and speedy decline, infused new promise into *Bentley's Miscellany*.'" [Droge, 41]
- "No **issue** of a serial ever exists on its own but calls up the memory of its predecessors while projecting its successors into the future." [James Mussell, "Repetition," 345, qtd. in Droge, 45]
- "An assignment to do a double-page spread, with illustrations, for the Sunday **issue** of the Globe brought him into close relationship with the art department." [Kwiat, 112]
- "Within every newspaper **issue** some elements can be found which are not directly part of the content but are only included for providing some basic information to the user." [Europeana Newspapers 2015, 13]
- "Lehden vuosikerta, **numero**, osa, sivu, päiväys/vuodenaika (painettu lehti, journaali)"

A special issue

- "... the first account of his activities does not appear in the *Mercury* until its **special issue** of January 9, 1813, devoted to the proceedings of the special commission at York." [Rosengarten, 596]
- "it continued to grow, hitting a new record in 1863 when its **special issue** marking the wedding of the Prince of Wales sold no less than 310,000." [Cranfield, 171]

Category Notes

This field is common. In F1ME and PPME, a string including the publication **date** and **newspaper title** is given rather than a numerical value.

Individual Collection Notes

B1GL: issue\id specifies that this is the source issue number. issue\id is composed of an abbreviated title and issue date. The issue\newspaperId is a unique

identifier assigned to the newspaper issue for online delivery.

B1GI: Specifies that this is the source issue number. The UID includes an abbreviation for the **newspaper title** and **date**.

B1GL: issue\id specifies that this is the source issue number. issue\id is composed of an abbreviated title and issue date. The issue\newspaperId is a unique identifier assigned to the newspaper issue for online delivery.

B2GI: Specifies that this is the source issue number. The UID includes an abbreviation for the **newspaper title** and **date**.

DEMP: Issue number as printed on the original.

TDAG: Specifies that this is the source issue number.

Instantiations			
Relative to newspaper title			
B1GL	issue\is	NUM	6209
B1GI	issue\metadataInfo\is	NUM	2608
B1JI	BL_newspaper\BL_page\issue_metadata\issueNumber	NUM	2731
B2GL	issue\is	NUM	6209
B2GI	issue\metadataInfo\is	NUM	1196
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:Identifier\mods:part\mods:detail[@TYPE="ISSUE"]\mods:number	NUM	5
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx:issuenumber	NUM	81
EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo\mods:partNumber	NUM	613
EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo@ID	UID	MODSMD_ISSUE1_TI1
EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo@ID	UID	MODSMD_ISSUE1_TI1
EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem@ID	UID	MODSMD_ISSUE1_RI1
EUME	METS:mets\mets:dmdSec@ID	UID	MODSMD_ISSUE1
F1ME	mets\dmdSec@ID	UID	MODSMD_ISSUE1
F1ME	mets\structMap\div\div\div@DMDID	UID	MODSMD_ISSUE1
F1ME	mets\dmdSec\mdWrap\xmlData\MODS:mods\MODS:titleInfo\MODS:partNumber	NUM	1
PPME	mets\structMap\div\div\div@DMDID	UID	MODSMD_ISSUE1
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:part\mods:detail[@type="issue"]\mods:number	STR	Stück 9
TDAG	issue\metadatainfo\is	NUM	8652
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:part\mods:detail\mods:number	NUM	2044

Unique identifier			
B1GL	issue\id	UID	WOJL-1822-01-03
B1GL	issue\newspaperId	UID	NCBL0081
B1GI	issue\metadataInfo\PSMID	UID	WO1_ANJO_1798_01_02
B2GL	issue\id	UID	WOJL-1822-01-03
B2GL	issue\newspaperId	UID	NCBL0081
B2GI	issue\metadataInfo\PSMID	UID	YOHD-1813-07-31
CAME	mets@LABEL	UID	National tribune (Washington, D.C.), 1898-11-10
DEMP	didl:DIDL\didl:Item\dc:identifier	UID	ddd:010419500:mpeg21
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:d-cx\dcx:recordIdentifier	UID	ddd:010419500:mpeg21
EUME	METS:mets@LABEL	STR	Hamburger Nachrichten no. 613 31.12.1932
EUME	METS:mets\mets:structMap[@LABEL="Logical Structure" @ TYPE="LOGICAL"]\mets:div\@LABEL	STR	Hamburger Nachrichten no. 613 31.12.1932
F1ME	mets@LABEL	STR	Suomi no. 1 1841
F1ME	mets\structMap[@LABEL="Logical Structure" @TYPE="LOGI-CAL"]\div\div\div@LABEL	STR	Suomi no. 1 1841
F1ME	mets\structMap[@LABEL="Logical Structure" @TYPE="LOGI-CAL"]\div@LABEL	STR	Suomi no. 1 1841
PPDI	abbreviatedTitle]_[yearmonthday]	UID	
PPME	mets\@LABEL	STR	Daily Southern Cross no. I 1843-04-22
PPME	mets\structMap[@LABEL="Logical Structure" @TYPE="LOGI-CAL"]\div@LABEL	STR	Daily Southern Cross no. I 1843-04-22
PPME	mets\structMap[@LABEL="Logical Structure" @TYPE="LOGI-CAL"]\div\div\div@LABEL	STR	Daily Southern Cross Southern Cross;New Zealand Herald no. I 1843-04-22
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\ mods:mods\mods:recordInfo\mods:recordIdentifier	UID	PPN1053698984_19340510_020
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\ mods:mods\mods:part@order	UID	19340510020
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\ mets:div\mets:div\mets:div\mets:div@ORDERLABEL	UID	19340510020
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\ mods:mods\mods:identifier	UID	PPN791048292
SBMY	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\ mets:mods\mets:recordInfo\mods:recordIdentifier	UID	PPN1053698984_1934
SBMY	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\ mets:mods\mods:relatedItem\mods:recordInfo\mods:recor-dIdentifier	UID	PPN1053698984_1934
TDAG	issue\metadatainfo\PSMID	UID	0FFO-1812-JUL14
URL			
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\ mods:mods\mods:identifier	URL	http://resolver.staatsbiblio-thek-berlin.de/SNP27779154-19340510-20-0-0-0

Page Number

Language Variants

Seitennummer; Número de Página; Sivunumero, Paginanummer.

Technical Definition

Provides an ID for the page. This is divided into: unique identifiers; page image numbers; identifiers across the database; URLs to web-accessible versions of the page; relative numerical identifiers, within the **issue**; and string descriptors.

Usage Notes

In contrast to the myth that Victorian writers were paid by the word, it was much more usual for newspapers to request an article of a certain number of **pages**. The size of a newspaper **issue** in the nineteenth century varied substantially depending on the publication **genre**, **format**, **periodicity**, **edition**, and **audience**. In 1859 the *Glasgow Herald*, for example, had eight pages Monday to Friday, and four pages Tuesday to Thrusday. Added to this, the newspapers expanded as improvements in technology reduced the price of paper. The decision was not always a straightforward expansion: the *Sheffield Weekly Telegraph* switched to a smaller paper size in 1887 but doubled the number of pages. In Australia, early newspapers mostly consisted of four pages. In Delpher, page numbers are generated semi-automatically and checked by an operator.

Examples:

- "**pages** have to be read sequentially (or, the skipping of pages has to be done in sequence); the entire contents of an article, or of a page, has to be read, for the user to avoid missing relevant information; copies of an article, a photograph, or of a page have to be paid for." [Edmund King, 167]
- "During that time period, at least, the paper was published in eight-**page** editions with each page containing four columns." [Simons, 389]
- "*Graphic* was published on Saturdays, and 'Place aux Dames' appeared every week, typically in the middle of the newspaper and on one **page**; often, her paragraphs make up the top and bottom third, with illustrations of some recent event, sometimes a sporting event, comprising the middle third; sometimes her paragraphs take up the left two-thirds of the page, with unrelated illustrations on the right." [Cogdill, 182-83]
- "The situation is even more complex, since—as Sue's very long note at the end of his feuilleton shows––if the top of the **page** fictionalizes, the bottom of the page defictionalizes by means of extra-narrative digressions, by the insertion of press articles within the feuilleton, and by the inscription of the novel's episodes in current public debates." [Thérenty, 38]
- "On the following **page** there is an outline of the series of decrees, British and French, which led up to the American embargo on trade with Britain in 1811." [Rosengarten, 592]
- "In the absence of anything like a working text, or even a rudimentary bibliography, it is essential to return to the original **pages** of the paper." [Garside, 505]

Category Notes

This field is found in all collections.

Individual Collection Notes

B1GL: The unique identifier is a page image ID. The issue\article\pi\pgref field specifies the relative page ID for the unique page image ID. The issue\article\ci\pgref specifies the page ID for the clipped article image, and issue\article\text\text.cr\pg@pgref specifies the relative page ID for the generic clip rectangle, which is the same as the article zone.

B2GL: The unique identifier is a page image ID. The issue\article\pi\pgref field specifies the relative page ID for the unique page image ID, the issue\article\ci\pgref specifies the page ID for the clipped article image, issue\article\text\text.cr\pg@pgref specifies the relative page ID for the generic clip rectangle (which is the same as the article zone), and issue\article\text\text.title\pg@pgref specifies the relative page ID for the article title. Clipref specifies the relative page image ID.

CAAL: The number of the page within the physical document.

DEMP: didl:DIDL\didl:Item\didl:Item\didl:Descriptor\didl:Statement contains the page number as a string (e.g. "page 1"). Each item and each component have a descriptor that describes the role of that element in the newspaper issue. The roles are listed in the statement element of the descriptor.

SBME: The "ORDER" field provides the page number in sequence, starting from 1. "ORDERLABEL" provides page numbers based on image file numbers, starting from 0.

Instantiations			
Unique identifier			
B1GL	issue\article\pi	UID	WOJL-1822-01-03-0002
B2GL	issue\article\pi	UID	WOJL-1822-01-03-0002
B2GI	issue\page\article\id	UID	YOHD-1813-07-31-0001-001
CADI	sn\year\month\day\edition\null[sequence]	STR	seq-1
DEMP	didl:DIDL\didl:Item\didl:Item@dc:identifier	UID	ddd:010419500:mpeg21:p001
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx:recordIdentifier	UID	ddd:010419500:mpeg21:p001
F2AL	pageOCRDATA\metadata\pageIdentifier	NUM	7687234
HNDM	$.publication.pagina.publicacion_id	STR	ObjectId("558075bd7d1e63c-9fea1a0eb")
HNDM	$.publication.pagina._id	STR	ObjectId("558a32a47d-1ed64f16885023")
TDAG	issue\page\pageid	UID	0FFO-1812-JUL14-001
TDAG	issue\page\article\pi	UID	0FFO-1812-JUL14-001
TDAG	issue\page\assetID	UID	cs16791278

Relative			
B1GL	issue\article\pi@pgref	NUM	1
B1GL	issue\article\ci@pgref	NUM	1
B1GL	issue\article\text\text.cr\pg@pgref	NUM	2
B2GL	issue\article\pi@pgref	NUM	2
B2GL	issue\article\ci@pgref	NUM	2
B2GL	issue\article\text\text.title\pg@pgref	NUM	2
B2GL	issue\article\text\text.cr\pg@clipref	NUM	1
B2GL	issue\article\text\text.cr\pg@pgref	NUM	2
B1JI	BL_newspaper\BL_page\pageImage\pageSequence	NUM	1
CAAL	alto\Layout\Page@PHYSICAL_IMG_NR	NUM	195
CAME	mets\dmdSec\mdWrap[@LABEL="Page metadata"\xmlData\mods:-mods\mods:part\mods:detail\mods:number	NUM	2
DEAL	alto\Layout\Page@PHYSICAL_IMG_NR	NUM	1
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ddd:nativePageNumber	NUM	1
EUAL	alto\Layout\Page@PHYSICAL_IMG_NR	NUM	1
F1AL	alto\Layout\Page@PHYSICAL_IMG_NR	NUM	1
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page@PHYSICAL_IMG_NR	NUM	
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page@PRINTED_IMG_NR	NUM	204
HNDM	$.publication.pagina.pagina	NUM	1
PPAL	alto\Layout\Page@PHYSICAL_IMG_NR	NUM	1
SBME	mets:mets\mets:structMap[@TYPE="PHYSICAL"]\mets:div\mets:div@ORDER	NUM	6
SBME	mets:mets\mets:structMap[@TYPE="PHYSICAL"]\mets:div\mets:div@ORDERLABEL	NUM	5
TDAG	issue\page\article\pi@pgref	NUM	1
TDAG	issue\page\article\ci@pgref	NUM	1
TDAG	issue\page\article\text\text.cr\pg@pgref	NUM	1
TDAG	issue\page\article\text\text.title\pg@pgref	NUM	1
TRAL	alto\Layout\Page@PHYSICAL_IMG_NR	NUM	1
TRAP	article\title\page	NUM	1
TRAP	article\title\page\pageSequence	STR	1 S

String descriptors			
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Descriptor\didl:Statement	STR	page 1
CAAL	alto\Layout\Page@ID	STR	PAGE.0
DEAL	alto\Layout\Page@ID	UID	P1
EUAL	alto\Layout\Page@ID	UID	P1
F1AL	alto\Layout\Page@ID	STR	P1
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page@ID	STR	P12
PPAL	alto\Layout\Page@ID	STR	P1
TRAL	alto\Layout\Page@ID	UID	PAGE1
TRME	mets:mets\mets:structMap\mets:div\mets:div@ID	UID	divpage1
URL			
DEMP	didl:DIDL\didl:Item\didl:Component\didl_Resource@ref	URL	http://resolver.kb.nl/resolve?urn=d-dd:010419500:mpeg21:pdf
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource@ref	URL	http://resolver.kb.nl/resolve?urn=d-dd:010419500:mpeg21:p001:image
TRAP	article\pdf	URL	http://trove.nla.gov.au/ndp/imageservice/nla.news-page2243325/print
TRME	mets:mets\mets:fileSec\mets:fileGrp\mets:file\mets:FLocat@xlink:href	URL	/pages/nla.news-issn22083111-s75000-b.tif

Edition

Language Variants

Auflage; Editie; Painos; Edición.

Technical Definition

Provides edition information for the **issue**, including morning, afternoon, evening, day, special and **supplemental**. In SBMA and SBME, it also specifies that it is an electronic edition of the issue.

Usage Notes

The term is used to denote geographical variants, to specify size, to indicate several issues of a newspaper published on the same day, and to distinguish between issues published on different days of the week—for example, a regular feature for a weekend edition. The term is not widely used because few collections contain multiple digitised editions. In the Gale collections, the final London edition of a newspaper is always **microfilmed**, but regional editions, such as Irish or Scottish editions, are excluded, as are editions which came out earlier in the day. For the British Library, since 1869 and the implementation of Legal Deposit legislation, the rule has been that a single issue per title is supplied (usually the last edition where relevant). Multiple editions were sometimes submitted by **publishers** for the 1820s–1860s period. Selecting which of these editions to digitise, or all of them, is an issue under discussion with the library's current Heritage Made Digital digitisation programme. The term was used, though not commonly, in the nineteenth century, where "edition" more usually referred to a specific printing of a book rather than a newspaper.

Examples:

As a geographical variant

- "Most were short-lived but the London **edition** of the *Detroit Free Press*, which was composed almost entirely of American humour, sold between 100,000 and 300,000 weekly copies during the 1880s and 1890s." [Nicholson 2012, 285]

To specify size

- "During that time period, at least, the paper was published in eight-page **editions**, with each page containing four columns." [Simons, 389]

Multiple issues on the same day

- "Daily newspapers and some weeklies had multiple **editions** in the nineteenth century, that is successive **editions**, usually with the same date." [DNCJ, LRB, 431]
- "More important in every way was the offer I received from Mr. J. R. Robinson, my former colleague on the *Weekly Chronicle*, who had now become **editor** of the *Express*, the then evening edition of the *Daily News*, to do some work for him…" [Yates, 1.285]

To distinguish between issues on different days

- In 1884 the *Preston Guardian* had cut the price of its Saturday **edition** from 2d to 1½d…" [Hobbs 2018, 45]
- "Produced in a regular rhythm of daily **editions**, containing a familiar staple of home news, foreign news, commercial information and political commentary, the newspaper embodied a Victorian sense of constancy and stability…" [Cronin, 586]
- "The *Caledonian Mercury* goes a step further, interpreting Ainsworth's narrative as being so enmeshed within its context as to warrant the extension of Dickens's child metaphor from *Bentley's* to Jack: 'There is no appearance in the **number** before us of any change in the periodical; all is sparkling and brilliant as in the most favoured days of its childhood and its pains taking attentive nurse. "Jack Sheppard" increases upon our affection as he grows in stature.'" ["Literature." *Caledonian Mercury* 18593 (March 11, 1839): 4, qtd. in Droge, 41]
- "From 'Die Presse', 13. Mai 1905. Professor Dr. R. v. Wettstein is mentioned as the author of an article about the German School Association. A personal opinion is expressed, as in every Saturday **edition** at that time." [Europeana Newspapers 2015, 44]
- "**Edition** statements are often found in the masthead or in the publisher's block." [Sagendorf and Moore, 20]

Category Notes

This field is uncommon and can include a string with the full edition title, a number, a multiple-choice option, or a unique identifier.

Individual Collection Notes

DEMP: The multiple-choice options are Ochtend (morning edition), Middag (afternoon edition), Avond (evening edition) and Dag (daily edition).

TRAP: If this article appeared in a special newspaper/periodical edition, the name of that edition is included here. In the Trove web interface, this is shown on the page view. e.g. http://trove.nla.gov.au/ndp/del/page/4298413.

Instantiations			
CADI	sn\year\month\day\null[edition]	STR	ed-1
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:Identifier\mods:part\mods:detail\mods:number	NUM	1
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcterms:temporal	MCH	Dag
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="digitization"]\mods:edition	STR	[Electronic ed.]
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="digitization"]\mods:edition	STR	[Electronic ed.]
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div\mets:div\mets:div\mets:div@LABEL	STR	Freundschaftsausgabe
SBMY	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:div\mets:div@LABEL	STR	Freundschaftsausgabe
TRAP	article\title\edition	STR	2, SPECIAL EDITION TO THE QUEANBEYAN AGE

Issue Date

Language Variants

Date; Datum; Datum; Päivämäärä; Fecha.

Technical Definition

Gives the **date** of the **issue**. May refer to the publication date, the date as printed on the issue, the ISO standard date or a part of the date, such as the day of the week, day, month or year. In some cases, this is normalised and in others it is the date as printed on the image.

Usage Notes

Within the newspaper layout, the **date** would appear centred underneath the masthead and in the **folio line** (a line at the top or bottom of a newspaper page that gives the newspaper name, **section** and **page number**, and publication date). As headlines were not in use until late in the century, some articles would appear under a date only. This could also indicate an article was reprinted from elsewhere or provide the date of a telegram. Date is a very common search filter across digital collections: Gale indicated that only around 5% of users make use of advanced search functions, but date is used very often to limit searches because researchers tend to know what period they want to look at and are looking for contextual information around a specific time. Nevertheless, the date was rarely the date the **issue** was printed: evening **editions** would obviously be printed earlier in the day, while newspapers published in the morning would be printed the night before.

Examples:

Publication Date

- "If anyone wishes to establish precise publication **dates** for issues of Victorian periodicals, the advertisements in the *Athenaeum* and the *Times* are among the most helpful and accessible pieces of evidence, yet a word of caution is needed." [Woodruff and de Groot, 104]
- "Morning newspapers were **printed the night before**, as were most weekly papers. Most editions of newspapers with 'evening' in their titles were published in the morning or afternoon." [Hobbs 2018, 113]

Date printed on the issue

- "The last issue in the Mitchell Library, Sydney, is marked, 'Vol. xl., No. 4503,' and **dated** 'October 20, 1842.' It was:--'Printed and Published by Richard Sanderson, Proprietor, at *The Sydney Gazette* Office, Lower George-street, New South Wales.' There is no intimation in this number that it is the last nor that it was about to cease publication." [Ferguson, Foster and Green, 69-70]
- "Publishers planned in advance a constructed 'now', a present, that lasted until the next issue, while the **date** at the top of the newspaper, 'the single most important emblem on it' according to Benedict Anderson, was rarely the date on which it was published." [Hobbs 2018, 112-13]

Category Notes

This field is common, but there are various date formats (e.g. month only, day of the week).

Individual Collection Notes

F2AL: Year and month only.

SBME: The separate fields for year, month and date build, so that "year" contains year only, "month" contains year-month, and "day" contains year-month-day.

SBMY: The separate fields for year, month and date build, so that "year" contains year only, "month" contains year-month, and "day" contains year-month-day.

Instantiations			
Publication date			
B1GL	issue\da	DAT	January 03, 1822
B2GL	issue\da	DAT	January 03, 1822
B1GI	issue\metadataInfo\da\composed	DAT	January 2, 1798
B2GI	issue\metadataInfo\da\composed	DAT	July 31, 1813
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ dc:date	DAT	18/04/1945
F2AL	pageOCRDATA\metadata\published	DAT	1881-11
HNDM	$.publication.pagina.fecha	DAT	ISODate("1900-01-01T18:00:00.000Z")
PPME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:originInfo\MODS:dateIssued	DAT	1843-04-22
PPME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:originInfo\MODS:dateIssued	DAT	1843-04-22
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="publication"]\mods:dateIssued	DAT	1934-05-10
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div\mets:div\mets:div\mets:div@ORDERLABEL	DAT	1934-05-10
SBMY	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:div\mets:-div@ORDERLABEL	DAT	1934-05-10
DTAD	issue\da\composed	DAT	July 14, 1812
TRAP	article\title\date	DAT	1876-05-27
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo\mods:dateIssued	DAT	19350706
Printed date			
B1GL	issue\printedDate	DAT	January 03, 1822
B1JI	BL_newspaper\BL_page\issue_metadata\printedDate	DAT	Monday, January 6, 1800
B2GL	issue\printedDate	DAT	January 03, 1822
Standardised date of publication			
B1GI	issue\metadataInfo\da\searchableDateStart	DAT	17980102
B1GL	issue\pf	DAT	18220103
BIJI	BL_newspaper\BL_page\issue_metadata\normalisedDate	DAT	1800.01.06
B2GI	issue\metadataInfo\da\searchableDateStart	DAT	18130731
B2GL	issue\pf	DAT	18220103
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:date	DAT	19240101
TDAG	issue\da\searchableDateStart	DAT	18120714
Day of the week of publication			
B1GI	issue\metadataInfo\dw	MCH	Tuesday
B1GL	issue\dw	MCH	Thursday
B2GI	issue\metadataInfo\dw	MCG	Friday
B2GL	issue\dw	MCH	Thursday

Day of publication			
B1GI	issue\metadataInfo\da\day	DAT	2
B2GI	issue\metadataInfo\da\day	DAT	31
CADI	sn\year\month\null[day]	DAT	4
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div\ mets:div\mets:div@ORDERLABEL	DAT	1934-05-10
SBMY	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:div@ ORDERLABEL	DAT	1934-05-10
TDAG	issue\da\day	DAT	14
Month of publication			
B1GI	issue\metadataInfo\da\month	DAT	January
B2GI	issue\metadataInfo\da\month	DAT	July
CADI	sn\year\null[month]	DAT	8
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div\ mets:div@ORDERLABEL	DAT	1934-05
SBMY	mets:mets\mets:structMap\mets:div\mets:div\mets:div@ORDERLABEL	DAT	1934-05
TDAG	issue\da\month	DAT	July
Year of publication			
B1GI	issue\metadataInfo\da\year	DAT	1798
B2GI	issue\metadataInfo\da\year	DAT	1813
CADI	sn\null[year]	DAT	1915
F1ME	mets\dmdSec\mdWrap\xmlData\MODS:mods\MODS:originInfo\MODS:-dateIssued	DAT	1841
PPDI	Newspaper Data\Papers Past\\[abbreviated newspaper title]\[year]	DAT	
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div@ ORDERLABEL	DAT	1934
SBMA	mets:mets\mets:structMap\mets:div\mets:div@ORDERLABEL	DAT	1934
SBMY	Mets:mets\mets:structMap\mets:div\mets:div@ORDERLABEL	DAT	1934
TDAG	issue\da\year	DAT	1812

Publisher

Language Variants

Herausgeber; Uitgever; Kustantaja; Editor.

Technical Definition

This category contains information about the **publisher** of the **publication**. It does so without date or other restrictions and should be considered to refer to the entire run of the newspaper as defined by category **Normalised Title**.

Usage Notes

The term "publisher" is used inconsistently across contemporary and academic accounts. In general, it refers to the owner, proprietor or principal financial stakeholder of a newspaper. When discussing the nineteenth century, it is usually identified as an individual, a family or named partners as they appear in the newspaper **imprint** (the legal information and address of the newspaper). Owing to the small management and staff sizes of many nineteenth-century newspapers, the term is often conflated, attached or used interchangeably with other recognisable leadership or production roles including **owner**, **proprietor**, **editor**, and **printer**.

Examples:

As a production role

- "newspapers tend to spring from printers and periodicals from **publishers**..." [DNCJ, MdW, 515]
- "Anyone with the price of a newspaper plant and a message to print may become a publisher overnight." [Lenox Lohr, "Broadcasting System", *London Times* (8 June 1939), 59, qtd. in Stamm, 56]
- "Linjan määrittelee siis kustantaja ja sen toteutumista valvoo **kustantajan** nimeämä päätoimittaja."

As a leadership role

- "Robert Chambers [...] was **co-proprietor** of a huge publishing firm" [DNCJ, vi]
- "An Appeal is now made [...] on behalf of Mr. ANDREW BENT, late Newspaper **Proprietor** of Van Diemen's Land" [Bent, 4]
- "For Whitelaw Reid, **owner** and **publisher** of the paper, was also the Republican candidate for the vice-presidency." [Kwiat, 106]
- "He was an **owner**, **editor**, and featured contributor of the *Buffalo Express*." [Branch, 584]
- "In a controversy between him and Hetherington, the **publisher** of the unstamped *Poor Man's Guardian*..." [Vizetelly, 1.74]
- "Editor and **publisher** notes are not generally input in newspaper records. If an editor's or **publisher**'s name is more widely known than the newspaper, a note and added entry may be included." [Sagendorf and Moore, 32]

As a conflated or ambiguous role

- "Edward Baines the elder [...] editor and **publisher** of the *Mercury* from 1801, was a vigorous proponent of the commercial interest..." [Rosengarten, 599]
- "As editors, **publishers**, and journalists in the wake of World War I reconsidered the social significance of their professions..." [Stamm, 63]
- "... many editors and **publishers** combined their work with extensive political involvement." [Djerf-Pierre and Weibull, 301]

Category Notes

This field is relatively uncommon. Information on the publisher for a given date is most likely to be obtained through cross reference with the MARC records of the holding library.

Individual Collection Notes

F2AL: Provides two fields for publisher information, the **original publisher** and the **latest publisher**.

Instantiations			
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:publisher	STR	J.H. Molenbroek
F2AL	pageOCRDATA\metadata\originalPublisher	STR	SLEY-Kirjat Oy
F2AL	pageOCRDATA\metadata\latestPublisher	STR	SLEY-Kirjat Oy
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="publication"]\mods:publisher	STR	Särchen

Place of Publication

Language Variants

Lugar de Publicación; Ort der Veröffentlichung; Plaats van Publicatie; Julkaisupaikka.

Technical Definition

Provides the **geographical location** associated with the printing or manufacture of the publication and generally listed as the city. They are determined by the imprint or cataloguer determination for the **publication** as a whole, except where specified as being the publication location of an **alternate title for this newspaper**.

Usage Notes

Researchers tend to group UK newspapers as **London Press, Provincial Press** or **Metropolitan Press**, based on where the newspaper was printed and where its reach was primarily felt. During the Victorian period, there was a recognised distinction between London/metropolitan and provincial press. In Germany, a typical small city would have one or two newspapers, while Berlin and Leipzig had dozens each; there were initially no national newspapers. Australia retained a similarly regional focus in the nineteenth century: Australia's first newspaper, the *Sydney Gazette and New South Wales Advertiser*, was printed in 1803, with The Australian launching in 1824. By the mid-1830s there were seven newspapers published in New South Wales and five weeklies in South Australia. By the 1850s, Tasmania had eleven newspapers. In nineteenth-century America, newspapers gained wider traction initially through partisan political affiliation, and then through the growth of the **Penny Press** of the 1830s. Many of these papers were referred to as "**city papers**". The US Postal Service Act of 1792 had provided substantial subsidies, meaning that newspapers could be delivered up to one hundred miles for a penny.

Examples:

As a formal part of the printed text

- "It has become the fashion of late years to discard the **imprint** from many jobs, such as programmes, bills of fare, and the like. The omission of it, however, is contrary to law, and renders the printer liable to the penalty above stated." [Powell,18]
- "Our journal [...] from headline to **imprint**, will strive to inculcate thoroughly English sentiments–respect for authority, attachment to the Church, and loyalty to the Queen..." [Wellesley Index 3, 387]
- "Within every newspaper issue some elements can be found which are not directly part of the content but are only included for providing some basic information to the user. These elements are mainly the title section, the **running title** and the **imprint**." [Europeana Newspapers 2015, 13]
- "In addition to naming the newspaper, the masthead may also state the edition, **place of publication**, designation, day of publication, the newspaper's motto or philosophy, and the price." [Sagendorf and Moore, 9]

As a general description of its origin

- "At the lower levels of the **London newspaper press** the position was more obscure. The legitimate cut-price thrice-weeklies, published between post days and circulating entirely in the London area, did not in the end fulfil their apparent growth potential" [Boyce, 87]
- "In some respects the **provincial weeklies** formed a lower stratum of cheap publication in the same way that the cut-price and unstamped papers did in London itself" [Boyce, 88]
- "Dr Cranfield cites the example of Peter Pass, employed to deliver both the *Manchester Mercury* and *Liverpool Advertiser*, who carried the Manchester paper to Bolton, Wigan, Preston, and Kendall a distance of about 80 miles" [Boyce, 90]
- "Kartoitettuani juttujen **julkaisupaikat**, etenin tarkastelemaan juttujen tyylejä."
- "To his great disgust he found the **London press** far less venal than the Parisian had proved..." [Vizetelly, 1.340]
- "Excellent in every department of journalism [...] so vast was his newspaper-reading that he never missed a noticeable point, not merely in **London**, but in the provincial press." [Yates, 1.283]
- "For newspapers, the **place of publication** is always used as the first element in the qualifier. USNP catalogers generally use the 'city, state' form of the place name." [Sagendorf and Moore, 11]

Category Notes

This field is relatively uncommon.

Individual Collection Notes

B1GP: The city listed refers to the place of publication of the **alternate title for this newspaper.**

B2GP: The city listed refers to the place of publication of the **alternate title for this newspaper.**

F2AL: An abbreviation of the country of publication is listed.

HDNM: A country, state and city are listed in separate fields.

HNME: A country, state and city are listed in separate fields.

Instantiations			
B1GP	PubInfo\VariantTitles\City	STR	Aberdeen
BIJI	BL_newspaper\BL_page\title_metadata\placeOfPublication	STR	Aberdeen
B2GP	PubInfo\VariantTitles\City	STR	York
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcterms:spatial	STR	Delft
F2AL	pageOCRDATA\metadata\publishingPlace@country	STR	fi
HNDM	$.publication.ciudad	STR	Ciudad de México
HNDM	$.publication.pagina.pais	STR	México
HNDM	$.publication.pagina.estado	STR	Distrito Federal
HNDM	$.publication.pagina.ciudad	STR	Ciudad de México
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:city	STR	Ciudad de México
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:state	STR	Distrito Federal
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:country	STR	Mexico
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="publication"]\mods:place\ mods:placeTerm	STR	Baruth, Mark
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="distribution"]\mods:place\mods:placeTerm	STR	Baruth, Mark

Bibliographic Metadata

Data in this section may be considered the traditional bibliographical metadata or cataloguing information for the **publication**. It includes information on the **title(s), date ranges, publication frequency, geographical coverage, language** and **genre classifications** of individual **articles** or the **publication** as a whole.

Newspaper Title

Language Variants

Titel der Zeitung; Krantentitel; (Sanoma)Lehden Nimi; Título del Periódico.

Technical Definition

There are two kinds of newspaper title provided: first, the title as it appears on that particular **issue**; second, the title in a **normalised** format. This may or may not be a version of the title as printed but rather an amalgamation chosen by the cataloguer. It is usually derived from the earliest available issue from that newspaper, after which **alternate newspaper titles** will be recorded.

Usage Notes

The term "newspaper title" is consistently applied across contemporary and academic literature. It is sometimes also used metonymically to refer to the newspaper or periodical as a whole. In the nineteenth century, the term "title" was used to refer to the title printed on the masthead for newspapers, magazines and other periodicals as well as for articles in reference to the printed title. Newspaper titles appear in the masthead on the front page (or title page) of a newspaper and in the **folio**, a line at the top of each subsequent page that also includes the date, the page number, and often a **section title**. The title at the top of the page could also be referred to as a **running title**. In the nineteenth century, mastheads could also graphically represent the **geographic coverage** of the publication, definitions of audience and the themes and topics of the publication. In the British Library collections, each title and title change is given a unique ID number, which is included in the metadata for scanned newspapers.

Examples:

Referring to the name of the newspaper

- "A common feature of many Victorian periodicals and newspapers, **mastheads**, along with the **title pages** of volume reissues and part-issue wrappers and covers, sought to establish an instant 'brand image' for the journals they represented." [DNCJ, BM/AGJ, 401]
- "**Masthead**: the **title** of the newspaper at the top of the front page." [Franklin, 201]
- "**Folio**: a line at the top of a page featuring the dateline, page number, the name of the newspaper and often a simple title indicating content, such as 'news', 'comment' or 'sport'." [Franklin, 201]
- "From that point the newspaper's **title**, which in previous issues appeared as it had since its inception, the Watchman, and Jamaica Free Press, ran simply as the *Jamaica Watchman*, dropping 'Free Press.'" [Ward, 2018]
- "The masthead includes the newspaper **title** statement found on the first, or front page. The masthead may also be called the nameplate, flag, or banner [...] In addition to naming the newspaper, the masthead may also state the edition, place of publication, designation, day of publication, the newspaper's motto or philosophy, and the price. If the newspaper lacks a masthead, take the **title** from any source within the issue, and note the source from which the title was taken." [Sagendorf and Moore, 9]
- "The **title** of the paper, *Illustrated London News*, had already been settled [...] the **title** heading to the paper engraved..." [Vizetelly, 1.226-27]
- "Within every newspaper issue some elements can be found which are not directly part of the content but are only included for providing some basic information to the user. These elements are mainly the **title** section, the **running title** and the imprint." [Europeana Newspapers 2015, 13]
- "Sometimes known as the **masthead**, the **titlepiece** is the most fiercely protected element of any newspaper." [Keeble and Reeves, 29]
- "The source of the **title** is the masthead of the first or earliest available issue." [Sagendorf and Moore, 9]
- "Åbo Tidningarin, joksi Suomen ensimmäisen **lehden nimi** lyheni, ..."

Referring to the newspaper or periodical as a whole

- "Happily, more **titles** have since appeared; while the number of pages across these platforms seems prodigious, the number of titles, as Leary noted, is only a tiny proportion of the 'offline penumbra.'" [Brake, 2015a, 249]
- "Some of the newspaper press directories—notably Mitchell's and May's—provided their readers with a huge, spectacular annual 'Newspaper Map of the United Kingdom' showing the density of **titles** in given geographical centres..." [Brake, 2015b, 570]
- "In 1861, one writer estimated that the number of newspapers in England had doubled from 562 to 1,102, although it was acknowledged that many of these **titles** were short lived." [O'Malley, 592]
- "This fluid network of print, in which titles change, merge, discontinue and spawn ever more **titles**,

challenges us to think more critically about the coherence of any single serial title amid the stuttering rhythms of the marketplace." [Turner, 121]

Category Notes

This field appears in all collections (though not all file types) except Chronicling America, which only includes the newspaper title in the issue data. Not all collections identify if the title is the printed title or a normalised one.

Individual Collection Notes

TRAP: Also contains **place of publication** and **date range** information.

PPME: The title associated with the bibliographic metadata for the article also contains the **date**.

Instantiations			
Printed title			
BIJI	BL_newspaper\BL_page\title_metadata\title	STR	The Aberdeen Journal
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of the printed version"]\mets:xmlData\mods:mods\mods:titleInfo\mods:title	STR	Hamburger Nachrichten
F1ME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\titleInfo\MODS:title	STR	Suomi
PPME	mets\dmdSec\mdWrap@LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:titleInfo\MODS:title	STR	Daily Southern Cross
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo\mods:title	STR	Baruther Heimatland
SBMY	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mets:mods\mods:relatedItem[@type=original"]\mods:titleInfo\mods:title	STR	Baruther Heimatland
Normalised title			
B1GL	issue\citation\titlegroup\marcTitle	STR	Berrow's Worcester Journal
BIGP	PubInfo\PublicationTitle	STR	Aberdeen Journal
BIJI	BL_newspaper\BL_page\title_metadata\normalisedTitle	STR	Aberdeen Journal
B2GL	issue\citation\titlegroup\marcTitle	STR	Berrow's Worcester Journal
B2GP	PubInfo\PublicationTitle	STR	The York Herald
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\ srw_dc:dcx\dc:title	STR	Delftsche courant : nieuwsblad voor Delft en Delfland
TRAP	article\title	STR	The Queenslander (Brisbane, Qld. : 1866-1939)

Unspecified			
B1GL	issue\jn	STR	Berrow's Worcester Journal
B2GL	issue\jn	STR	Berrow's Worcester Journal
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of the issue"]\mets:xmlData\mods:mods\mods:titleInfo\mods:title	STR	Hamburger Nachrichten
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of article"]\mets:xmlData\mods:mods\mods:titleInfo\mods:title	STR	Hamburger Nachrichten
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of table"]\mets:xmlData\mods:mods\mods:titleInfo\mods:title	STR	Hamburger Nachrichten
EUME	METS:mets\mets:structMap[@LABEL="Physical Structure" @TYPE="PHYSICAL"]\mets:div@LABEL	STR	Hamburger Nachrichten
EUME	METS:mets\mets:structMap[@LABEL="Logical Structure" @TYPE="LOGICAL"]\mets:div\mets:div\mets:div\mets:div\mets:div@LABEL	STR	Hamburger Nachrichten
F1ME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of the electronic version"]\xmlData\MODS:mods\titleInfo\MODS:title	STR	Suomi
F1ME	mets\structMap[@LABEL="Physical Structure" @TYPE="PHYSICAL"]\div@LABEL	STR	Suomi
F1ME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:titleInfo\MODS:title	STR	Suomi
F2AL	pageOCRDATA\metadata\title	STR	Kirkollinen kuukauslehti : uskonnollista lukemista perheille
HNDM	$.publication.titulo	STR	Actas de Cabildo
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:title	STR	Historia de los Descubrimientos Antiguos y Modernos de la Nueva Espana
PPME	mets\structMap[@LABEL="Physical Structure" @TYPE="PHYSICAL"]\div@LABEL	STR	Southern Cross;New Zealand Herald
PPME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of article"]\xmlData\MODS:mods\MODS:titleInfo\MODS:title	STR	Southern Cross. Saturday, April 22.
PPME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of the electronic version"]\xmlData\MODS:mods\MODS:titleInfo\MODS:title	STR	Daily Southern Cross
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:relatedItem\mods:titleInfo\mods:title	STR	Baruther Heimatland
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div@LABEL	STR	Baruther Heimatland
SBMY	mets:mets\mets:structMap\mets:div@LABEL	STR	Baruther Heimatland
SBMA	mets:mets\mets:structMap\mets:div@LABEL	STR	Baruther Heimatland
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:titleInfo\mods:title	STR	mullewamail

Newspaper Subtitle

Language Variants

Untertitel der Zeitung; Ondertitel van de Krant; Sanomalehden Tekstitys; Subtítulo de Periódico.

Technical Definition

The subtitle, which is intended to provide clarification of the **newspaper title**, may be taken from the physical object, or an amalgamation chosen by the cataloguer.

Usage Notes

It was more common for periodicals in the nineteenth century to have subtitles than newspapers, though the blurring of **publication genres** makes this distinction difficult. The subtitle would be printed on the **masthead** of periodicals, and could provide further information about **geographic coverage**, topics covered, or definitions of **audience**. However, most collections do not contain fields for newspaper subtitles.

Examples:

- "**Mottos**, related titles, or statements of geographic or other intended audience may appear with the title in the masthead, but these statements are generally not considered to be part of the **title** and are not transcribed…" [Sagendorf and Moore, 14]
- "This inherent hybridity was registered in the *Observer's* **subtitle**, *A Record and Review.*" [Hughes, 203]
- "In 1841 *Punch's* **subtitle** 'the London Charivari' had recalled and appropriated a Parisian model at a time when translations of Sue, Dumas and Féval were selling well: *London Journal* likewise preferred to look towards Paris." [King 2017, 67]

Category Notes

This field only appears in SBME and SBMA.

Instantiations			
Printed title			
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo\mods:subTitle	STR	Blätter zur Pflege der Heimatkunde
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:relatedItem\mods:titleInfo\mods:subTitle	STR	Blätter zur Pflege der Heimatkunde

Alternate Newspaper Title

Language Variants

Alternativer Zeitungstitel; Alternatieve Krantentitel; Vaihtoehtoinen Lehden Nimi; Título de Periódico Alternativo.

Technical Definition

Provides an alternate title for the publication, where the title may have changed during its run. Occasionally this is a minor change, such as dropping the article, but this can also be a more radical restyling of the publication. Standardised title information can be found in **newspaper title**.

Usage Notes

The term "alternate newspaper title" encompasses changes to titles over time. Title changes were very common; these changes might take place owing to a change of editor or publisher, reflect an aspirational change in intended audience, indicate a move from daily to weekly or vice versa, or signal an attempt to broaden readership by widening geographical focus. Where a newspaper changes title, the British Library catalogues the new title separately. Records for title changes are linked to the former title and to the continuation title (where this exists). For digitisation, however, they bring these titles together as one set. Each title and title change is given a unique ID number which is included in the metadata for scanned newspapers. In Delpher, when a newspaper title changes, it will be given a new ID in the catalogue. In Trove, the decision whether to digitise the change in title as a new publication mirrors curatorial changes by the National Library of Australia for its physical collection.

Examples:

- "**Title changes** can also reflect the consolidation of a journal within a particular niche market, such as the *London, Provincial, and Colonial Press News: A Literary and Business Journal* which eventually dropped the subtitle as it specialized solely on the press trade." [DNCJ, MBT, 631]
- "Newspaper 'record sets,' or groups of related records, typically involve more titles and corresponding records than those for other serials, due to the proliferation of **title changes** and related editions. Quite often a newspaper cataloger will sketch out the record set to create a diagram of related titles." [Sagendorf and Moore, 26]
- "The paper continued as a weekly in the twentieth century, undergoing a third and last **title change** to *Y Faner* before folding after 135 years." [DNCJ, AGJ, 38]
- "If a **title change** lasts for less than one year, after which it reverts back to its original title, do not consider this to be a major change…" [Sagendorf and Moore, 16]

Category Notes

This field appears most commonly in the adapted Dublin Core format.

Individual Collection Notes

PPME: Multiple alternate titles are listed within the string, separated by semi-colons.

DEMP: The alternate title listed in dcterms:alternative can take one of two forms: either the value starts with "Ook bekend onder de naam" (also known as), or "Eerder verschenen onder de naam" (previously published as). dcterms:isVersionOf is the same as "Ook bekend onder de naam", but in this case only the alternate newspaper title is listed, without the leading text, giving a **normalised title**.

Instantiations			
B1GP	PubInfo\VariantTitles\Title	STR	Aberdeen Weekly Journal
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\name	STR	Aberdeen Weekly Journal
B2GP	PubInfo\VariantTitles\Title	STR	The York Herald, and General Advertiser
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcterms:alternative	STR	Eerder verschenen onder de naam Delflandsche courant : nieuws- en advertentieblad
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcterms:isVersionOf	STR	Delflandsche courant : nieuws- en advertentieblad
PPME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:titleInfo\MODS:title	STR	Southern Cross;New Zealand Herald
PPME	mets\dmdSec\mdWrap[@LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:relatedItem\MODS:titleInfo\MODS:title	STR	Southern Cross; New Zealand Herald

Abbreviated Newspaper Title

Technical Definition

A standardised abbreviation of the **newspaper title** which may also appear in unique IDs for the newspaper and article.

Usage Notes

During the nineteenth century, commentators might drop the article ('the') when referring to newspaper titles but, as several newspapers might have similar titles (e.g. *The Guardian, Preston Guardian, Poor Man's Guardian, The Guardian of Education*), the entire title was generally used – or a standardised shortened version (e.g. *The Times* as compared to *The Times of London*).

Category Notes

This field only appears in the British Library collections and the directory structure of Papers Past.

Individual Collection Notes

B1GL: An ad hoc, unique, four-character abbreviation of the title.

B1JI: An ad hoc, unique, four-character abbreviation of the title.

B2GL: An ad hoc, unique, four-character abbreviation of the title.

Instantiations			
B1GL	issue\titleAbbreviation	STR	WOJL
B1JI	BL_newspaper\BL_page\title_metadata\titleAbbreviation	STR	ANJO
B2GL	issue\titleAbbreviation	STR	WOJL
PPDI	[abbreviated newspaper title]	NUL	

Publication Date Range

Language Variants

Veröffentlichungsdatum; Verschijningsperiode; Julkaisupäivä; Intervalo de Fechas de Publicación.

Technical Definition

Provides the **date range**, in either years or full dates, of the publication. It does so without date or other restrictions and should be considered to refer to the newspaper as defined by ontological subcategory **normalised title**. There are three variants, the dates included in the collection (collection range), those that to our knowledge existed (full range), and all the individual days, months and years of publication.

Newspaper **start date** and **end date** provide the full ISO **date** for the **publication's** first and last **issue**. Instantiations are divided into container elements Instantiations are divided into container elements, which hold no specific data, and attributes or specific elements, which hold the year, month and day separately.

Usage Notes

Globally, the spread of newspapers exploded in the nineteenth century. This also led to many new publications that were short-lived. In the UK, the repeal of stamp duty on newspapers in 1855 saw an increase in the number of provincial newspapers, many of which would quickly fold. In the 1880s, New Journalism and the mass-market approach led to a further explosion of new publications. In Mexico, many newspapers sprang up during the Mexican War of Independence to spread propaganda or gain support but were usually quickly quashed. Most digital collections have aimed to digitise complete runs: for example, Trove evaluates recommendations for **titles**, and then aims for a complete run. Delpher prefers to digitise newspaper titles in a complete run, but sometimes has had to exclude issues because of **copyright**. Additionally, particularly with seventeenth- or eighteenth-century newspapers, they digitise only part of the run because they focus on the periods when the newspaper is most fashionable, most read, and most interesting from a historical point of view. As such, the date range recorded refers to only the selection present in the collection.

Examples:

- "**Between 1856 and 1914** the number of newspapers published in Britain and Ireland increased more than eightfold, from 274 to 2,205, with London numbers tripling from 151 to 478." [Williams, 99]
- "**Between 1855 and 1861**, 137 newspapers were established in 123 English towns where previously there had been no local newspaper. In 1861, one writer estimated that the number of newspapers in England had doubled from 562 to 1,102, although it was acknowledged that many of these titles were short lived. This expansion continued into the twentieth century: in 1856 there were thirty daily and evening London and provincial newspapers in England, but in 1900, there were 203 and in 1914, 153. The total number of national, London, and provincial morning and evening papers rose from 197 in 1920 to 202 in 1922, dropping to 164 in 1931 and down to 126 in 1944." [O'Malley, 592-93]
- "Aineistoa kerättäessä huomioitiin sekä Integrumissa mainittu ilmestymispäivä (lehden painettu versio) että internet-version **julkaisupäivä**"
- "On a sabbatical year in the United Kingdom from 1965 to 1966, he immersed himself in the journalism of the 1860s but was frustrated by the difficulties of ascertaining the location of copies, the **lengths of runs**, as well as changes in title, editorial staff, contributors, and affiliations of these little-known publications." [Shattock 2017a, 3-4]
- "Entire **runs** of newspapers have disappeared, and although we know something of Clemens's early experience and the books he read, no record of his newspaper reading exists." [Branch, 600]

Category Notes

These fields are uncommon; more usually, only the date for the specific issue is given. Start and end dates only appear in B1JI.

Individual Collection Notes

B1GL: Several date ranges can be listed, separated by a semicolon. Refers to the digitised holdings, rather than the newspaper's actual run.

B1JI: Several date ranges can be listed, separated by a semicolon. The start and end date refer to a specific **alternate title for this newspaper** and only to the digitised holdings, rather than the newspaper's actual run.

DEMP: The ddd:yearsDigitized field refers only to years of publication held in the "Databank Digitale Dagbladen" collection.

HNDM: All the individual days, months and years of publication are listed.

Instantiations			
Full publication date range			
B1JI	BL_newspaper\BL_page\title_metadata\datesOfPublication	DAR	1 Jan 1800 - 23 Aug 1876; 30 Aug 1876 - 31 Dec 1900
B1GL	issue\citation\datesOfPublication	DAR	1 Jan 1800 - 27 Aug 1859; 29 Aug 1859 - 21 Feb 1860; 22 Feb 1860 - 20 Apr 1867
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ddd:-yearsDigitized	DAR	1940-1945
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcterms:issued	DAR	1856-2002
HNDM	$.publication.fecha	DAR	"fecha":[{"aa":"1900","mes":[{"mm":"01","dia":["01"]}]}]
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:note[@type="date\sequential designation"]	STR	1932,4(14.Nov.) - 1936,39(26. Okt.); 1939,45(27.Jan.); mehr nicht digitalisiert
Publication start date			
Container			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\startDate	NUL	
Day			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\startDate@day	DAT	30
Month			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\startDate@month	DAT	8
Year			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\startDate@year	DAT	1876
Combined			
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="publication"]\mods:dateIssued[@point="start"]	DAT	1930-01-01
Publication end date			
Container			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\endDate	NUL	
Day			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\endDate@day	DAT	30
Month			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\endDate@month	DAT	8
Year			
B1JI	BL_newspaper\BL_page\title_metadata\changeToTitle\endDate@year	DAT	1876
Combined			
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:originInfo[@eventType="publication"]\mods:dateIssued[@point="end"]	DAT	1945-12-31

Publication Frequency

Language Variants

Veröffentlichungshäufigkeit; Publicatiefrequentie; Julkaisutiheys; Frecuencia de Publicación.

Technical Definition

Specifies the frequency of publication as a whole and should not be confused with the **edition** of a specific issue. Across these databases, it is usually listed as daily, weekly or quarterly.

Usage Notes

Publication frequency is an important term in understanding nineteenth-century newspapers, as it dictated and worked together with length, content and format to establish the character of the publication; for example, newspapers published in multiple editions per day were intended to be consumed more quickly, while in Australia, paper, ink, compositor or printer shortages meant that some newspapers would have to suspend a planned publication. Different editors could be assigned to different editions and the content could be updated depending on the events of the day. Very few digitised collections, however, include these multiple **editions**.

Examples:

- "In the nineteenth century, there were many frequencies – **morning**, **evening**, **fortnightly**, **monthly**, **quarterly** and **annual**." [DNCJ, MWT, 234]
- "... general **weeklies** for the middle classes [...]; monthlies for men [...] and women [...]; illustrated titles [...]; **monthly** class papers [...], and class **weeklies** [...]." [DNCJ, v-vi]
- "In June of 1841 Hume took over the *Daily Calcutta Intelligencer* and *Commercial Advertiser* and repositioned it as the *Calcutta Star*, a daily newspaper..." [Simons, 389]
- "Hume announced that he was starting a new **weekly** newspaper, the *Eastern Star*, with the first number to be published on January 5, 1840." [Simons, 389]
- "The emphasis on **annuals** also obscures the significance of **monthly** periodicals and **weekly** provincial and metropolitan newspapers..." [Easley 2016a, 707]
- "Journalisti-lehden **julkaisutiheys** on vaihdellut tarkasteluun ottamani ajanjakson."
- "Mr Robinson, editor of the **evening edition**..." [Yates, 1.288]
- "**Frequency** is generally used as a qualifier to distinguish titles from the same place and time and is sometimes included as an edition statement (e.g., 'Weekly ed.')." [Sagendorf and Moore, 11]
- "A newspaper's **frequency** is most often found in the publisher's statement or in the masthead." [Sagendorf and Moore, 28]

Category Notes

This is an uncommon field, existing only in the British Library and Hemeroteca Nacional Digital de México collections.

Individual Collection Notes

B1GP: The field refers to the publication frequency of the alternate title for this newspaper.

B2GP: The field refers to the publication frequency of the alternate title for this newspaper.

HNDM: The multiple-choice values are diario (daily), semanal (weekly) and trimestral (quarterly).

B1JI: The subCollection field can, but does not necessarily, provide frequency information, e.g. "Regional Daily".

Instantiations

B1GP	PubInfo\VariantTitles\Frequency	MCH	Daily
B2GP	PubInfo\VariantTitles\Frequency	MCH	Weekly
B1JI	BL_newspaper\BL_page\title_metadata\subCollection	MCH	Regional Daily
HNDM	$.publication.frecuencia	MCH	Diario
HNDM	$.publication.pagina.frecuencia	MCH	Diario

Publication Genre

Language Variants

Género de Publicación; Veröffentlichungsgenre; Publicatie Genre; Julkaisun lajityyppi.

Technical Definition

Specifies the **genre** of the **publication** at the broadest level, i.e. whether a newspaper, periodical or magazine. The British Library Newspaper Programme used the following ISO standard definition of a newspaper to decide which titles formerly held at the Newspaper Library were newspapers: "a serial publication which contains news of current events of special or general interest. The individual parts are listed chronologically or numerically and usually appear at least once a week. Newspapers usually appear without a cover, with a masthead, and are normally larger than 297mm x 420mm in size". Newspapers usually appear without a cover, with a masthead, and are normally larger than 297mm x 420mm in size" [Sagendorf and Moore, 7].

For Delpher, a newspaper must meet the following criteria: it must be a product from the printing press, in multiple identical copies (therefore not handwritten); it must be published **periodically** (in the seventeenth century that normally starts at once a week, but later more frequently to even twice a day in the early twentieth century, and on a fixed day of the week); its content has to have news and has to be for a wide audience; and it must be publicly available ("Selection Criteria").

The Australian Newspaper Plan libraries have adopted the following definition of a newspaper, similarly drawn from the ISO standard definition but with several expansions: "A newspaper is a serial publication which contains news on current events of special or general interest. The individual parts are listed chronologically or numerically and appear **frequently** usually at least once a week but sometimes fortnightly or monthly. Traditionally newspapers are printed on newsprint paper, usually appear without a cover, folded rather than bound together, with a masthead, and are normally larger than 297 mm x 420 mm in size. They may include supplements such as colour magazines, or other inserts for special features or events."

Usage Notes

Genre is often used in academic accounts interchangeably with **type** and **format**, although these terms can also be used to specify sub-divisions such as broadsheet versus quarto. **Genre** can also refer to **periodicity** and **frequency**, and distinct types such as family papers, police gazettes and society papers. The term **genre** was not commonly used in this sense in the nineteenth century.

In Trove, there are some journals that have found their way into the newspaper digitisation project owing to practicalities at the time of digitisation. Publications such as government gazettes, which are technically journals, were digitised with the newspapers because they did not have the process for journals at the time. Similarly, **publications** that changed genre (such as *The Bulletin*, which started as a newspaper but became a magazine) are listed under the first genre for the run.

Examples:

At the broadest level (newspaper, periodical, magazine)

- "... the periodical as a **publishing genre** has sustained a remarkable development since the 18th century." [Beetham, 96]
- "The second example points out the interdependence of various **publishing genres** in the periodical marketplace, emphasizing how this informed the reputations of editors and contributors." [Tilley, 210]
- "A newspaper can change its form and content to become a periodical. Do not make a new record to reflect a change in **format** from newspaper to periodical (or vice versa). Instead, give the information in a note. In the case of a newspaper that has changed to a periodical, leave the Type of Serial (008/21) code as 'n.' Existing subject headings should retain their form subdivisions with a change in format, and additional headings may be input to reflect the new format." [Sagendorf and Moore, 42]
- "Young Ireland responded quickly to political events, and the newspaper **format**, with its cheapness and topicality, was crucial to the success of the movement." [Tilley, 210]

As a sub-category of journalism

- "Although it seems a commonplace, the nineteenth-century press is characterized by a number of distinct **genres** beyond that of the usual distinction between periodicals and newspapers." [DNCJ, LRB, 245]
- "The *Journal* was inserted into the market as a 'cheap illustrated miscellany', and to understand the expectations this aroused in the implied reader, it is necessary to sketch a history of the **genre**." [King 2004, 49]

- "**Journalistic genres** such as news articles, editorial comment, foreign correspondence, political debate, court reports, financial bulletins, illustrations, sports coverage and even advertisements..." [Nicholson 2012, 277]
- "Tiettyä **lajityyppiä** ohjaavat tietyt säännönmukaisuudet."

Category Notes

This field is common in METS files and the British Library adapted Dublin Core files. In METS, it is an optional attribute rather than an element.

Individual Collection Notes

B1GP: In this collection, the genre specified refers to alternate title for this newspaper.

B2GP: In this collection, the genre specified refers to alternate title for this newspaper.

HNDM: The multiple-choice options are periodico and monografia.

TRME: The same field name is used to describe both the file and also related items.

Instantiations			
B1GL	issue\citation\typeOfPublication\	MCH	Newspaper
B1GP	PubInfo\VariantTitles\Format	MCH	Newspaper
B1JI	BL_newspaper\BL_page\title_metadata\typeofPublication	MCH	Newspaper
B2GL	issue\citation\typeOfPublication	MCH	Newspaper
B2GP	PubInfo\VariantTitles\Format	MCH	Newspaper
EUME	METS:mets@TYPE	MCH	Newspaper
EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods\ mods:genre	STR	newspaper issue
EUME	METS:mets\mets:structMap[@LABEL="Physical Structure" @TYPE="-PHYSICAL"]\mets:div@TYPE	MCH	Newspaper
EUME	METS:mets\mets:structMap[@LABEL="Logical Structure" @TYPE="LOG-ICAL"]\mets:div@TYPE	MCH	Newspaper
F2AL	pageOCRDATA\metadata\contentType	STR	serial
HNDM	$.publication.tipoPublicacion	MCH	Monografía
HNDM	$.publication.pagina.tipoPublicacion	MCH	Monografía
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:category	MCH	Monografias
PPME	mets@TYPE	MCH	Newspaper
PPME	mets\structMap[@LABEL="Logical Structure" @TYPE="LOGICAL"]\div@ TYPE	MCH	Newspaper
PPME	mets\structMap\div@TYPE	MCH	Newspaper
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\ mods:classification	MCH	Zeitungen
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div@TYPE	MCH	newspaper
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\ mods:classification	MCH	Zeitungen
SBMA	mets:mets\mets:structMap\mets:div@TYPE	MCH	newspaper

SBMY	mets:mets\mets:structMap\mets:div@TYPE	MCH	newspaper
TDAG	issue@contentType	MCH	Newspaper
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:genre	STR	newspaper issue
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:genre	MCH	newspaper

Geographic Coverage

Language Variants

Geographische Abdeckung; Verspreidingsgebied; Maantieteellinen Kattavuus; Cobertura Geografica.

Technical Definition

Classifies newspapers depending on their wider **geographic area** of **publication** and readership; it is listed as regional, local, or a specific territory. If not indicated, it can often be presumed to coincide with **place of publication**. It can be used to distinguish between different **editions** of the same paper aimed at different cities, towns or regions.

Usage Notes

Researchers of the British and Australian press tend to group newspapers as Rural, Provincial or Metropolitan Press, based on its **place of publication** and where its reach was primarily felt. The Delpher collection includes newspapers that were published in the Netherlands or parts of the world that once belonged to the Kingdom of the Netherlands, which includes the former colonies, the Dutch Indies, Suriname and the Caribbean. There are exceptions, including newspapers of Dutch migrants to the United States in the nineteenth and twentieth century, partly in Dutch and later in English. The Trove collection began with one title from each state and territory to start the archive, in order to be representative. These were often the oldest or biggest newspaper from each state. They are also in the process of expanding their collection beyond Australia, digitising newspapers from countries including Indonesia, Papua New Guinea, and the Pacific islands.

Examples:

- "The letter campaign of 1894 was just the beginning of a life spent agitating for representation and equality for poor women. Nield's career as an author began by writing anonymous letters to the **provincial press**..." [Bunting, 142]
- "Lady Greville likely did not attend the ball personally. Except for the by-line on her own article, her name is not mentioned in any of the (approximately 200) articles about the ball found so far, not even in the stories from newspapers like the ones that later published her obituary, with **regional** connections to her and the rest of her family." [Cogdill, 185]
- "The emphasis on annuals also obscures the significance of monthly periodicals and weekly **provincial** and **metropolitan** newspapers in the history of nineteenth-century women's poetry." [Easley 2016a, 707]
- "The *Chronicle*, like many **town papers** of the era, brought national issues and trends to **regional** readers and carried general reports on progressive women's issues of the day." [Bunting, 143]
- "Excellent in every department of journalism [...] so vast was his newspaper-reading that he never missed a noticeable point, not merely in **London**, but in the **provincial press.**" [Yates, 1.283]
- "A **geographical edition** is one of multiple publications issued at the same time by the same publisher and usually having the same title. The contents of the editions generally vary to include news from different cities, towns, or regions." [Sagendorf and Moore, 6]

Category Notes

This is a rare field and is only available in B1JI and DEMP.

Individual Collection Notes

B1JI: The subCollection field can, but does not necessarily, provide geographical information, e.g. "Regional Daily".

DEMP: Multiple choice options include Landelijk (national newspaper title), Regionaal/lokaal (local newspaper title), Nederlandse Antillen, Suriname, and Nederlands-Indië / Indonesië.

Instantiations

B1JI	BL_newspaper\BL_page\title_metadata\subCollection	MCH	Regional Daily
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcterms:spatial	MCH	Regional/lokaal

Document Type

Language Variants

Dokumentenart; Documenttype; Dokumentti Tyyppi; Tipo de Documento.

Technical Definition

Specifies the nature of the piece of writing, generally "article".

Usage Notes

In the nineteenth century, the word "**article**" was commonly used to describe each individual piece of writing in a newspaper. As well as articles there might be poetry, fiction, advertisements, illustrations, announcements, errata, letters and obituaries; in periodical and historical research, some of these other categories are also occasionally referred to interchangeably as articles. "**Item**" and "**piece**" are also used as synonyms, the former being more popular in mid-twentieth century research than today. Modern readers should be mindful of selective preservation that has sometimes meant that **advertisements** and other **supplementary** matter are missing from **issues**, having been ripped out or removed. In the context of Library and Information Sciences, a document is simply a record that contains information content. It could include books, manuscripts, articles, audio-visual material, computer files and databases. See **article category** for further subdivisions.

Examples:

- "Under her usual byline, Greville's **article** for *The Graphic* describing this event competed with more substantial articles in London's *Morning Post, Times,* and *Daily News*; her article nonetheless reveals her importance as a writer, at least in terms of what she called fashionable Society (*Gentlewoman* 108)." [Cogdill, 184-85]
- "The **article**'s tone, especially in its opening passages, is not unlike that of many early-to-mid-century newspaper articles that detail accidental deaths, for its text lingers not only on the deep sadness of the event (its melancholic, lamentable, mournful, grief-filled, and awful qualities) but also on its potential to teach readers a lesson." [Fieldberg, 14]
- "In December 1881, roughly thirty years after the foundation of the Hospital for Sick Children in Great Ormond Street, the *Daily Telegraph* published a **lengthy** article about the institution." [Boehm, 154]
- "The attack on Cartwright's mill at Rawfolds, which took place on the night of April 11, 1812, is described at length in the *Mercury* of the following Saturday; almost all the **article** is quoted by Wroot, and need not be repeated here." [Rosengarten, 594]
- "Consider the following newspaper **items** dating from the eight years between the summer of 1853 and the summer of 1861." [Branch, 576]
- "Though these **items** are probably the source of the references in the novel…" [Rosengarten, 593]
- "In March 1819, the *Alexandria Herald* published the following anonymous **piece**…" [Gelmi, 151]
- "Between 1851 and 1861 he contributed several dozen **pieces** of varying kinds to newspapers." [Branch, 583]
- "'A lady who has much time on her hands' would read and re-read the morning paper throughout the day, according to an **article** in *The Journalist and Newspaper Proprietor of 1900*." [Hobbs 2018, 55]
- "It so fortuned that I was as innocent of writing the **article** in question as I am of having murdered Eliza Grimwood, set the Thames on fire, or eaten the puppy pie under Marlow Bridge." [Sala 1895, 1.360-61]

Category Notes

This field appears in only three databases and always as a multiple-choice option.

Individual Collection Notes

DEMP: Value is always "Text" for elements in this collection.

PPAL: A string to identify the type of composed block (e.g. table, advertisement, …).

Instantiations			
B1GI	issue\page\article@type	MCH	Article
B2GI	issue\page\article@type	MCH	Article
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:type	STR	Text
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:type	STR	Text
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\srw_dc:dcx\dc:-type	STR	Text
EUME	METS:mets\mets:structMap\mets:div\mets:div\mets:div\mets:div\mets:-div@TYPE	STR	ARTICLE
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@TYPE	STR	Advertisement
PPME	mets\structMap\div\div\div\div\div\div\div@TYPE	MCH	Advertisement
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div\mets:div\mets:div\mets:div\mets:div\mets:div@TYPE	MCH	article
TDAG	issue\page\article@type	MCH	Article
TRME	mets:mets\mods:dmdSec\mods:mdWrap\mods:xmlData\mods:mods\mods:genre	MCH	Article
TRME	mets:mets\mets:structMap[@type="Logical"]\mets:div\mets:div@TYPE	MCH	Article

Article Category

Language Variants

Artikelkategorie; Artikelcategorie; Artikkeliluokka; Categoría de Artículo.

Technical Definition

Specifies the genre of the article, such as Advertisements or News Article.

Usage Notes

There are a wide range of different **categories of article**, broken down into subject matter and format. The use of fixed **headings** from **issue** to issue for topics such as provincial news, reviews, advertisements and notices means that these are the most commonly used categories in the databases. However, journalism and media history provide a much more detailed list of categories of journalism that owes much to nineteenth-century newspaper press directories, which first appeared in a sustained way in the UK in the 1840s. *Mitchell's Newspaper Press Directory*, first published in 1846 (becoming an annual publication in 1856), eventually included an index of topics. Palmer's *Index to The Times*, covering the period from 1790 to 1905, provided a systematic index of the topics in that newspaper in which articles can be located by broad subject headings, **date of publication**, or **title** keywords.

The Gale collections have used the standard section headings form *The Times* as a model for metadata categorisation of other papers because of its relatively clear structure. In Delpher, section headings are generated semi-automatically and checked by an operator.

Examples:

- "A listing of the holdings of the National Library of India at Calcutta notes that the *Calcutta Star* contained: **Advertisements, Notices, Domestic occurrences, Commercial Intelligence, Shipping Intelligence, Bank shares, Price of Bullion, Rates of interest and discount, Literary articles, Sporting intelligence, Original correspondence, Editorial paragraphs, Orders of the Governor General in Council, European intelligence with special reference to England, House of Commons reports, Parliamentary miscellanea, Precis of miscellaneous events, Europe–births, marriages and deaths**. ('South Asian Library and Research Notes' 134–35)" [Simons, 389]
- "Journalistic genres such as **news articles, editorial comment, foreign correspondence, political debate, court reports, financial bulletins, illustrations, sports coverage and even advertisements**." [Nicholson 2012, 277]
- "These authors wrote about **Society, social gossip**, and such women's issues as **suffrage, temperance, poverty, married women's property, health**, and **athleticism**, as well as **fashion, marriage, travel, fine art,** and **photography**." [Cogdill, 176]
- "On that day, the reader could find above the *feuilleton*, the 'premier-Paris,' an **editorial column** addressing the question of the Orient, as well as a collection of **news stories** about French foreign policies, **readers' letters, news copied from foreign sources**, a series of **local crime reports, miscellaneous news items**, a mix of **advertisements for shows and bookstores, a list of those appointed to public office**, a summary of the **stock market activity**, a **list of entertainments**, and a series of **other advertisements**—all spread over four pages." [Thérenty, 35]
- "In a similar way as the five main classes of newspaper content, UIBK specifies text types or **genres of news** which appear in nearly every newspaper. Some typical examples are the '**Editorial**', a (subjective) statement of the publisher, the editor in chief with regard to a recent event or the political, economic or cultural situation in general. Another example are **book reviews**, which were first introduced in the early 19th century and are until today one of the classical articles in the 'Feuilleton' section of a newspaper. Other examples are **Death notices**, or **Job** and **Real Estate offers**." [Europeana Newspapers, 28]
- "artikkelin perustiedot ja artikkeliluokat"

Category Notes

This field appears in several databases, always as a multiple-choice option; Dublin Core guidance suggests using a controlled vocabulary of keywords.

Individual Collection Notes

DEMP: Multiple choice options are Familiebericht (family notice), Advertentie (advertisement or small notice), Illustratie met onderschrift (illustration which is not part of an article, e.g. a cartoon), and Artikel (regular article).

TRAP: Multiple choice options are: Article; Advertising; Detailed lists, results, guides; Family Notices; Literature.

Instantiations			
B1GI	issue\page\article\ct	MCH	Classified
B1GL	issue\article\ct	MCH	Advertisements & Notices
B2GI	issue\page\article\ct	MCH	Classified ads
B2GL	issue\article\ct	MCH	News
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:subject	MCH	Artikel
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\srw_dc:dcx\dc:subject	MCH	Artikel
TDAG	issue\page\article\ct	MCH	Classified ads
TRAP	article\category	MCH	Article
TRME	mets:mets\mods:dmdSec\mods:mdWrap\mods:xmlData\mods:mods\mods:genre[@TYPE="articleCategory"]	MCH	Article

Language

Language Variants

Sprache; Taal; Kieli; Idioma.

Technical Definition

Specifies the language of the textual unit, often, but not always, using the ISO language code. This can refer to the language of the newspaper, the article text, or the specific block of text.

Usage Notes

The **language** in which newspapers were published was shaped by various factors including the culture of reprinting, which meant that articles published in Europe were often simply reproduced for a colonial audience, and vice versa. This practice, along with class-based language preferences and the efficiency of reusing text without translation, tended to reinforce the centrality of particular dialects or languages within multi-lingual states. For example, though the first newspaper to be published in Wales appeared in 1804 (*The Cambrian*), the first Welsh language weekly was not published until 1814 (*Seren Gomer,* founded by Joseph Harris). Similarly, in Finland, all three of the oldest newspapers still retain their original Swedish titles, while the oldest daily newspaper in Finland is the Swedish language publication *Åbo Underrättelser*, founded 1824. In colonial spaces, language homogeneity was likewise encouraged by the dominance or early appearance of government gazettes. As the provincial and colonial press expanded, so did the use of dialect and different languages, including the printing of non-European language newspapers within European colonies. Nor were newspapers necessarily monolingual; Spanish, German and Italian newspapers in Australia and the United States included English-language articles and multilingual communities, such as New Orleans, support bilingual publications. Editions of a newspaper in other languages are often considered separate publications in databases because they will be published on separate **dates**.

English has become the *lingua franca* for XML, and all databases other than HNDM use English to categorise all their metadata.

In regard to a sense of **style** and word choice, each newspaper attempted to maintain a consistent tone throughout its pages. This is exemplified by the uneven move from anonymity to signature in the nineteenth century; established daily newspapers were some of the last to adopt signed articles. Debates about the status of newspapers as literature shaped the kind of language employed in the press, depending on how each publication positioned itself.

Examples:

- "… the **dialect speech** of middle-class speakers was more likely to be transcribed into 'Queen's English', apart from the heckles of the working-class crowd. Vernon sees this as a sign of the closer association between Standard English and power, and the growing power of print over oral media." [Hobbs 2018, 309]
- "Place trumped class in the changing discourse of local newspapers, as a theoretically neutral technique—**dialect**—was used to include rather than exclude. This was part of complex changes in the relationship between spoken and written language, and the crossover between literary and journalistic techniques." [Hobbs 2018, 302]
- "Because of the timely nature of newspapers, very few newspapers are considered to be **translations**." [Sagendorf and Moore, 40]

Category Notes

This field appears in all databases.

Individual Collection Notes

B1GI: issue\dcLanguage specifies the language of the **issue** (given in abbreviated form), issue\metadataInfo\language specifies the language of the **newspaper title**, issue\metadataInfo\language@ocr specifies the OCR language of the newspaper, and issue\page\article\ocrLanguage specifies the OCR language of the article.

B1GP: The field refers to the language of the **alternate title for this newspaper**.

B2GL: Specifies the language of the **issue** (given in abbreviated form)

B2GI: issue\metadataInfo\language specifies the language of the **newspaper title**, issue\metadataInfo\language@ocr specifies the OCR language of the newspaper, issue\page\article\ocrLanguage specifies the OCR language of the article.

B2GP: The field refers to the language of the **alternate title for this newspaper**.

CAAL: Declares the language of the textblock. On a practical level, this declaration applies to the contents of the @content attribute on the <STRING> element.

DEMP: Provides a UID for the language.

EUME: METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods\language\languageTerm refers to the language of the printed version. All instantiations of @xml:lang provide an ISO code for the language.

F1ME: Fields give an ISO code for the language.

HNDM: Language of the country of publication. See collection "catLanguage" to see options. Idioma del país de la Publicación. Ver colección "catIdioma" para conocer las opciones.

PPME: Fields give an ISO code for the language.

TDAG: issue\metadatainfo\language provides the language for the **issue**, issue\metadatainfo\language@ocr specifies the OCR language for the issue, and issue\page\article\ocrLanguage provides the OCR language for the article.

TRME: Fields give an ISO code for the language.

Instantiations			
Language of the text block			
CAAL	alto\Layout\Page\PrintSpace\TextBlock@language	UID	en
Language of the article			
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of article"]\mets:xmlData\mods:mods\mods:titleInfo@xml:lang	UID	de
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of article"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	en
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of article"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	en
TDAG	issue\page\article\ocrLanguage	STR	English
Language of the section			
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of section"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	en
Language of the page			
F2AL	pageOCRDATA\metadata\language	STR	fin
HNDM	$.publication.pagina.idioma	MCH	Español
Language of the issue			
B1GL	issue\dcLanguage	STR	eng
B1GL	issue\metadataInfo\language	STR	English
B1GI	issue\metadataInfo\language@ocr	STR	English
B1GI	issue\page\article\ocrLanguage	STR	English
B2GL	issue\dcLanguage	STR	eng
B2GI	issue\metadataInfo\language	STR	English
B2GI	issue\metadataInfo\language@ocr	STR	English
B2GI	issue\page\article\ocrLanguage	STR	English

DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:language@	xsi:-type	dcterms:ISO639-1
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of the issue"]\mets:xmlData\mods:mods\mods:titleInfo@xml:lang	UID	de
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	sv
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	sv
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the electronic version"]\xmlData\MODS:mods\titleInfo@xml:lang	UID	sv
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the electronic version"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	sv
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	sv
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:language	STR	SPANISH
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	en
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	en
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	en
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the printed version"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	en
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the electronic version"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	en
PPME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of the electronic version"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	en
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:language\mods:languageTerm	UID	ger
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:language\mods:languageTerm	UID	ger
TDAG	issue\metadatainfo\language	STR	English
TDAG	issue\metadatainfo\language@ocr	STR	English
Language of the publication as a whole			
B1GP	PubInfo\VariantTitles\language	STR	English
B2GP	PubInfo\VariantTitles\Language	STR	English
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:language	UID	n1

EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods\language\languageTerm	STR	German
EUME	METS:mets\mets:dmdSec\mets:mdWrap[@LABEL="Bibliographic meta-data of table"]\mets:xmlData\mods:mods\mods:titleInfo@xml:lang	UID	de
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of Issue"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	sv
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of list"]\xmlData\MODS:mods\MODS:titleInfo@xml:lang	UID	sv
F1ME	mets\dmdSec\mdWrap[LABEL="Bibliographic meta-data of list"]\xmlData\MODS:mods\MODS:language\MODS:languageTerm	UID	fi
HNDM	$.publication.idioma	MCH	Español
TRME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:language\mods:languageTerm	UID	en

Illustration Information

Technical Definition

Provides information about any illustrations, including whether one is present, its type, and its colour information.

Usage Notes

1806 saw the first illustration ever published in *The Times*, and improved technology and reduced taxation made illustrations a central feature of the nineteenth-century press. Illustrations became central to new journalistic genres such as investigative journalism, though **photography** did not become a central part of newspapers in the UK until 1904, with the *Daily Mirror*. Illustration technology was most significantly used in advertising around the world, with improvements in technology ensuring that illustration played a greater role in the press globally in the second half of the nineteenth century.

Examples:

- "Drawing on emergent reprographic media, initially wood engraving (which allowed text and **illustration** to be printed together) and, to a more limited extent, lithography and then a variety of photo-reprographic techniques, **illustration** was a central attribute of many nineteenth-century periodicals…" [DNCJ, BM/AGJ, 304]
- "*Graphic* was published on Saturdays, and 'Place aux Dames' appeared every week, typically in the middle of the newspaper and on one page; often, her paragraphs make up the top and bottom third, with **illustrations** of some recent event, sometimes a sporting event, comprising the middle third; sometimes her paragraphs take up the left two-thirds of the page, with unrelated **illustrations** on the right." [Cogdill, 182-83]
- "An assignment to do a double-page spread, with **illustrations**, for the Sunday issue of the *Globe* brought him into close relationship with the art department." [Kwiat, 112]
- "The ActivePaper Archive, created by Olive Software, used an image processing technique called 'segmentation,' which breaks each page of newspaper text down into its smaller information units (articles, **pictures**, advertisements)." [Edmund King, 175]

Category Notes

These fields are relatively uncommon.

Individual Collection Notes

B1GI: Colour information multiple choice options are: bitonal, grayscale or colour. Multiple choice options for type: "Cartoon", "Chart", "Coat_of_Arms", "Diagram", "Drawing", "Engraving", "Genealogical_table," "Graph", "Map", "Musical_work", "Painting", "Photograph", "Plan", "Portrait", "Seal", "Table", "Image".

Instantiations			
Indicating presence of illustration			
B1GL	issue\article\il@indicator	BOO	no
TRAP	article\illustrated	BOO	y
Illustration type			
B1GI	issue\page\article\il@type	MCH	image
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Illustration@TYPE	MCH	
Colour information			
B1GI	issue\page\article\il@colorimage	MCH	bitonal
Illustration ID			
SBAT	PcGts\Page\GraphicRegion@id	UID	r25

Holdings Metadata

Data in this section refers to how the various instantiations of the items are held, in physical, microfilm and digital collections. This includes information on **holding library**, **shelf mark**, **microfilm reel**, **database**, **sub-collection** and **filename** as well as information on **microfilm quality** and **copyright status**.

Copyright

Language Variants

Urheberrecht; Tekijänoikeus; Derechos de Autor.

Technical Definition

Specifies the copyright holder of the **issue**. **Access conditions** provides additional information about the status of a physical object, i.e. any restrictions on access.

Usage Notes

Copyright in the nineteenth century was a controversial topic. In the UK, the 1842 Copyright Act, which focused on copyright for books, did not mention newspapers directly; as such, the reprinting of articles from other newspapers continued unabated. This was also complicated by the 1842 Act, binding copyright to the author, when the majority of newspapers published articles anonymously. An 1881 judgement classed newspapers as books for copyright purposes, but there was still need for further rulings to specify which **article categories** were protected. In America, there was no copyright protection for foreign authors until the 1891 Chace Act. This meant that the reprinting of articles remained an integral feature of nineteenth-century American journalism. The Berne Convention of 1886, signed initially by Belgium, France, Germany, Haiti, Italy, Liberia, Spain, Switzerland, Tunisia, and the United Kingdom, attempted to coordinate copyright protections at an international level, although discussion about its applicability to newspaper content continued into the twentieth century. It also ended the pirating of works published in the British Empire and Commonwealth (Australia had created its own copyright protections, but this did not protect Australian authors from being reprinted in Britain until the Berne Convention and subsequent British legislation). The US did not join until 1988.

With regard to modern copyright of the digital images and texts, all newspaper materials in Delpher have been cleared for third-party copyright claims and released into the public domain, while Trove and Papers Past have done so with much of their collections; any data from in-copyright works are clearly identified on the web interface and in API results. Although all nineteenth-century newspapers would now technically be in the public domain in the UK, the British Library retains copyright for the digital files because of the skill, judgement and labour used in the creation of the digital image. Each database history contains additional details about copyright within their collection.

Examples:

- "The extent of **copyright** in news material itself was still at issue in the 1890s, when there were important decisions on the protection of press agency information." [DNCJ, GL, 143]
- "The 1842 Copyright Act protected **copyright** in 'books,' defined in a broad manner. However, section 18 contained special rules applying to the following genres: encyclopaedias, reviews, magazines, periodical works, or works published in a series of parts." [Cooper, 662]
- "The count hinted that he expected a trifle of £10,000 for his **copyright**, but Cole's friend, Minton, did not quite see this, and proposed a royalty upon every copy sold." [Vizetelly, 1.393]
- "**Tekijänoikeus** jaetaan moraalisiin ja taloudellisiin oikeuksiin."

Category Notes

The "copyright" field appears in the adapted Dublin Core file structures, always as a string. Access condition information is only included in the SBB collections. One field gives a string that describes its status, and a second field provides a persistent uniform resource locator (PURL) link to explain the given status.

Individual Collection Notes

DEMP: There are different fields for the copyright holder for the **newspaper title** (didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:rights) and the digital record (DEMP::didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx:recordRights).

Instantiations			
Copyright holder			
B1GI	issue\metadataInfo\sourceLibrary\copyrightStatement	STR	Copyright © The British Library Board
B1GL	issue\cp	STR	Copyright © The British Library Board
B2GL	issue\cp	STR	Copyright © The British Library Board
B2GI	issue\metadataInfo\sourceLibrary\copyrightStatement	STR	Copyright © The British Library Board
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:rights	STR	AD NieuwsMedia B.V.
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx-:recordRights	STR	Koninklijke Bibliotheek Den Haag
SBME	mets:mets\mets:amdSec\mets:rightsMD\mets:mdWrap\mets:xml-Data\dv:rights\dv:owner	STR	Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
SBMA	mets:mets\mets:amdSec\mets:rightsMD\mets:mdWrap\mets:xml-Data\dv:rights\dv:owner	STR	Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
SBMY	mets:mets\mets:amdSec\mets:rightsMD\mets:mdWrap\mets:xml-Data\dv:rights\dv:owner	STR	Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
TDAG	issue\sourceLibrary\copyrightStatement	STR	© Times Newspapers Limited
Access conditions			
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:accessCondition	STR	open access
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:accessCondition@xlink:href	URL	http://purl.org/coar/access_right/c_abf2
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:accessCondition@displayLabel	STR	open access
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:accessCondition@xlink:href	URL	http://purl.org/coar/access_right/c_abf2
SBMY	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mets:mods\mods:accessCondition	STR	open access
SBMY	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mets:mods\mods:accessCondition@xlink:href	URL	http://purl.org/coar/access_right/c_abf2

Holding Library

Language Variants

Bibliothek; Bibliotheek; Kokoelman Omistava Kirjasto.

Technical Definition

Provides the name and details of the library or archive that contained the digitised material at the time of digitisation. For some databases, it is separated into library name and library location.

Usage Notes

The commercial digitiser Gale works with newspaper publishers directly, including *The Times, The Independent* and *The Mirror*. However, where the microfilm or the originals are not of sufficient quality or there are gaps, they source missing editions from the British Library. Each database history contains additional details about the libraries consulted for their collection.

Examples:

- "Even by 1889, men outnumbered women in the segregated news rooms of Barrow's public **library** by a factor of fifteen to one, 'a fair average of the number of persons who enter the rooms daily, for the purpose of reading the Newspapers and Periodicals'." [Hobbs 2018, 55]
- "According to the 1881 *Macmillan's* article, the principal purpose of the free **library** is 'employing the leisure time of the working classes in a more rational way, and weaning them from the degrading haunts of drink and vice through its newspapers,' and thus through newspapers and books 'carrying on the education of the coming race'." [Bernstein, n.pag.]
- "The **Library**'s digital collection is heterogeneous, comprising content such as eBooks, eJournals and the UK Web Archive, digitised newspapers and manuscripts, digital maps and digital sheet music, electoral registers, patents, and personal digital archives." [The British Library, 3]

Category Notes

This field is relatively uncommon despite most collections consulting multiple physical collections.

Individual Collection Notes

Standardised across all collections.

Instantiations			
Library name			
B1GL	issue\sourceLibrary	STR	British Library
B1GI	issue\metadataInfo\sourceLibrary\libraryName	STR	British Library
B2GL	issue\sourceLibrary	STR	British Library
B2GI	issue\metadataInfo\sourceLibrary\libraryName	STR	British Library
HNME	METS:mets\METS:dmdSec\METS:mdWrap\METS:xmlData\DC:collection	STR	Hemeroteca
TDAG	issue\sourceLibrary\libraryName	STR	Times Newspapers Limited
TRME	mets:mets\mets:amdSec\mets:techMD\mets:mdWrap\mets:xmlData\premis:object\premis:objectIdentifier\premis:objectIdentifierType	STR	National Library of Australia
TRME	mets:mets\mets:amdSec\premis:object\premis:objectIdentifier\premis:objectIdentifierType	STR	National Library of Australia

Library location			
B1GI	issue\metadataInfo\sourceLibrary\libraryLocation	STR	London, United Kingdom
B2GI	issue\metadataInfo\sourceLibrary\libraryLocation	STR	London, United Kingdom
TDAG	issue\sourceLibrary\libraryLocation	STR	London, United Kingdom
Library name and location			
CAME	mets\metsHdr\agent\name	STR	Library of Congress, Washington, DC
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:location\mods:physicalLocation@displayLabel	STR	Library of Congress, Washington, DC
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:location\mods:physicalLocation@displayLabel	STR	Library of Congress, Washington, DC
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:location\mods:physicalLocation@displayLabel	STR	Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, Berlin, Germany
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:location\mods:physicalLocation@displayLabel	STR	Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, Berlin, Germany

Shelf Mark

Language Variants

Signatur; Signatuur; Nimekkeen; Numero.

Technical Definition

This category contains information linking the **publication**, as a conceptual unit, to an item or record in an external database or catalogue. It does so without specific **date** or other restrictions and should be considered to refer to a specific physical volume rather than a **volume** as numbered by the original publisher.

Usage Notes

This is a common term within library science, complicated only by the fact that there is not a one-to-one relationship between the physical object and the digital one in the case of newspapers. Because most digitised newspapers come from volumes of several issues bound together, and the majority of **libraries** do not disbind them to scan (except, notably, the British Library), the shelf mark should be taken to refer to the physical volume from which the newspaper issue has come, if not the full run as defined by **normalised newspaper title**. These volumes can be dedicated to one specific newspaper, or issues from a specific year taken from a range of newspapers.

Examples:

- "Absolute and unchanging fixed locations are normally only found in older libraries and collections that are kept in their original rooms or buildings. This is known as a **shelf mark** or **class mark**, the latter term reflecting the fact that CLASSIFICATION numbers (sometimes abbreviated in the case of complex FACETED CLASSIFICATIONS are often used to derive the location mark. Fixed location is normally associated with CLOSED-ACCESS libraries." [Feather and Sturges, 202]
- "banden van nieuwe **signature** voorzien en ondergebracht ..."

Category Notes

This field only appears in DEMP and SBME.

Instantiations			
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:source	UID	KB C 22
SBME	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:location\mods:shelfLocator	UID	Ztg 1424

Microfilm Reel

Language Variants

Microfilm Haspel; Mikrofilmrolle; Mikrofilmikela; Carrete de Microfilm.

Technical Definition

Most often provides a unique 4-digit ID to the microfilm reel used in the creation of the image associated with this XML. This does not translate to a MARC or library-based record number.

Usage Notes

Many collections are digitised entirely from microfilm, at least initially: for example, *The Times* Digital Archive was digitised from microfilm up to 1985 but now *The Times'* digital PDF copy of each issue can be used. For the British Library collections, from the early 1940s to 2010 the usual method of preserving newspapers was microfilming and approximately 30% of the newspaper collection was microfilmed during this time. Delpher's policy is to digitise from paper, unless microfilm of an acceptable quality is available. Trove has begun to digitise from physical copies, particularly for titles that have not previously been microfilmed. Microfilm digitisation has led to the majority of newspapers being digitised in greyscale.

Category Notes

This metadata field is uncommon, despite the general use of microfilm as a digitisation intermediary.

Instantiations			
B1GL	issue\reelID	NUM	2359
B1JI	BL_newspaper\BL_page\issue_metadata\reelID	NUM	3633
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:identifier[@type="reel number"]	UID	211102366
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:identifier[@type="reel sequence number"]	UID	529
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:identifier[@type="reel number"]	UID	211102366
CAME	mets\dmdSec\mdWrap\xmlData\mods:mods\mods:relatedItem\mods:identifier[@type="reel sequence number"]	UID	530

Quality

Technical Definition

The preservation status of the physical object.

Usage Notes

When microfilming or digitising, many collections rely on bound **volumes** of many newspaper issues, as the volume format better preserves them. For the British Library collections, in those cases where volumes that were tightly bound, they were partially disbound and then shrink-wrapped and returned to the shelves in a partially bound condition for microfilming. For Trove, some state libraries have used formerly bound copies of newspapers, in order to get the best quality for digitisation. The State Library of South Australia disbinds its copies when **microfilming**, as does the Library of Victoria, or they loosen the bindings during the microfilming process in order to better capture text near the margins. Other digitisers indicate that they do not usually disbind, as it is not usually possible to rebind them, and this threatens the physical object. In those cases, the pages can be harder to read as the binding interferes with the scanning process. Owing to policies attached to federal funding for microfilming in the United States, many of the physical collections were damaged during disbinding for the microfilming process and were not preserved after filming.

Examples:

- "Other examples of problems the OCR will have to overcome include; problems with 'set through', (able to see the printing on the reverse of the page through the paper), often due to poor **quality** paper, heavy inking or a combination of the two. Also there are printers' errors, due to paper slippage or creased/folded paper causing breaks in the font during printing." [Ed King, 63]
- "The most brittle papers encountered were from 1860 onwards which typically used poorer **quality** wood pulp paper which is identifiable by its orange brown tinge." [Shaw, 13]

Category Notes

This field is uncommon.

Individual Collection Notes

B1JI: Multiple choice options are "Poor", "Fair" or "Good".

DEAL: String-Choice: "OK", "Missing", "Missing in original", "Damaged", "Retained", "Target" or "As in original".

TRAL: String-Choice: "OK", "Missing", "Missing in original", "Damaged", "Retained", "Target" or "As in original".

Instantiations			
B1JI	BL_newspaper\BL_page\issue_metadata\qualityRating	MCH	Fair
DEAL	alto\Layout\Page@QUALITY	MCH	As in original
TRAL	alto\Layout\Page@QUALITY	MCH	OK

Database

Technical Definition

Provides the name of the digital database in which the issue is stored. Distinct from the Holding Library information.

Usage Notes

Many commercial newspaper digitisers were initially driven to produce complete runs of newspapers. As the major newspapers have now largely been digitised, there is a push to create thematic collections. Gale includes these collections as part of their wider databases but tag them separately as a collection, which is independently searchable. In their partnership with the British Library, for example, they undertook digitisation in two parts, with the second part representing newspapers more marginal, and less well preserved. The database name therefore indicates which batch the issue is from.

Category Notes

This field is rare, appearing only in Gale's B2GL and TDAG.

Individual Collection Notes

B2GL: Present in full and abbreviated form.

TDAG: Provides a UID for the collection.

Instantiations			
B2GL	issue\module	STR	19th Century British Library Newspapers: Part II
B2GL	issue\ztag	STR	BLD
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:classification@authority	UID	ZVDD
SBMA	mets:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:relatedItem\mods:identifier@type	UID	zdb
TDAG	issue\metadatainfo\dviCollectionID	UID	TDAOC0001

Sub-Collection

Technical Definition

Provides details of the sub-collection within which the item has been placed.

Usage Notes

As Trove's sub-collections can be created by end-users, their variant of this field includes the number, names, URLs and the name of creator. In the BL collections, this indicates a broader but standardised sub-category of publication. In the nineteenth century, newspapers were referred to using **geographic** and **genre distinctions**, such as the provincial press, or family papers, though such collections did not coalesce around themes in the way that digitised collections can.

Examples:

- "The letter campaign of 1894 was just the beginning of a life spent agitating for representation and equality for poor women. Nield's career as an author began by writing anonymous letters to the **provincial press**..." [Bunting, 142]
- "The emphasis on **annuals** also obscures the significance of **monthly periodicals** and **weekly provincial and metropolitan newspapers** in the history of nineteenth-century women's poetry." [Easley 2016a, 707]
- "The *Chronicle*, like many **town papers** of the era, brought national issues and trends to regional readers and carried general reports on progressive women's issues of the day." [Bunting, 143]

Category Notes

This field appears in B1JI and the Trove API only, which provide substantially different types of information.

Individual Collection Notes

B1JI: subCollection, specifying e.g. frequency of publication and area (i.e. regional).

TRAP: The name of a public, user created list this article belongs to. There may be more than one list element.

Instantiations			
B1JI	BL_newspaper\BL_page\title_metadata\subCollection	STR	Regional Daily
TRAP	article\list	STR	Beacher-Craggs Family

Filename

Technical Definition

Provides the filename of the image file attached to the XML text. This can take the form of file names, URLs or relative paths with filenames.

Individual Collection Notes

B1JI: The name of the page image file.

CAAL: The name of the image file from which the OCR text was created.

DEAL: The name of the image file from which the OCR text was created.

DEMP: dcx:coordinates@image is the access image file name, while dc:identifier is the article PDF.

F1AL: The name of the image file from which the OCR text was created.

HNDM: Page image file name.

SBME: Page image file name.

TRAL: The name of the image file from which the OCR text was created.

TRAP: A link leading to the PDF for the page this article appears on. There may be more than one pdf element if the article appears over multiple pages. PDFs are not available for some articles, such as the Australian Women's Weekly, and articles with a status of "coming soon".

TRME: Page image file URL.

Instantiations			
B1JI	BL_newspaper\BL_page\pageImage\pageImageFile	FIN	WO1_ANJO_1800_01_06-0001.tif
CAAL	alto\Description\sourceImageInformation\fileName	FIN	/mnt/lustre01.iarchives.com/lustre01/root/projects/production/newspaper/ASLAPR/NDNP_2010_hybrid/batch_datura/200_200_4/ocr/0195.tif
DEAL	alto\Description\sourceImageInformation\fileName	FIN	DDD_010419500_001.jp2
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Component\didl:Resource\dcx:zoning\dcx:coordinates@image	FIN	DDD_010419500_001_access.jp2
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component@dc:identifier	FIN	ddd:010069811:mpeg21:a0001:ocr
DEMP	didl:DIDL\didl:Item\didl:Component@dc:identifier	FIN	ddd:010419500:mpeg21:pdf
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource@ref	FIN	DDD_010419500_001.jp2
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\@ref	FIN	DDD_010097935_001_techmeta.xml
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\@dcx:filename	FIN	DDD_010419500_001_access.jp2
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component@dc:identifier	FIN	ddd:010419500:mpeg21:p001:alto
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\@dcx:filename	FIN	DDD_010419500_001_alto.xml
EUAL	alto\Description\sourceImageInformation\fileName	FIN	//PROC2-NAS1/IN3/EU-NPO_HH/HH_00013/1932/19321231/00000001.tif
EUME	METS:mets\mets:fileSec\mets:fileGrp\mets:file\mets:FLocat@xlink:href	FIN	file://./Viewing/19321231_1-0001.jp2

F1AL	alto\Description\sourceImageInformation\fileName	FIN	//Themis/IN/aikakausleh-det/0355-0257_suomi/1841/suomi_1841 - 0001.tif
F2AL	pageOCRDATA\content\altoXML\alto\Description\sourceImage-Information\fileName	FIN	//hyperion/in/Teolo-gia/1100001713/0001/0341_L000000.TIF
HNDM	$.publication.pagina.imagen	FIN	p0001.tif
HNME	METS:mets\METS:dmdSec\METS:amdSec\METS:mdWrap\METS:xmlData\hiddentext@object	FIN	00213-01.tif
PPAL	alto\Description\sourceImageInformation\fileName		//nasserver-9/impdata$/IN/NZ_5/HBH/1863/01/03/00100.TIF
PPME	mets\structMap\div\div\fptr\par@FILEID	UID	IMG00001
PPME	mets\structMap\div\div\fptr\par@FILEID	UID	ALTO0001
PPME	mets\structMap\div\div\div\div\div\fptr\area@FILEID	UID	ALTO0001
PPME	mets\structMap\div\div\div\div\div\fptr\area@FILEID	UID	ALTO0001
PPME	mets\structMap\div\div\div\div\div\fptr\area@FILEID	UID	ALTO0001
PPME	mets\structMap\div\div\div\div\div\div\div\fptr\area@FILEID	UID	ALTO0001
SBAT	PcGts\Page@imageFilename	FIN	SNP27450028-19340510-0-1-0-0.jpg
SBME	mets:mets\mets\fileSec\mets\fileGrp\mets:file\mets:Flocat@xlink:href	FIN	file:///data/goobi10/tiff001/sbb/1053698984/00000001.tif
SBMY	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:-div\mets:div\mets:mptr	URL	http://content.staatsbib-liothek-berlin.de/zefys/SNP27779154_19340510_020.xml
TRAL	alto\Description\sourceImageInformation\fileName	FIN	nla.news-issn22083111-s75003-b.tif
TRAP	article\pdf	URL	http://trove.nla.gov.au/ndp/imag-eservice/nla.news-page2243325/print
TRME	mets:mets\mets:fileSec\mets:fileGrp\mets:file\mets:FLocat@xlink:href	URL	/pages/nla.news-issn22083111-s75000-b.tif

Descriptive Metadata

The data in this section provides physical descriptive information about the items at various levels. It includes information on **word count**, **page count**, **paragraph style** and **position within the issue**.

Word Count of Page

Language Variants

Anzahl der Wörter der Seite; Aantal Woorden op Pagina; Sivun Sanamäärä.

Technical Definition

The number of words on the page, as identified through Optical Character Recognition.

Usage Notes

The number of words on the average newspaper page decreased over the nineteenth century, partly owing to changes in the law; in the UK, for example, the removal of Stamp Duty, previously applied to any page containing printed news, led to advertising becoming increasingly integrated into the standard newspaper **page**. More generally, lower printing costs led to newspapers having more pages while decreasing in price. The rise of New Journalism in the UK led to the use of larger **headlines** and more **subheadings**, further reducing the space on the page for articles, though some broadsheets, such as *The Times* and the *Daily News*, retained their dense columns of text. Long leader articles, common at mid-century, were replaced with much shorter front-page articles. The Victorian newspaper reader would not have been aware of the exact number of words on the page and traditional press historians have referred instead to "column inches"; the word count is now available through OCR.

Examples:

- "A rough estimation shows that the **number** of distinguishable content pieces within one newspaper issue increases dramatically over the years. In a newspaper of the 18th century some dozens of articles and advertisements can be found, already in the 19th century hundreds of single news can be found and at the heyday of newspapers in the 20th century even some thousand news are included in one issue. This development goes along with an enlargement of the paper size and a reduction of the font size so that more and more text can be delivered to the reader." [Europeana Newspapers 2015, 19]

Category Notes

This field appears only in HNDM.

Individual Collection Notes

HNDM: Total words in field OCR. Total de palabras del campo OCR.

Instantiations			
HNDM	$.publication.pagina.:totalPalabrasOCR	NUM	51

Word Count of Article

Language Variants

Anzahl der Wörter des Artikels; Aantal Woorden in Artikel; Artikkelin Sanamäärä.

Technical Definition

The number of words in the article.

Usage Notes

Leader articles would usually contain around 1,200 words. However, digital analysis has enabled researchers to analyse word counts and revealed that different newspapers had their own standards: for example, a leader for *The Times* in 1855 averages around 1500 words. In contrast to the myth that Victorian writers were paid by the word, it was much more usual for newspapers to request an article of a certain number of **pages**. Although the abolition of Stamp Duty and advances in printing made it cheaper to print more pages, articles did not simply get longer: the introduction of headlines, integrated advertisements and illustrations all influenced article length. Added to this, there was a stark difference between broadsides, with their densely packed columns, and the New Journalism publications of 1880s Britain. Similar developments took place elsewhere, with factors including paper and ink shortages in young Australasian colonies causing publishers to temporarily suspend publication or make use of non-standard paper stock.

Examples:

- "... the **word count** at which *Times* leaders stabilize by the 1850s centres at just over 1500 words." [Liddle, 236]
- "But when I joined the staff of the *Daily Telegraph* and had a free hand in writing at least **three thousand words** every day..." [Sala 1895, 1.363]
- "I thought myself, however, sufficiently well paid with the sum he originally offered me, which was at the **rate of some five or six shillings a line**." [Vizetelly, 1.272]

Category Notes

This field is uncommon, because often issues are zoned by **page** rather than by article.

Individual Collection Notes

B2GI: It is unclear if the word count applies only to the article as zoned on the page, or if it takes into account any other pages including article text.

TDAG: It is unclear if the word count applies only to the article as zoned on the page, or if it takes into account any other pages including article text.

Instantiations

B1GI	issue\page\article\wordCount	NUM	20605
B2GI	issue\page\article\wordCount	NUM	15396
TDAG	issue\page\article\wordCount	NUM	26407
TRAP	article\wordCount	NUM	1514

Page Count of Article

Language Variants

Seitenzahl des Artikels; Aantal Pagina's van Artikel; Artikkelin Sivumäärä.

Technical Definition

Provides the number of pages over which the particular article, as computationally zoned, is spread.

Usage Notes

Article segmentation, where it occurs, is usually done semi-automatically and then checked by an operator. Articles are zoned rather than pages in Gale, Delpher, Trove, and Papers Past in order to support users searching for specific topics and the creation of a search relevance ranking.

Examples:

- "I was debating in my mind on what subject I could most amusingly descant in a paper of from **sixteen to twenty pages in length**..." [Sala 1895, 358]

Category Notes

Although article zoning is relatively common, this specific field is uncommon, appearing only in British Library collections and *The Times* Digital Archive.

Instantiations			
B1GL	issue\article\pc	NUM	1
B1GI	issue\page\article\pc	NUM	1
B2GL	issue\article\pc	NUM	1
B2GI	issue\page\article\pc	NUM	4
TDAG	issue\page\article\pc	NUM	1

Page Count of Issue

Language Variants

Seitenzahl der Ausgabe; Aantal Pagina's van Aflevering; Numeron Sanamäärä.

Technical Definition

Total number of pages of the newspaper **issue**.

Usage Notes

Advances in printing made it cheaper to print more pages, and this is reflected in the growing length of issues throughout the nineteenth century. However, there were stark differences in issue length between different genres of publications; daily publications were often shorter than weekly or monthly titles and provincial papers were often shorter than metropolitan or national publications. Issue length might have also been temporarily reduced in response to shortages of paper or ink, or if the publisher was suffering financial hardship, as was the case of the *Caledonian Mercury* near the end of its run.

Examples:

- "Its **four pages** carried classified advertisements and *London Gazette* announcements, shipping news, London, foreign and local news, Parliamentary reports, prices and advertising." [DNCJ, WHF, 250]
- "All he cared for, in those days of heavily-taxed paper [...] was that a volume of a given **number of pages** should not cost more than a certain sum." [Vizetelly, 1.89]
- "On 1 October 1887 the paper switched to a smaller format, halving in size to a tabloid and doubling the **number of pages** to 16 at the same time." [DNCJ, MaT, 572]
- "In the 19th century newspapers not only the **number of pages** increases significantly but several new elements appear, e.g. pages are structured into several columns, the usage of sections, articles, advertisements, classified advertisements, becomes more sophisticated and complex." [Europeana Newspapers, 24]

Category Notes

This field is relatively uncommon, appearing only in the British Library collections and *The Times* Digital Archive.

Individual Collection Notes

It is always a numeric field.

Instantiations			
B1GL	issue\ip	NUM	4
B1GI	issue\metadataInfo\ip	NUM	4
B1JI	BL_newspaper\BL_page\issue_metadata\pageCount	NUM	4
B2GL	issue\ip	NUM	4
B2GI	issue\metadataInfo\ip	NUM	4
TDAG	issue\metadatainfo\ip	NUM	4

Starting Column for Article

Technical Definition

Provides the column on the page (given as a letter) in which the article begins.

Usage Notes

A daily broadsheet newspaper in the nineteenth century would typically have six or seven columns, though in the UK the abolition of Stamp Duty in 1855 meant that the structure of the newspaper changed quite dramatically, as advertisements were no longer used to bulk out the **page count** but instead were integrated into the paper itself. Illustrated newspapers might have two or three columns and were generally published on smaller paper.

Examples:

As a synonym for regular feature or article:

- "In order to answer this question we must track the journey of the joke from its origins in New York, through its refinement in the papers of the mid-West, its perfection in a Wild West weekly, its journey across the Atlantic, its circulation in the humour **columns** of the British press and finally to its leap into wider popular discourse" [Nicholson 2012, 274]
- "The **column** consists of about a thousand words, dominating an entire page and giving it a substantial weekly presence." [Cogdill, 183]
- "*The Times*, the *Daily News* and the *Morning Chronicle* are typical of the daily broadsheet, having either six or seven **columns** to a page with densely set news items." [DNCJ, MBT, 576]

As a physical layout indicator:

- "This delay roughly mirrors the life cycle of two other jokes published in the same **column**, both of which circulated in the American press during the first half of January." [Nicholson 2012, 274]
- "**Columns** are basic units of design in newspapers and periodicals that distinguish serials from most printed books and predate later additions of page design such as tiered headlines, subheads, font variation and imaginative layout." [DNCJ, LRB/AK, 134]
- "Yet, in the same **column**, Crane could write straightforwardly and even respectfully of the professional fighter, James J. Corbett, then training at Asbury Park." [Kwiat, 104]
- "During that time period, at least, the paper was published in eight-page editions, with each page containing four **columns**." [Simons, 389]
- "Henley paid well, a guinea a **column** (Peterson 180)." [Gray, 150]
- "Two **columns** over from 'The Horrors of Delirium Tremens,' in the same number of the *Missouri Democrat*, we read of a man bitten by a rattlesnake." [Branch, 578]
- "Mr. Brontë had also made several contributions to the *Mercury's* **columns**, the earliest being a letter printed on December 15, 1810, in which he protested against the wrongful imprisonment of a Dewsbury man." [Rosengarten, 591]
- "Like *The Daily Courant*, *The Observator* and *The Flying Post*, it was printed in double **columns**, on both sides of a single folio sheet of paper; and it came out, like *The Evening Post*, *The Post Boy* and *The Review*, three times a week–on Tuesdays, Thursdays and Saturdays." [Bateson, 155]

As a synonym for section (used in plural)

- "There must not be 'unpleasant' jokes; otherwise the babies' food and the condensed milks will not come into the advertising **columns**." [E.T. Raymond, *Portraits of the Nineties* (T. Fisher Unwin, 1921), 305, qtd. in Nicholson 2012, 282]

Category Notes

This field is uncommon, appearing only in two BL legacy collections and the TDA.

Instantiations			
B2GL	issue\article\sc	MCH	B
B2GI	issue\page\article\sc	MCH	A
TDAG	issue\page\article\s	MCH	a

Page Position

Technical Definition

Indicates the position of the page within the issue.

Usage Notes

Unlike the newspapers of today, many nineteenth-century newspapers, particularly the earliest ones, did not have breaking news items on the front pages; news stories would be added to the inside to accommodate later submissions, while the front page would be designed in advance. When *El Universal* was introduced in Mexico in 1888, it made the revolutionary decision of putting the news section on the front page. In general, newspapers were contained in wrappers that were made up of advertisements, so the external pages would not indicate the content. As such, knowing the situation of an article within a paper reveals something about the structure of that newspaper. In addition, some newspapers were made up of only one foldout sheet.

Examples:

- "Like *The Daily Courant, The Observator* and *The Flying Post*, it was printed in double columns on both sides of **a single folio sheet of paper**; and it came out, like *The Evening Post, The Post Boy* and *The Review*, three times a week–on Tuesdays, Thursdays and Saturdays." [Bateson, 155]
- "The differences between these editions and the original tri-weekly **sheet** are not limited to the differences of appearance." [Bateson, 155]
- "They were not appended to any of the material in the original folio **sheets**, and although the newspaper continued publication to 1752 there were no further reprint volumes." [Lockwood, 91]

Category Notes

This field is uncommon.

Individual Collection Notes

DEAL: String-Choice: "Left", "Right", "Foldout", "Single", "Cover".

TDAG: Options are "Single" or "Double".

TRAL: String-Choice: "Left", "Right", "Foldout", "Single", "Cover".

TRAP: Sometimes used to indicate what is printed on the page, or perhaps the page number within a section/supplement. Not displayed in the Trove web interface. Sometimes numeric and sometimes not. e.g. http://api.trove.nla.gov.au/newspaper/45649893?reclevel=full&key=<insert key here>.

Instantiations

DEAL	alto\Layout\Page@POSITION	MCH	Single
TDAG	issue\page\pageid@pageIndicator	MCH	Single
TRAL	alto\Layout\Page@POSITION	MCH	Single
TRAP	article\title\page\pageLabel	STR	Front cover
F1ME	mets\structMap\div\div\div\div@TYPE	STR	FRONT

Paragraph

Technical Definition

Information about paragraphs, including XML containers, text alignment, and UIDs.

Usage Notes

A paragraph is the default unit of a running text, and usually provides a single thought or narrative.

Examples:

- "*Graphic* was published on Saturdays, and 'Place aux Dames' appeared every week, typically in the middle of the newspaper and on one page; often, her **paragraphs** make up the top and bottom third, with illustrations of some recent event, sometimes a sporting event, comprising the middle third; sometimes her paragraphs take up the left two-thirds of the page, with unrelated illustrations on the right." [Cogdill, 182-83]
- "In this, it is 'illustrative,' and one need go no further than the opening **paragraphs** in order to understand just what the instruction might be: in the midst of life we are in death'." [Fieldberg, 14]
- "For one thing, most of the short **paragraphs** of news, which are so conspicuous in the folio edition, were not reprinted." [Bateson, 155]
- "One of them is a proof sheet of the leader published on 18 July 1730 in *Craftsman* No. 218 (reprinted as No. 211 and marked 'A'), with corrections in Amhurst's hand and two manuscript sheets corresponding to part of the printed leader, together with a separate 'Foreign Affairs' **paragraph**." [Lockwood, 95]
- "A report of a meeting of ministers at Ocean Grove, for example, reflects in the lead **paragraph** the correspondent's tongue-in-cheek attitude toward the guardians of moral righteousness." [Kwiat, 104]
- "The correct separation of the text into **paragraphs** is important for many reasons, especially for any kind of Natural Language Processing where tree- and part-of-speech taggers are used. They will in many cases rely on the correct start and end of sentences." [Europeana Newspapers, 48]

Category Notes

These fields appear irregularly in ALTO files and in some of the other file structures. They are primarily used as XML containers, rather than holding data, to which style attributes can be attached.

Individual Collection Notes

DEAL: String-Choice: "Left", "Right", "Center", "Block".

EUAL: String-Choice: "Left", "Right", "Center", "Block".

F1AL: String-Choice: "Left", "Right", "Center", "Block".

F2AL: String-Choice: "Left", "Right", "Center", "Block".

PPAL: String-Choice: "Left", "Right", "Center", "Block".

TRAL: String-Choice: "Left", "Right", "Center", "Block".

Instantiations

As an XML container

B1GL	issue\article\text\text.title\p	NUL	
B1GL	issue\article\text\text.cr\p	NUL	
B2GL	issue\article\text\text.title\p	NUL	
SBAT	PcGts\Page\TextRegion	NUL	
TDAG	issue\page\article\text\text.cr\p	NUL	

The description of text alignment for paragraph text			
DEAL	alto\Styles\ParagraphStyle@ALIGN	MCH	Right
EUAL	alto\Styles\ParagraphStyle@ALIGN	MCH	Right
F1AL	alto\Style\ParagraphStyle@ALIGN	MCH	Left
F2AL	pageOCRDATA\content\altoXML\alto\Styles\ParagraphStyle@ALIGN	MCH	Center
PPAL	alto\Styles\ParagraphStyle@ALIGN	MCH	Right
TRAL	alto\Styles\ParagraphStyle@ALIGN	MCH	Left
Unique identifier for paragraph text			
DEAL	alto\Styles\ParagraphStyle@ID	UID	PAR_RIGHT
EUAL	alto\Styles\ParagraphStyle@ID	UID	PAR_CENTER
F1AL	alto\Style\ParagraphStyle@ID	UID	PAR_LEFT
F2AL	pageOCRDATA\content\altoXML\alto\Styles\ParagraphStyle@ID	UID	PAR_CENTER
PPAL	alto\Styles\ParagraphStyle@ID	UID	PAR_RIGHT
SBAT	PcGts\Page\TextRegion@id	UID	R0
TRAL	alto\Styles\ParagraphStyle@ID	UID	PAR1

Social Interaction Data

This section contains data relevant to social interaction with the databases, namely **comments, tagging,** or division into **sub-collections by users**.

Comments and Social Tagging

Technical Definition

Information about tags and comments added to an **article** by online users.

Usage Notes

Only Trove currently has the functionality to support user-generated tags or to provide this information to other users via its API.

Category Notes

These fields only appear in the Trove API. All dates are recorded in UTC time zone notation. There may be more than one tag element.

Instantiations			
TRAP	article\tag@tag@lastupdated	DAT	"2012-02-20T23:57:57Z"
TRAP	article\comment	STR	Robert Fuy should probably read Robert Fuz
TRAP	article\commentCount	NUM	1
TRAP	article\correctionCount	NUM	1
TRAP	article\tag	STR	fashion
TRAP	article\tagCount	NUM	3
TRAP	article\comment@comment@by	STR	public:annmanley
TRAP	article\comment@comment@lastupdated	DAT	2012-02-20T23:57:57Z
TRAP	article\list@list@by	STR	"dance" (NLA staff accounts do not include the 'public' prefix)
TRAP	article\list@list@lastupdated	DAT	2012-02-29T01:02:15Z
TRAP	article\list@list@url	URL	/list/21922
TRAP	article\lastCorrection@lastCorrection@by	STR	"anonymous" (no user account identified)
TRAP	article\lastCorrection@lastCorrection@lastupdated	DAT	2011-12-06T18:05:17Z

Technical Metadata

This section refers to all data that may be understood as relating to the digitisation or digital display processes. It includes information relating to **OCR** and **page skew**, various **IDs** for linking digital files and **metadata standards**, algorithmically derived font and style information, and **dimensions** or **display coordinates** of objects.

OCR Information

Technical Definition

Provides technical information on the OCR software used, including Description ID, Agency, Date and Time, Step Description, Step Settings, Creator, Name, Version, Relevance and Confidence Level.

Usage Notes

Optical Character Recognition (OCR) software recognises text from scanned pages and is also used to populate metadata fields by identifying title information, layout, etc. OCR first pre-processes the image into component parts such as text blocks, sentence blocks and word blocks. This is **zoning** and is done slightly differently by different digitisers (sometimes by article, for example). Common OCR software is most effective on modern Roman type; older newspapers and those using Gothic font are difficult to OCR successfully. Delpher worked with volunteers to improve the quality of the text of newspapers, particularly seventeenth-century newspapers and Second World War illegal resistance newspapers, while Trove crowdsources the correction of article text to improve accuracy.

Category Notes

This field is very common, appearing in all except B1GP and B2GP, which only provide information about **alternate newspaper titles**. Only DEAL and TRAL have a processing agency and date.

Individual Collection Notes

B1GI: @primary specifies the OCR language is the primary language (see **language**).

B2GL: Relevance is a string choice "yes" or "no".

B2GI: Relevance is a numerical score. @primary specifies the OCR language is the primary language (see **language**).

CAAL: sourceImageInformation contains information to identify the image file from which the OCR text was created. processingStepSettings: A description of any setting of the processing application. For example, for a multi-engine OCR application this might include the engines which were used. Ideally, this description should be adequate so that someone else using the same application can produce identical results. Confidence is given as a numerical value for page and word, with a value between 0 (unsure) and 1 (sure). processingStepDescription provides an ordinal listing of the image processing steps performed, for example "image despeckling".

DEAL: **Page confidence**, with a value between 0 (unsure) and 1 (sure); string confidence, as a list of numbers, one number between 0 (sure) and 9 (unsure) for each character; word confidence, with a value between 0 (unsure) and 1 (sure).

DEMP: **Confidence level** of the correctness of the character recognition on this page, 0 is no confidence at all, 1 is highest confidence. (This page confidence level is an aggregate of the confidence levels for each word and character, as listed in the alto files.)

EUAL: Page confidence, with a value between 0 (unsure) and 1 (sure); string confidence, as a list of numbers, one number between 0 (sure) and 9 (unsure) for each character; word confidence, with a value between 0 (unsure) and 1 (sure).

F1AL: String confidence as a list of numbers, with one number between 0 (sure) and 9 (unsure) for each character; word confidence, with a value between 0 (unsure) and 1 (sure).

F2AL: Page confidence, with a value between 0 (unsure) and 1 (sure); string confidence, with a value 0-9 for each character; word confidence, with a value 0-1; character confidence, with a value between 0 (unsure) and 1 (sure).

PPAL: Estimated percentage of OCR Accuracy in range from 0 to 100; character confidence level 0-9 for each character in string; word confidence for whole string.

SBAT: Word confidence, with a value between 0 (unsure) and 1 (sure).

TRAL: Page confidence, with a value between 0 (unsure) and 1 (sure); string confidence with a value 0-9 for each character; word confidence for whole string.

TRAP: Relevance is a string choice, with the options "very relevant", "likely to be relevant", "may have relevance" or "vaguely relevant", and a numerical representation.

Instantiations			
Relevance			
B2GL	issue\article\ocr\relevant	STR	yes
B2GI	issue\page\article\ocr	NUM	41.29
TRAP	article\relevance	MCH	very relevant
TRAP	article\relevance@relevance@score	NUM	0.009942865
Confidence level			
CAAL	alto\Layout\Page@PC	NUM	1
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WC	NUM	1
DEAL	alto\Layout\Page@PC	NUM	0.885
DEAL	:alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@CC	NUM	80150190
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WC	NUM	0.99
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ddd:OCRConfidencelevel	NUM	0.885
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ddd:OCRConfidencelevel	NUM	0.885
EUAL	alto\Layout\Page@PC	NUM	0.853
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@CC	NUM	77800000000000
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WC	NUM	0.98
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@CC	NUM	2030626684413
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WC	NUM	0.65
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page@PC	NUM	0.825
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block\TextLine\String@CC	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block\TextLine\String@WC	NUM	0.49
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@CC	NUM	4444
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@WC	NUM	0.53
PPAL	alto\Layout\Page@ACCURACY	NUM	93.58
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@CC	NUM	0
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WC	NUM	1
SBAT	PcGts\Page\TextRegion\Textline\Word\TextEquiv@conf	NUM	0.68334
TDAG	issue\metadatainfo\ocr	NUM	32.73
TDAG	issue\page\article\ocr	NUM	13.28
TRAL	alto\Layout\Page@PC	NUM	0
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block\TextLine\String@CC	NUM	3
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block\TextLine\String@WC	NUM	0.7

Primary language			
B1GI	issue\metadataInfo\language@primary	BOO	Y
B2GI	issue\metadataInfo\language@primary	BOO	Y
TDAG	issue\metadatainfo\language@primary	BOO	Y
Unique identifiers			
CAAL	alto\Description\OCRProcessing@ID	UID	OCR.0
CAAL	alto\Layout\Page@PROCESSING	UID	OCR.0
DEAL	alto\Description\OCRProcessing@ID	UID	OCRPROCESSING_1
EUAL	alto\Description\OCRProcessing@ID	UID	
F1AL	alto\Description\OCRProcessing@ID	UID	OCRPROCESSING_1
F2AL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing@ID	UID	OCRPROCESSING_1
PPAL	alto\Description\OCRProcessing@ID	UID	OCRPROCESSING_1
TRAL	alto\Description\OCRProcessing@ID	UID	OCR1
TRAL	alto\Layout\Page@PROCESSING	UID	OCR1
Processing step settings			
CAAL	alto\Description\OCRProcessing\ocrProcessingStep\processing-StepSettings	STR	Conf:1 abbyy9.option.analyze-manual-zones:false Lang:en_US Inverted:false […]
DEAL	alto\Description\OCRProcessing\preProcessingStep\processing-StepSettings	STR	CCS OCR Processing
Software creator			
CAAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareCreator	STR	iArchives
DEAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareCreator	STR	CCS Content Conversion Specialists GmbH, German
DEAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareCreator	STR	ABBYY (BIT Software), Russia
EUAL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareCreator	STR	CCS Content Conversion Specialists GmbH, Germany
EUAL	alto\Description\OCRProcessing\OCRProcessingStep\processing-Software\softwareCreator	STR	ABBYY (BIT Software), Russia
F1AL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareCreator	STR	CCS Content Conversion Specialists GmbH, Germany
F1AL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareCreator	STR	ABBYY (BIT Software), Russia
F2AL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\preProcesssingStep\processingSoftware\softwareCreator	STR	CCS Content Conversion Specialists GmbH, Germany

PPAL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareCreator	STR	CCS Content Conversion Specialists GmbH, Germany
PPAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareCreator	STR	ABBYY (BIT Software), Russia
TRAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareCreator	STR	ABBYY
TRAL	alto\Description\OCRProcessing\postProcessingStep\processingSoftware\softwareCreator	STR	iSolve Technologies Pvt Ltd
Software name			
CAAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	iArchives OCR Framework
DEAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	CCS docWORKS
DEAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	FineReader
EUAL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareName	STR	CCS docWorks
EUAL	alto\Description\OCRProcessing\OCRProcessingStep\processingSoftware\softwareName	STR	FineReader
F1AL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareName	STR	CCS docWORKS
F1AL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	Finereader
F2AL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\preProcesssingStep\processingSoftware\softwareName	STR	CCS docWORKS
F2AL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	FineReader
PPAL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\preProcesssingStep\processingSoftware\softwareName	STR	CCS docWORKS
PPAL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	FineReader
TRAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareName	STR	FineReader
TRAL	alto\Description\OCRProcessing\postProcessingStep\processingSoftware\softwareName	STR	iClip
Software version			
CAAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareVersion	STR	Multiple
DEAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareVersion	STR	6.3-0.91
DEAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareVersion	STR	8.1

EUAL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareVersion	STR	6.7-1.35
EUAL	alto\Description\OCRProcessing\OCRProcessingStep\processingSoftware\softwareVersion	STR	10
F1AL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareVersion	STR	6.0-8.19
F2AL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\preProcesssingStep\processingSoftware\softwareVersion	STR	6.4-0.29
F2AL	pageOCRDATA\content\altoXML\alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareVersion	STR	7.1
PPAL	alto\Description\OCRProcessing\preProcessingStep\processingSoftware\softwareVersion	STR	6.2-1.5
TRAL	alto\Description\OCRProcessing\ocrProcessingStep\processingSoftware\softwareVersion	STR	12
TRAL	alto\Description\OCRProcessing\postProcessingStep\processingSoftware\softwareVersion	STR	5
Application description			
DEAL	alto\Description\OCRProcessing\preProcessingStep\applicationDescription	STR	
Processing agency			
DEAL	alto\Description\OCRProcessing\preProcessingStep\processingAgency	STR	CCS Content Conversion Specialists GmbH, www.content-conversion.com
TRAL	alto\Description\OCRProcessing\ocrProcessingStep\processingAgency	STR	iSolve Technologies Pvt Ltd
TRAL	alto\Description\OCRProcessing\postProcessingStep\processingAgency	STR	iSolve Technologies Pvt Ltd
Processing date			
DEAL	alto\Description\OCRProcessing\preProcessingStep\processingDateTime	DAT	13/10/2009
TRAL	alto\Description\OCRProcessing\ocrProcessingStep\processingDateTime	DAT	2017-09-13T05:47:57
TRAL	alto\Description\OCRProcessing\postProcessingStep\processingDateTime	DAT	2017-09-13T05:47:58
Processing step description			
DEAL	alto\Description\OCRProcessing\preProcessingStep\processingStepDescription	STR	align
TRAL	alto\Description\OCRProcessing\postProcessingStep\processingStepDescription	STR	iSolve -> ALTO

ID

Technical Definition

Provides a unique ID for the component of the image.

Usage Notes

This can be a URL, a **filename**, a numerical string, or a string incorporating an **abbreviated title**.

Category Notes

These field are very common and used for linking individual newspapers, issues and articles as well as style information.

Individual Collection Notes

SBMY: There are two newspaper URLs, one referring to the catalogue page of the SBB itself and one referring to the catalogue page of the ZEFYS.

Instantiations			
Collection			
B1GI	issue\metadataInfo\dviCollectionID	UID	BNCNC0001
B2GI	issue\metadataInfo\dviCollectionID	UID	BNCNC0002
Newspaper			
B1GI	issue\metadataInfo\assetID	UID	3405609147
B1GI	issue\page\assetID	UID	3305609147
B1GI	issue\page\article\assetID	UID	3305609147
B1GI	issue\metadataInfo\newspaperID	UID	BNCN0001
B2GI	issue\metadataInfo\newspaperID	UID	NCBL0127
B2GI	issue\metadataInfo\assetID	UID	3411020871
B2GI	issue\page\assetID	UID	3500902290
B2GI	issue\page\article\assetID	UID	3211020871
CADI	ROOT\null[sn]	UID	sn85032923
DEMP	didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dc:identifier	URL	http://resolver.kb.nl/resolve?urn=d-dd:010419500:mpeg21
HNDM	$.publication._id	STR	ObjectId("558075bd7d1e63c-9fea1a0eb")
SBMY	mets:mets\mets:amdSec\mets:digitprovMD\mets:mdWrap\mets:xmlData\dv:links\dv:reference	URL	http://www.stabikat.de/DB=1/PPN?PPN=791048292
SBMY	mets:mets\mets:amdSec\mets:digitprovMD\mets:mdWrap\mets:xmlData\dv:links\dv:presentation	URL	http://zefys.staatsbiblio-thek-berlin.de/list/title/zdb/27450028/
TDAG	issue\metadatainfo\newspaperID	UID	TDAO0001
TRAP	article\title@title@id	UID	42
TRME	mets:mets\mets:amdSec\mets:techMD\mets:mdWrap\mets:xmlDa-ta\premis:object\premis:objectIdentifier\premis:objectIdentifierValue	FIN	nla.news-issn22083111-s75000-b.tif

Newspaper (alternate title)			
B1GP	PubInfo\VariantTitles\NewspaperID	UID	BNCN0001
B2GP	PubInfo\VariantTitles\NewspaperID	UID	NCBL0126
Article (unique/database)			
B1GI	issue\page\article\id	UID	WO1_ANJO_1798_01_02-0001-001
B1GL	issue\article\id	UID	WOJL-1822-01-03-0002-003
B1GT	articles\artInfo@id	UID	WO1_ANJO_1798_01_02-0001-001
B2GL	issue\article\id	UID	WOJL-1822-01-03-0002-003
B2GT	articles\artInfo@id	UID	YOHD-1813-07-31-0002-003
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item@dc:identifier	UID	ddd:010069811:mpeg21:a0001
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item@ddd:article_id	UID	ddd:010419500:mpeg21:a0001
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx:recordIdentifier	UID	ddd:010069811:mpeg21:a0001
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item@ddd:article_id	UID	ddd:010419500:mpeg21:a0001
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\srw_dc:dcx\dcx:recordIdentifier	UID	ddd:010419500:mpeg21:a0001
TDAG	issue\page\article\assetID	UID	cs17301582
DTAG	issue\page\article\id	UID	0FFO-1785-FEB14-001-004
TRAP	article@article@id	UID	18341291
Article (URL)			
B1GT	articles\artInfo\ProductLink	URL	http://gdc.galegroup.com/gdc/artemis/NewspapersDetailsPage/NewspapersDetailsWindow?prodId=BNCN&windowstate=normal&mode=view&displayGroupName=DVI-Newspapers&p=GDCS&action=e&documentId=GALE%7CBA3205609147
B2GT	articles\artInfo\ProductLink	URL	http://gdc.galegroup.com/gdc/artemis/NewspapersDetailsPage/NewspapersDetailsWindow?prodId=BNCN&windowstate=normal&mode=view&displayGroupName=DVI-Newspapers&p=GDCS&action=e&documentId=GALE%7CR3211020873

DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item\didl:Component\srw_dc:dcx\dc:identifier	URL	http://resolver.kb.nl/resolve?urn=d-dd:010419500:mpeg21:a0001
SBME	mets:mets\mets:structMap[@TYPE="LOGICAL"]\mets:div@CONTENTIDS	URL	http://resolver.staatsbibliothek-berlin.de/SNP27779154-00000000-0-0-0-0
TRAP	article\identifier	URL	http://nla.gov.au/nla.news-article18341291
TRAP	article\troveUrl	URL	http://trove.nla.gov.au/ndp/del/article/18342701
Article (relative/issue)			
EUME	METS:mets\mets:structMap[@LABEL="Logical Structure"]\mets:div\mets:div\mets:div\mets:div\mets:div[@TYPE="ARTICLE"]@ID	UID	DIVL7
EUME	METS:mets\mets:dmdSec@ID	UID	MODSMD_ARTICLE1
EUME	METS:mets\mets:dmdSec\mets:mdWrap\mets:xmlData\mods:mods\mods:titleInfo@ID	UID	MODSMD_ARTICLE1_TI1
PPME	mets\structMap\div\div\div@DMDID	UID	MODSMD_ARTICLE1
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock@ID	UID	ART1
TRME	mets:mets\mets:structMap[@TYPE="logical"]\mets:div\mets:div@DMDID	UID	modsarticle1
Article zone ID			
Provides an ID for the article, as zoned on the image. These fields are highly inconsistent and include unique IDs, across the database, relative IDs within the issue, and URLs to the web-accessible version of the image.			
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Component@dc:identifier	UID	ddd:010419500:mpeg21:p001:a0001:zoning
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Item@dc:identifier	UID	:010419500:mpeg21:p001:a0001
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@ID	UID	P1_CB00001
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock@ID	UID	ZONE1-1
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:div@ID	UID	divarticle1-1
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:fptr\mets:area@FILEID	FIN	nla.news-issn22083111-s75003-b.tif
TRME	mets:mets\mets:structMap\mets:div\mets:div@ID	UID	artzone1-1
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:fptr\mets:area@FILEID	FIN	nla.news-issn22083111-s75003-b.tif
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:fptr\mets:area@FILEID	FIN	nla.news-issn22083111-s75003-b.xml
Clipped article image (unique)			
Provides a unique identifier for the article, as segmented on a single page. These are unique to each segment of an article spanning multiple pages.			

B1GL	issue\article\ci	UID	WOJL-1822-01-03-0002-003-001
B2GL	issue\article\ci	UID	WOJL-1822-01-03-0002-003-001
TDAG	issue\page\article\ci	UID	0FFO-1812-JUL14-001-001-001
Clipped article image (relative)			
Provides a relative identifier for the article, as segmented on a single page. These are sequential to each segment of an article.			
B1GL	issue\article\ci\clip	NUM	1
B1GL	issue\article\text\text.cr\pg@clipref	NUM	1
B1GI	issue\page\article\il@clipref	NUM	1
B2GL	issue\article\ci\clip	NUM	1
B2GL	issue\article\text\text.title\pg\clipref	NUM	1
TDAG	issue\page\article\text\text.cr\pg@clipref	NUM	1
TDAG	issue\page\article\text\text.title\pg@clipref	NUM	1
Text or paragraph style			
Individual Collection Notes			
The contents of @STYLEREFS declare the Text or Paragraph style of its parent element via linking. The value of this @STYLEREFS attribute matches the value of the @ID attribute on the <TextStyle> or <ParagraphStyle> element (both children of the <Styles> element.			
CAAL	alto\Styles\TextStyle@ID	UID	TS_10.0
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@STYLEREFS	UID	TS_10.0
DEAL	alto\Styles\TextStyle@ID	UID	TXT_0
DEAL	alto\Layout\Page\PrintSpace\TextBlock@STYLEREFS	UID	TXT_0 PAR_RIGHT
EUAL	alto\Styles\TextStyle@ID	UID	TXT_0
EUAL	alto\Layout\Page\PrintSpace\TextBlock@STYLEREFS	UID	TXT_0 PAR_CENTER
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@STYLEREFS	UID	TXT_2
F1AL	alto\Style\TextStyle@ID	UID	TXT_0
F1AL	alto\Layout\Page\PrintSpace\TextBlock@STYLEREFS	UID	TXT_0 PAR_LEFT
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@STYLEREFS	UID	TXT_1
F2AL	pageOCRDATA\content\altoXML\alto\Styles\TextStyle@ID	UID	TXT_0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock@STYLEREFS	UID	TXT_0 PAR_CENTER

F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock@STYLEREFS	UID	TXT_1 PAR_Block
PPAL	alto\TextStyle@FONTFAMILY	UID	TXT_0
PPAL	alto\Layout\Page\PrintSpace\TextBlock@STYLEREFS	UID	TXT_0 PAR_CENTER
TRAL	alto\Styles\TextStyle@ID	UID	TS7.7
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock@STYLEREFS	UID	PAR1
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@STYLEREFS	UID	PAR1
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine@STYLEREFS	UID	TS.10.2
Print space			
CAAL	alto\Layout\Page\PrintSpace@ID	UID	PS.0
DEAL	alto\Layout\Page\PrintSpace@ID	UID	P1_PS00001
EUAL	alto\Layout\Page\PrintSpace@ID	UID	P1_PS00001
F1AL	alto\Layout\Page\PrintSpace@ID	UID	P1_PS00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace@ID	UID	P12_PS00001
PPAL	alto\Layout\Page\PrintSpace@ID	UID	P1_PS00001
TRAL	alto\Layout\Page\PrintSpace@ID	UID	SPACE1
Text block			
CAAL	alto\Layout\Page\PrintSpace\TextBlock@ID	UID	TB.0195.1
DEAL	alto\Layout\Page\PrintSpace\TextBlock@ID	UID	P1_TB00001
EUAL	alto\Layout\Page\PrintSpace\TextBlock@ID	UID	P1_TB00001
F1AL	alto\Layout\Page\PrintSpace\TextBlock@ID	UID	P1_TB00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock@ID	UID	P12_TB00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock@ID	UID	P12_TB00002
PPAL	alto\Layout\Page\PrintSpace\TextBlock@ID	UID	P1_TB00001
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\@ID	UID	BLOCK1
Text line			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@ID	UID	TB.0195.1_0
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@ID	UID	P1_TL00001
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@ID	UID	P1_TL00001

F1AL	:alto\Layout\Page\PrintSpace\TextBlock\TextLine@ID	UID	P1_TL00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block\TextLine@ID	UID	P12_TL00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine@ID	UID	P12_TL00002
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@ID	UID	P1_TL00001
SBAT	PcGts\Page\TextRegion\Textline@id	UID	l1
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block\TextLine@ID	UID	LINE1
String			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ID	UID	TB.0195.1_0_0
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ID	UID	P1_ST00001
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ID	UID	P1_ST00001
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ID	UID	P1_ST00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block\TextLine\String@ID	UID	P12_ST00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@ID	UID	P12_ST00002
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@ID	UID	P1_ST00001
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block\TextLine\String@ID	UID	S1
Word			
SBAT	PcGts\Page\TextRegion\Textline\Word@id	UID	w2
White space			
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@ID	UID	P1_SP00002
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@ID	UID	P1_SP00304
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@ID	UID	P1_SP00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block\TextLine\String\SP@ID	UID	P12_SP00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\SP@ID	UID	P12_SP00006
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@ID	UID	P1_SP00001
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block\TextLine\SP@ID	UID	SP1
Margin			
DEAL	alto\Layout\Page\TopMargin@ID	UID	P1_TM00001
DEAL	alto\Layout\Page\LeftMargin@ID	UID	P1_LM00001

DEAL	alto\Layout\Page\RightMargin@ID	UID	P1_RM00001
DEAL	alto\Layout\Page\BottomMargin@ID	UID	P1_BM00001
EUAL	alto\Layout\Page\TopMargin@ID	UID	P1_TM0001
EUAL	alto\Layout\Page\LeftMargin@ID	UID	P1_LM0001
EUAL	alto\Layout\Page\RightMargin@ID	UID	P1_RM0001
EUAL	alto\Layout\Page\BottomMargin@ID	UID	P1_BM0001
F1AL	alto\Layout\Page\TopMargin@ID	UID	P1_TM00001
F1AL	alto\Layout\Page\LeftMargin@ID	UID	P1_LM00001
F1AL	alto\Layout\Page\RightMargin@ID	UID	P1_RM00001
F1AL	alto\Layout\Page\BottomMargin@ID	UID	P1_BM00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin@ID	UID	P12_TM00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\LeftMargin@ID	UID	P12_LM00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\RightMargin@ID	UID	P12_RM00001
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin@ID	UID	P12_BM00001
PPAL	alto\Layout\Page\TopMargin@ID	UID	P1_TM00001
PPAL	alto\Layout\Page\LeftMargin@ID	UID	P1_LM00001
PPAL	alto\Layout\Page\RightMargin@ID	UID	P1_RM00001
PPAL	alto\Layout\Page\BottomMargin@ID	UID	P1_BM00001
TRAL	alto\Layout\Page\TopMargin@ID	UID	MARGIN1
TRAL	alto\Layout\Page\LeftMargin@ID	UID	MARGIN2
TRAL	alto\Layout\Page\RightMargin@ID	UID	MARGIN3
TRAL	alto\Layout\Page\BottomMargin@ID	UID	MARGIN4

Coordinates

Technical Definition

Provides coordinates for a component of the image.

Individual Collection Notes

CAAL: Expressed in inch1200 rather than default 1/10 of mm.

DEAL: Expressed in pixels rather than default 1/10 of mm.

Instantiations			
Page (4-point)			
B1JI	BL_newspaper\BL_page\pageImage\pageCoordinates	COO	79,915,344,925,613
B2GL	issue\article\text\text.cr\pg@pos	COO	1219, 1491, 1298, 1528
TDAG	issue\page\article\text\text.cr\pg@pos	COO	34,406,3060,4943
Title			
TDAG	issue\page\article\text\text.title\pg@pos	COO	49,403,763,449
Print space (horizontal)			
CAAL	alto\Layout\Page\PrintSpace@HPOS	NUM	0
DEAL	alto\Layout\Page\PrintSpace@HPOS	NUM	20
EUAL	alto\Layout\Page\PrintSpace@HPOS	NUM	161
F1AL	alto\Layout\Page\PrintSpace@HPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space@HPOS	NUM	42
PPAL	alto\Layout\Page\PrintSpace@HPOS	NUM	8
TRAL	alto\Layout\Page\PrintSpace@HPOS	NUM	0
Print space (vertical)			
CAAL	alto\Layout\Page\PrintSpace@VPOS	NUM	0
DEAL	alto\Layout\Page\PrintSpace@VPOS	NUM	13
EUAL	alto\Layout\Page\PrintSpace@VPOS	NUM	224
F1AL	alto\Layout\Page\PrintSpace@VPOS	NUM	264
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space@VPOS	NUM	75
PPAL	alto\Layout\Page\PrintSpace@VPOS	NUM	27
TRAL	alto\Layout\Page\PrintSpace@VPOS	NUM	0

Article zone			
B1GL	issue\article\text\text.cr\pg\@pos	COO	1219, 1491, 1298, 1528
B2GL	issue\article\text\text.title\pg@pos	COO	1219, 1491, 1298, 1528
EUAL	alto\Layout\Page\PrintSpace\TextBlock\Shape\Polygon@POINTS	COO	3778,594 2392,594 2392,527 1607,527 1607,594 228,594 228,284 3778,284
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@HPOS	NUM	1894
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@VPOS	NUM	546
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@HPOS	NUM	1894
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@VPOS	NUM	546
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@HPOS	NUM	2632
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@VPOS	NUM	1520
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@HPOS	NUM	2632
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@VPOS	NUM	1520
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:fptr\mets:area@COORDS	COO	187,152,026,328,128
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:div\mets:fptr\mets:area@COORDS	COO	187,152,026,328,128
TRME	mets:mets\mets:structMap\mets:div\mets:div\mets:fptr\mets:area@COORDS	COO	187,152,026,328,128
Article zoning starting (top right) coordinates			
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Component\didl:Resource\dcx:zoning\dcx:coordinates@hpos	NUM	22
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Component\didl:Resource\dcx:zoning\dcx:coordinates@vpos	NUM	305
Paragraph			
SBAT	PcGts\Page\TextRegion\Coords@points	COO	190,106 190,148 213,148 213,149 216,149 216,150 226,150 226,151 228,151 228,152 229,152 229,153 231,153 231,154 233,154 233,155 234,155 234,156 235,156 235,157 [...]
Text block (horizontal)			
CAAL	alto\Layout\Page\PrintSpace\TextBlock@HPOS	NUM	2560
DEAL	alto\Layout\Page\PrintSpace\TextBlock@HPOS	NUM	824
EUAL	alto\Layout\Page\PrintSpace\TextBlock@HPOS	NUM	1819
F1AL	alto\Layout\Page\PrintSpace\TextBlock@HPOS	NUM	212

F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block@HPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space\TextBlock@HPOS	NUM	42
PPAL	alto\Layout\Page\PrintSpace\TextBlock@HPOS	NUM	8
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block@HPOS	NUM	2632
Text block (vertical)			
CAAL	alto\Layout\Page\PrintSpace\TextBlock@VPOS	NUM	20596
DEAL	alto\Layout\Page\PrintSpace\TextBlock@VPOS	NUM	13
EUAL	alto\Layout\Page\PrintSpace\TextBlock@VPOS	NUM	224
F1AL	alto\Layout\Page\PrintSpace\TextBlock@VPOS	NUM	957
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\Text-Block@VPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space\TextBlock\@VPOS	NUM	111
PPAL	alto\Layout\Page\PrintSpace\TextBlock@VPOS	NUM	27
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block@VPOS	NUM	2742.67
Text line			
SBAT	PcGts\Page\TextRegion\Textline\Coords@points	COO	63,146 2198,146 2198,407 63,407
Text line (horizontal)			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HPOS	NUM	2560
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HPOS	NUM	829
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HPOS	NUM	1819
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HPOS	NUM	686
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine@HPOS	NUM	643
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space\TextBlock\TextLine@HPOS	NUM	48
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HPOS	NUM	8
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\Text-Block\TextLine@HPOS	NUM	2655.98
Text line (vertical)			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@VPOS	NUM	2096
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@VPOS	NUM	17

EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@VPOS	NUM	226
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@VPOS	NUM	925
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine@VPOS	NUM	48
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine@VPOS	NUM	115
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@VPOS	NUM	27
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine@VPOS	NUM	2821.22
String (horizontal)			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HPOS	NUM	2560
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@HPOS	NUM	7824
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HPOS	NUM	829
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HPOS	NUM	1819
F2AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HPOS	NUM	686
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String@HPOS	NUM	643
PPAL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@HPOS	NUM	48
TRAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HPOS	NUM	8
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\String@HPOS	NUM	2886
String (vertical)			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@VPOS	NUM	2096
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@VPOS	NUM	2272
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@VPOS	NUM	17
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@VPOS	NUM	226
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@VPOS	NUM	925
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String@VPOS	NUM	48
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@VPOS	NUM	120
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@VPOS	NUM	27
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\String@HPOS	NUM	2821.22

Word			
B1GL	issue\article\text\text.title\p\wd@pos	COO	1219, 1491, 1298, 1528
B1GL	issue\article\text\text.cr\pg\wd@pos	COO	1219, 1491, 1298, 1528
B1JI	BL_newspaper\BL_page\pageText@coord	COO	125,695,208,732
B2GL	issue\article\text\text.title\pg\wd@pos	COO	1219, 1491, 1298, 1528
B2GL	issue\article\text\text.cr\p\wd@pos	COO	1219, 1491, 1298, 1528
HNME	METS:mets\METS:dmdSec\METS:amdSec\METS:mdWrap\METS:xmlData\hiddentext\pagecolumn\region\paragraph\line\word@coords	COO	324,43,691,14
SBAT	PcGts\Page\TextRegion\Textline\Word\Coords@points	COO	63,148 1099,148 1099,382 63,382
TDAG	issue\page\article\text\text.cr\p\wd@pos	COO	107,451,394,490
TDAG	issue\page\article\text\text.title\p\wd@pos	COO	90,422,194,449
White space (horizontal)			
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@HPOS	NUM	1113
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@HPOS	NUM	2798
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@HPOS	NUM	877
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\SP@HPOS	NUM	695
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\SP@HPOS	NUM	524
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\SP@VPOS	NUM	153
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@HPOS	NUM	931
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\SP@HPOS	NUM	3105.01
White space (vertical)			
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@VPOS	NUM	46
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@VPOS	NUM	1675
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@VPOS	NUM	1099
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\SP@VPOS	NUM	72
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@VPOS	NUM	28
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\SP@VPOS	NUM	2821.22
Hyphen			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@HPOS	NUM	3364
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@VPOS	NUM	23244

DEAL	Alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@HPOS	NUM	554
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@VPOS	NUM	374
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@HPOS	NUM	2873
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@VPOS	NUM	1647
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\HYP@HPOS	NUM	640
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\HYP@VPOS	NUM	124
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\HYP@HPOS	NUM	640
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\HYP@VPOS	NUM	124
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@HPOS	NUM	588
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@VPOS	NUM	795
Margin (horizontal)			
DEAL	alto\Layout\Page\TopMargin@HPOS	NUM	0
DEAL	alto\Layout\Page\LeftMargin@HPOS	NUM	0
DEAL	alto\Layout\Page\RightMargin@HPOS	NUM	1364
DEAL	alto\Layout\Page\BottomMargin@HPOS	NUM	0
EUAL	alto\Layout\Page\TopMargin@HPOS	NUM	0
EUAL	alto\Layout\Page\LeftMargin@HPOS	NUM	0
EUAL	alto\Layout\Page\RightMargin@HPOS	NUM	3818
EUAL	alto\Layout\Page\BottomMargin@HPOS	NUM	0
F1AL	alto\Layout\Page\TopMargin@HPOS	NUM	0
F1AL	alto\Layout\Page\LeftMargin@HPOS	NUM	0
F1AL	alto\Layout\Page\RightMargin@HPOS	NUM	1482
F1AL	alto\Layout\Page\BottomMargin@HPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin@HPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\LeftMargin@HPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\RightMargin@HPOS	NUM	1299
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin@HPOS	NUM	0
PPAL	alto\Layout\Page\TopMargin@HPOS	NUM	0
PPAL	alto\Layout\Page\LeftMargin@HPOS	NUM	0
PPAL	alto\Layout\Page\RightMargin@HPOS	NUM	2538

PPAL	alto\Layout\Page\BottomMargin@HPOS	NUM	0
TRAL	alto\Layout\Page\TopMargin@HPOS	NUM	0
TRAL	alto\Layout\Page\LeftMargin@HPOS	NUM	0
TRAL	alto\Layout\Page\RightMargin@HPOS	NUM	6600
TRAL	alto\Layout\Page\BottomMargin@HPOS	NUM	0
Margin (vertical)			
DEAL	alto\Layout\Page\TopMargin@VPOS	NUM	0
DEAL	alto\Layout\Page\LeftMargin@VPOS	NUM	13
DEAL	alto\Layout\Page\RightMargin@VPOS	NUM	13
DEAL	alto\Layout\Page\BottomMargin@VPOS	NUM	1892
EUAL	alto\Layout\Page\TopMargin@VPOS	NUM	0
EUAL	alto\Layout\Page\LeftMargin@VPOS	NUM	224
EUAL	alto\Layout\Page\RightMargin@VPOS	NUM	224
EUAL	alto\Layout\Page\BottomMargin@VPOS	NUM	5432
F1AL	alto\Layout\Page\TopMargin@VPOS	NUM	0
F1AL	alto\Layout\Page\LeftMargin@VPOS	NUM	264
F1AL	alto\Layout\Page\RightMargin@VPOS	NUM	264
F1AL	alto\Layout\Page\BottomMargin@VPOS	NUM	1938
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin@VPOS	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\LeftMargin@VPOS	NUM	75
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\RightMargin@VPOS	NUM	75
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin@VPOS	NUM	2028
PPAL	alto\Layout\Page\TopMargin@VPOS	NUM	0
PPAL	alto\Layout\Page\LeftMargin@VPOS	NUM	27
PPAL	alto\Layout\Page\RightMargin@VPOS	NUM	
PPAL	alto\Layout\Page\BottomMargin@VPOS	NUM	3975
TRAL	alto\Layout\Page\TopMargin@VPOS	NUM	0
TRAL	alto\Layout\Page\LeftMargin@VPOS	NUM	1520
TRAL	alto\Layout\Page\RightMargin@VPOS	NUM	1520
TRAL	alto\Layout\Page\BottomMargin@VPOS	NUM	8128
Illustration			
SBAT	PcGts\Page\GraphicRegion\Coords@points	UID	1154,484 1154,485 1203,485 1203,486 1205,486 1205,487 [...]

Dimensions

Technical Definition

Provides the dimensions of a component of the image. See **measurement unit** for the specific unit used; this is usually "mm10".

Category Notes

These fields appear commonly in ALTO files.

Instantiations			
White space width			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@WIDTH	NUM	560
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@WIDTH	NUM	20
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@WIDTH	NUM	23
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@WIDTH	NUM	44
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\SP@WIDTH	NUM	0
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\SP@WIDTH	NUM	18
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\SP@WIDTH	NUM	132
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\SP@WIDTH	NUM	22.89
Hyphen width			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@WIDTH	NUM	32
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@WIDTH	NUM	26
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@WIDTH	NUM	17
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String\HYP@WIDTH	NUM	17
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String\HYP@WIDTH	NUM	17
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\HYP@WIDTH	NUM	11
Page height			
CAAL	alto\Layout\Page@HEIGHT	NUM	27112
DEAL	alto\Layout\Page@HEIGHT	NUM	1912
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Component\didl:Resource\dcx:zoning\dcx:coordinates@height	NUM	184
EUAL	alto\Layout\Page@HEIGHT	NUM	5697

EUME	METS:mets\mets:amdSec\mets:techMD\mets:mdWrap\mets:xmlData\mix:mix\mix:BasicImageInformation\\mix:BasicImageCharacteristics\mix:imageHeight	NUM	6729
F1AL	alto\Layout\Page@HEIGHT	NUM	2211
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page@HEIGHT	NUM	2061
PPAL	alto\Layout\Page@HEIGHT	NUM	3976
SBAT	PcGts\Page@ImageHeight	NUM	3624
TDAG	issue\page\pageid@height	NUM	5247
TRAL	alto\Layout\Page@HEIGHT	NUM	8233.8
Page width			
CAAL	alto\Layout\Page@WIDTH	NUM	20664
DEAL	alto\Layout\Page@WIDTH	NUM	1395
DEMP	didl:DIDL\didl:Item\didl:Item\didl:Component\didl:Component\didl:Re-source\dcx:zoning\dcx:coordinates@width	NUM	628
EUAL	alto\Layout\Page@WIDTH	NUM	4104
EUME	METS:mets\mets:amdSec\mets:techMD\mets:mdWrap\mets:xmlData\mix:mix\mix:BasicImageInformation\\mix:BasicImageCharacteristics\mix:imageWidth	NUM	4847
F1AL	alto\Layout\Page@WIDTH	NUM	1482
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page@WIDTH	NUM	1341
PPAL	alto\Layout\Page@WIDTH	NUM	2539
SBAT	PcGts\Page@imageWidth	NUM	2400
TDAG	issue\page\pageid@width	NUM	3701
TRAL	alto\Layout\Page@WIDTH	NUM	6745
Print space height			
CAAL	alto\Layout\Page\PrintSpace@HEIGHT	NUM	27112
DEAL	alto\Layout\Page\PrintSpace@HEIGHT	NUM	1879
EUAL	Ealto\Layout\Page\PrintSpace@HEIGHT	NUM	5208
F1AL	alto\Layout\Page\PrintSpace@HEIGHT	NUM	1697
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space@HEIGHT	NUM	1953
PPAL	alto\Layout\Page\PrintSpace@HEIGHT	NUM	0
TRAL	alto\Layout\Page\PrintSpace@HEIGHT	NUM	8233.89

Print space width			
CAAL	alto\Layout\Page\PrintSpace@WIDTH	NUM	20664
DEAL	alto\Layout\Page\PrintSpace@WIDTH	NUM	1344
EUAL	alto\Layout\Page\PrintSpace@WIDTH	NUM	3657
F1AL	alto\Layout\Page\PrintSpace@WIDTH	NUM	1482
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\Print-Space@WIDTH	NUM	1257
PPAL	alto\Layout\Page\PrintSpace@WIDTH	NUM	2538
TRAL	alto\Layout\Page\PrintSpace@WIDTH	NUM	6745
Margin height			
DEAL	alto\Layout\Page\TopMargin@HEIGHT	NUM	13
DEAL	alto\Layout\Page\LeftMargin@HEIGHT	NUM	1879
DEAL	alto\Layout\Page\RightMargin@HEIGHT	NUM	1879
DEAL	alto\Layout\Page\BottomMargin@HEIGHT	NUM	20
EUAL	alto\Layout\Page\TopMargin@HEIGHT	NUM	224
EUAL	alto\Layout\Page\LeftMargin@HEIGHT	NUM	5208
EUAL	alto\Layout\Page\RightMargin@HEIGHT	NUM	508
EUAL	alto\Layout\Page\BottomMargin@HEIGHT	NUM	2655
F1AL	alto\Layout\Page\TopMargin@HEIGHT	NUM	264
F1AL	alto\Layout\Page\LeftMargin@HEIGHT	NUM	1674
F1AL	alto\Layout\Page\RightMargin@HEIGHT	NUM	1674
F1AL	alto\Layout\Page\BottomMargin@HEIGHT	NUM	273
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin@HEIGHT	NUM	75
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\LeftMargin@HEIGHT	NUM	1953
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\RightMargin@HEIGHT	NUM	1953
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin@HEIGHT	NUM	33
PPAL	alto\Layout\Page\TopMargin@HEIGHT	NUM	27
PPAL	alto\Layout\Page\LeftMargin@HEIGHT	NUM	3948
PPAL	alto\Layout\Page\RightMargin@HEIGHT	NUM	3948
PPAL	alto\Layout\Page\BottomMargin@HEIGHT	NUM	0
TRAL	alto\Layout\Page\TopMargin@HEIGHT	NUM	1520
TRAL	alto\Layout\Page\LeftMargin@HEIGHT	NUM	8233.89

TRAL	alto\Layout\Page\RightMargin@HEIGHT	NUM	8233.89
TRAL	alto\Layout\Page\BottomMargin@HEIGHT	NUM	105.89
Margin width			
DEAL	alto\Layout\Page\TopMargin@WIDTH	NUM	1395
DEAL	alto\Layout\Page\LeftMargin@WIDTH	NUM	20
DEAL	alto\Layout\Page\RightMargin@WIDTH	NUM	31
DEAL	alto\Layout\Page\BottomMargin@WIDTH	NUM	1395
EUAL	alto\Layout\Page\TopMargin@WIDTH	NUM	4104
EUAL	alto\Layout\Page\LeftMargin@WIDTH	NUM	161
EUAL	alto\Layout\Page\RightMargin@WIDTH	NUM	286
EUAL	alto\Layout\Page\BottomMargin@WIDTH	NUM	4104
F1AL	alto\Layout\Page\TopMargin@WIDTH	NUM	1482
F1AL	alto\Layout\Page\LeftMargin@WIDTH	NUM	0
F1AL	alto\Layout\Page\RightMargin@WIDTH	NUM	0
F1AL	alto\Layout\Page\BottomMargin@WIDTH	NUM	1482
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin@WIDTH	NUM	1341
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\LeftMargin@WIDTH	NUM	42
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\RightMargin@WIDTH	NUM	42
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin@ WIDTH	NUM	1341
PPAL	alto\Layout\Page\TopMargin@WIDTH	NUM	2539
PPAL	alto\Layout\Page\LeftMargin@WIDTH	NUM	2538
PPAL	alto\Layout\Page\RightMargin@WIDTH	NUM	0
PPAL	alto\Layout\Page\BottomMargin@WIDTH	NUM	0
TRAL	alto\Layout\Page\TopMargin@WIDTH	NUM	6745
TRAL	alto\Layout\Page\LeftMargin@WIDTH	NUM	186.67
TRAL	alto\Layout\Page\RightMargin@WIDTH	NUM	145
TRAL	alto\Layout\Page\BottomMargin@WIDTH	NUM	6745
Text block height			
CAAL	alto\Layout\Page\PrintSpace\TextBlock@HEIGHT	NUM	116
DEAL	alto\Layout\Page\PrintSpace\TextBlock@HEIGHT	NUM	34
EUAL	alto\Layout\Page\PrintSpace\TextBlock@HEIGHT	NUM	56
F1AL	alto\Layout\Page\PrintSpace\TextBlock@HEIGHT	NUM	578

F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock@HEIGHT	NUM	75
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock@HEIGHT	NUM	698
PPAL	alto\Layout\Page\PrintSpace\TextBlock@HEIGHT	NUM	357
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock@HEIGHT	NUM	2660
Text block width			
CAAL	alto\Layout\Page\PrintSpace\TextBlock@WIDTH	NUM	208
DEAL	alto\Layout\Page\PrintSpace\TextBlock@WIDTH	NUM	524
EUAL	alto\Layout\Page\PrintSpace\TextBlock@WIDTH	NUM	365
F1AL	alto\Layout\Page\PrintSpace\TextBlock@WIDTH	NUM	957
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock@WIDTH	NUM	13415
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock@WIDTH	NUM	621
PPAL	alto\Layout\Page\PrintSpace\TextBlock@HEIGHT	NUM	2507
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock@HEIGHT	NUM	753.33
Text line height			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HEIGHT	NUM	116
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HEIGHT	NUM	32
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HEIGHT	NUM	53
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HEIGHT	NUM	24
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine@HEIGHT	NUM	24
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine@HEIGHT	NUM	38
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@HEIGHT	NUM	1
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine@HEIGHT	NUM	56.67
Text line width			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@WIDTH	NUM	208
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@WIDTH	NUM	518
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@WIDTH	NUM	364
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@WIDTH	NUM	13

F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine@WIDTH	NUM	52
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine@WIDTH	NUM	609
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine@WIDTH	NUM	2507
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine@WIDTH	NUM	15.6
String height			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HEIGHT	NUM	116
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HEIGHT	NUM	30
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HEIGHT	NUM	53
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HEIGHT	NUM	24
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String@HEIGHT	NUM	24
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@HEIGHT	NUM	25
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@HEIGHT	NUM	1
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\String@HEIGHT	NUM	56.67
String width			
CAAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WIDTH	NUM	208
DEAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WIDTH	NUM	227
EUAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WIDTH	NUM	364
F1AL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WIDTH	NUM	13
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\TopMargin\TextBlock\TextLine\String@WIDTH	NUM	52
F2AL	pageOCRDATA\content\altoXML\alto\Layout\Page\BottomMargin\PrintSpace\TextBlock\TextLine\String@WIDTH	NUM	68
PPAL	alto\Layout\Page\PrintSpace\TextBlock\TextLine\String@WIDTH	NUM	923
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock\TextBlock\TextLine\String@HEIGHT	NUM	56.1
Article zone height			
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@HEIGHT	NUM	3429
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock@HEIGHT	NUM	6608
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@HEIGHT	NUM	1222.67

Article zone width			
PPAL	alto\Layout\Page\PrintSpace\ComposedBlock@WIDTH	NUM	644
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock@WIDTH	NUM	2445.33
TRAL	alto\Layout\Page\PrintSpace\ComposedBlock\ComposedBlock@WIDTH	NUM	753.33

Measurement Unit

Technical Definition

Provides the measure unit for all values except the **font size**.

Usage Notes

The default value is 1/10 of mm.

Category Notes

This field is common in ALTO files. For all, string-choice: “pixel”, “mm10” or “inch1200”.

Instantiations			
CAAL	alto\Description\MeasurementUnit	MCH	inch1200
DEAL	alto\Description\MeasurementUnit	MCH	pixel
EUAL	alto\Description\MeasurementUnit	MCH	mm10
F1AL	alto\Description\MeasurementUnit	MCH	mm10
F2AL	pageOCRDATA\content\altoXML\alto\Description\MeasurementUnit	MCH	mm10
PPAL	alto\Description\MeasurementUnit	MCH	mm10
TRAL	alto\Description\MeasurementUnit	MCH	pixel

Font Information

Technical Definition

Provides information about the font of the text, as recognised by the OCR software. This includes font size, font style (whether bold, italics, underlined, small caps, etc.), font type (whether serif or sans serif), font width (whether proportional or fixed), and font family.

Usage Notes

It is unlikely that the font registered will accurately reflect the specific typeface used, though the weight, style and size may be more accurate. Font styles and sizes in the nineteenth century often reflected something about the genre of the newspaper. Most newspapers would use different typefaces for the masthead, adverts and editorials, and some might choose to use a nondescript font for ease. As technology made it easier to standardise, these decisions became formalised. OCR software has difficulty with older font styles and will not always recognise them correctly.

Category Notes

Font size, style and family are quite common fields in ALTO files. Font type and font width appear in CAAL only.

Individual Collection Notes

CAAL: The font size is in points (1/72 of an inch). Font style: list of any combination of font styles as a string choice: "bold", "italics", "subscript", "superscript", "smallcaps", "underline". Font type: String choice of "serif" or "sans-serif". Font width: String choice of "proportional" or "fixed".

DEAL: Font style: list of any combination of font styles. String-Choice: "bold", "italics", "subscript", "superscript", "smallcaps", "underline".

EUAL: Font style: list of any combination of font styles. String-Choice: "bold", "italics", "subscript", "superscript", "smallcaps", "underline".

F1AL: Font style: list of any combination of font styles. String-Choice: "bold", "italics", "subscript", "superscript", "smallcaps", "underline".

F2AL: Though labelled Font Style, font size contains numeric data.

PPAL: Font style: list of any combination of font styles. String-Choice: "bold", "italics", "subscript", "superscript", "smallcaps", "underline".

Instantiations			
Font size			
CAAL	alto\Styles\TextStyle@FONTSIZE	NUM	10
DEAL	alto\Styles\TextStyle@FONTSIZE	NUM	9
EUAL	alto\Styles\TextStyle@FONTSIZE	NUM	15
F1AL	alto\Style\TextStyle@FONTSIZE	NUM	15
F2AL	pageOCRDATA\content\altoXML\alto\Styles\TextStyle@FONTSTYLE	NUM	9
PPAL	alto\TextStyle@FONTSIZE	NUM	10
TRAL	alto\Styles\TextStyle@FONTSIZE	NUM	7.7
Font style			
CAAL	alto\Styles\TextStyle@FONTSTYLE	STR	small caps
DEAL	alto\Styles\TextStyle@FONTSTYLE	STR	bold
EUAL	alto\Styles\TextStyle@FONTSTYLE	STR	bold
F1AL	alto\Style\TextStyle@FONTSTYLE	STR	bold
PPAL	alto\TextStyle@FONTSTYLE	STR	italics

Font type			
CAAL	alto\Styles\TextStyle\FONTTYPE1	MCH	Serif
Font width			
CAAL	alto\Styles\TextStyle\FONTWIDTH	MCH	proportional
Font family			
DEAL	alto\Styles\TextStyle@FONTFAMILY	STR	Times New Roman
EUAL	alto\Styles\TextStyle@FONTFAMILY	STR	Fraktur
F1AL	alto\Style\TextStyle@FONTFAMILY	STR	Times New Roman
F2AL	pageOCRDATA\content\altoXML\alto\Styles\TextStyle@FONTFAMILY	STR	Fraktur
PPAL	alto\TextStyle@FONTFAMILY	STR	Times New Roman

Page Skew

Technical Definition

Provides the degree to which the page image is skewed from the perpendicular.

Usage Notes

The skewing may be due to the scanned newspapers being part of a bound **volume**. As newspaper issues were intended to be printed and consumed quickly rather than kept and revisited, publishing volumes offered readers a more expensive, lasting edition of the articles. Sometimes, the new volumes would include illustrations. Many digitisers rely on bound volumes of many newspaper issues, as the volume format better preserves them. For the British Library collections, in some cases volumes that were tightly bound have had to be partially disbound and then shrink-wrapped and returned to the shelves in a partially bound condition for **microfilming**. The physical items are bound in volumes, so there is not a one-to-one relationship between the physical object, containing many issues, and the digital object. For Trove, some state libraries have used formerly bound copies of newspapers, in order to get the best quality for digitisation. The State Library of South Australia disbinds its copies when **microfilming**, as does the Library of Victoria, or they loosen the bindings during the microfilming process in order to deal with margins and so on. Other digitisers indicate that they do not usually disbind, as it is not usually possible to rebind them, and this threatens the physical object. In those cases, the pages can be harder to read as the binding interferes with the scanning process.

Category Notes

This field appears in B1JI only.

Instantiations			
B1JI	BL_newspaper\BL_page\pageImage\pageSkew	NUM	-10

Metadata Type

Technical Definition

Defines the metadata type.

Usage Notes

For many of the collections, this is a standard format: when labelling descriptive metadata, MODS or MET is used, and for administrative metadata NISOIMG is used.

Category Notes

This field is common in METS files.

Instantiations			
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	NISOIMG
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	NISO still image metadata
CAME	mets\dmdSec\mdWrap@MDTYPE	STR	MET
CAME	mets\dmdSec\mdWrap@LABEL	STR	Issue metadata
CAME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
CAME	mets\dmdSec\mdWrap@LABEL	STR	Page metadata
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	PREMIS object metadata
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	OTHER
CAME	mets\amdSec\techMD\mdWrap@OTHERMDTYPE	STR	PREMIS
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	NISO still image metadata
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	NISOIMG
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	OTHER
CAME	mets\amdSec\techMD\mdWrap\mdWrap@OTHERMDTYPE	STR	PREMIS
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	PREMIS object metadata
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	PREMIS object metadata
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	OTHER
CAME	mets\amdSec\techMD\mdWrap@OTHERMDTYPE	STR	PREMIS
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	PREMIS object metadata
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	OTHER
CAME	mets\amdSec\techMD\mdWrap@OTHERMDTYPE	STR	PREMIS
CAME	mets\amdSec\techMD\mdWrap@LABEL	STR	NISO still image metadata
CAME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	NISOIMG

EUME	METS:mets\mets:dmdSec\mets:mdWrap@LABEL	STR	Bibliographic meta-data of the printed version
EUME	METS:mets\mets:dmdSec\mets:mdWrap@MDTYPE	STR	MODS
EUME	METS:mets\mets:dmdSec\mets:mdWrap@LABEL	STR	Bibliographic meta-data of the issue
EUME	METS:mets\mets:dmdSec\mets:mdWrap@MDTYPE	STR	MODS
EUME	METS:mets\mets:dmdSec\mets:mdWrap@LABEL	STR	Bibliographic meta-data of article
EUME	METS:mets\mets:dmdSec\mets:mdWrap@MDTYPE	STR	MODS
EUME	METS:mets\mets:dmdSec\mets:mdWrap@LABEL	STR	Bibliographic meta-data of table
EUME	METS:mets\mets:dmdSec\mets:mdWrap@MDTYPE	STR	MODS
EUME	METS:mets\mets:amdSec\mets:techMD\mets:mdWrap@MDTYPE	STR	NISOIMG
F1ME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	NISOIMG
F1ME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
F1ME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of the printed version
F1ME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of the electronic version
F1ME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
F1ME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of Issue
F1ME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
F1ME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of list
F1ME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
PPME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of the printed version
PPME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
PPME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of the electronic version
PPME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
PPME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of Issue
PPME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
PPME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of article
PPME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS

PPME	mets\dmdSec\mdWrap@LABEL	STR	Bibliographic meta-data of section
PPME	mets\dmdSec\mdWrap@MDTYPE	STR	MODS
PPME	mets\amdSec\techMD\mdWrap@MDTYPE	STR	NISOIMG
SBME	mets:mets\mets:dmdSec\mets:mdWrap@MDTYPE	STR	MODS
TRME	mets:mets\mets:dmdSec\mets:mdWrap@MDTYPE	STR	MODS
TRME	mets:mets\mods:dmdSec\mods:mdWrap@MDTYPE	STR	MODS
TRME	Tmets:mets\mets:amdSec\mets:techMD\mets:mdWrap@MDTYPE	STR	PREMIS
TRME	mets:mets\mets:amdSec\mets:techMD\mets:mdWrap@MDTYPE	STR	PREMIS
TRME	mets:mets\mets:amdSec\mets:digiprovMD\mets:mdWrap@MDTYPE	STR	PREMIS
TRME	mets:mets\mets:amdSec\mets:digiprovMD\mets:mdWrap@MDTYPE	STR	PREMIS

Bibliography

Abruzzi, Ray, Luisa Calè and Ana Parejo Vadillo. "Gale Digital Collections: Ray Abruzzi Interviewed by Luisa Calè and Ana Parejo Vadillo." *19: Interdisciplinary Studies in the Long Nineteenth Century*, vol. 21, 2015, n.pag., doi:10.16995/ntn.753.

"Access." National Library of Australia, 2014. Web. https://www.nla.gov.au/australian-newspaper-plan/about/access. Accessed 29 November 2019.

"Access Policy." *Strategy and Policy*. National Library of New Zealand, 2003. Web. https://natlib.govt.nz/about-us/strategy-and-policy/access-policy. Accessed 29 November 2019.

Adams, David. "Papers Past: Browse Access for Online New Zealand Newspapers." *Microform & Digitization Review*, vol. 34, no. 1, 2007, doi:10.1515/MFIR.2005.22.

Alam, Sultana Lubna and John Campbell. "Crowdsourcing Motivations in a Not-for-Profit GLAM Context: The Australian Newspapers Digitisation Program." *ACIS 2012: Location, Location, Location: Proceedings of the 23rd Australasian Conference on Information Systems*. ACIS, 2012. 1–11. http://hdl.handle.net/10536/DRO/DU:30049107.

"Alexander Turnbull Library Collections." *A–Z of our Collections*. National Library of New Zealand. National Library of New Zealand. Web. https://natlib.govt.nz/collections/a-z/alexander-turnbull-library-collections. Accessed 29 November 2019.

Allingham, Philip V. "Nineteenth-Century British and American Copyright Law." *The Victorian Web*, 10 August 2013. Web. www.victorianweb.org/authors/dickens/pva/pva74.html. Accessed 4 December 2019.

"Antecedentes." *Quiénes somos*. Universidad Nacional Autónoma de México, 2018. Web. www.hnm.unam.mx/index.php/quienes-somos/antecedentes. Accessed 29 November 2019.

"API – Newspaper Article Record Structure." Trove. Web. https://trove.nla.gov.au/general/api-record-article. Accessed 4 December 2019.

"API Version 2 Technical Guide." *Trove Help Centre*. National Library of Australia, 2019. Web. https://help.nla.gov.au/trove/building-with-trove/api-version-2-technical-guide. Accessed 29 November 2019.

Aspinall, A. "The Social Status of Journalists at the Beginning of the Nineteenth Century." *The Review of English Studies*, vol. 21, no. 83, 1945, pp. 216–32. JSTOR, www.jstor.org/stable/509176.

---. "The Circulation of Newspapers in the Early Nineteenth Century." *The Review of English Studies*, vol. 22, no. 85, 1946, pp. 29–43. JSTOR, www.jstor.org/stable/509326.

"The Atlas." Oceanic Exchanges, 2019. Web. https://www.digitisednewspapers.net. Accessed 29 November 2019.

Ayres, Marie-Louise. "'Singing for their supper': Trove, Australian Newspapers, and the Crowd." *IFLA WLIC 2013, Singapore*. 31 July 2013. http://library.ifla.org/245/1/153-ayres-en.pdf. Accessed 1 December 2019.

---. "Trove at 5: Are We There Yet." National Library of Australia, 2014. Web. https://www.nla.gov.au/our-publications/staff-papers/trove-at-5are-we-there-yet. Accessed 1 December 2019.

---. "Trove and the World: Cultural Collections in a Global Environment – A Presentation at ALIA Online 2015, 5 February 2015." National Library of Australia, 5 February 2015. Web. https://www.nla.gov.au/our-publications/staff-papers/trove-and-the-world. Accessed 2 December 2019.

Bagnall, Kate. "Trove Presents a New Perspective on Australian History." National Library of Australia, 19 February 2015. Web. https://www.nla.gov.au/stories/blog/trove/2015/02/19/early-chinese-newspapers. Accessed 29 November 2019.

Barrall, Kelly and Christine Guenther. *Microfilm Selection for Digitization*. OCLC Preservation Service Centers, 17 May 2005. http://www.loc.gov/ndnp/guidelines/NEH_MicrofilmSelectionNDNP.pdf. Accessed 6 December 2019.

Bateson, F. W. "The Errata in The Tatler." *The Review of English Studies*, vol. 5, no. 18, 1929, pp. 155–66, doi:10.1093/res/os-V.18.155.

Beals, M. H. "Close Readings of Big Data: Triangulating Patterns of Textual Reappearance and Attribution in the Caledonian Mercury, 1820–40." *Victorian Periodicals Review*, vol. 51, no. 4, 2018, pp. 616–39, doi:10.1353/vpr.2018.0046.

Beals, M. H. and Emily Bell (eds.). *Map of Digitised Newspaper Metadata* v.1.0.0. Dataset. Figshare. doi:10.6084/m9.figshare.11560110. Accessed 3 March 2020.

Beals, M. H. and Emily Bell (eds.). *Map of Digitised Newspaper Metadata (Dynamic)*. Dataset. Github. https://github.com/AtlasOfDigitisedNewspapers/AtlasOfDigitisedNewspapers.github.io/tree/master/downloads. Accessed 3 March 2020.

Beekman, Grace. "Emotional Density, Suspense, and the Serialization of The Woman in White in All the Year Round." *Victorian Periodicals Review*, vol. 50 no. 1, 2017, pp. 9–38, doi:10.1353/vpr.2017.0002.

Beetham, Margaret. "Open and Closed: The Periodical as a Publishing Genre." *Victorian Periodicals Review*, vol. 22, no. 3, 1989, pp. 96–100. JSTOR, www.jstor.org/stable/20082400.

Bent, Andrew. *An Appeal to the Sympathies and Benevolence of the Australasian Public, For Relief for Mr. Andrew Bent, (Late Newspaper Proprietor of Van Diemen's Land,) And His Large Family, Now in a State of Utter Destitution!* Andrew Bent, 1844.

Bernstein, S.D. "Reading Room Geographies of Late-Victorian London: The British Museum, Bloomsbury and the People's Palace, Mile End." *19: Interdisciplinary Studies in the Long Nineteenth Century*, vol. 13, 2012, n.pag., doi:10.16995/ntn.632.

Berthon, Hilary and Julia Hickie. "Digitising Journals on Trove: A National Approach to Sharing Content, Engaging Communities and Collaboration in the Digital World." National Library of Australia, 22 March 2018. Web. https://www.nla.gov.au/our-publications/staff-papers/digitising-journals-on-trove-a-national-approach-to-sharing-content. Accessed 2 December 2019.

Boehm, Katharina. "'A Place for More than the Healing of Bodily Sickness': Charles Dickens, the Social Mission of Nineteenth-Century Pediatrics, and the Great Ormond Street Hospital for Sick Children." *Victorian Review*, vol. 35 no. 1, 2009, pp. 153–74, doi:10.1353/vcr.2009.0031.

Booth, Keitha and Julia Napier. "Linking People and Information: Website Access to National Library of New Zealand Information and Services." *The Electronic Library*, vol. 21, no. 3, pp. 227–33, 2003, doi:10.1108/02640470310480470.

Boyce, George, James Curran and Pauline Wingate. *Newspaper History from The Seventeenth Century to the Present Day*. Constable, 1978.

Brake, Laurel, Chandrika Kaul and Mark W. Turner. *The 'News of the World' and the British Press, 1843–2011*. Palgrave, 2016.

Brake, Laurel, Aled Jones and Lionel Madden (eds.). *Investigating Victorian Journalism*. Palgrave, 1990.

Brake, Laurel and Marysa Demoor (eds.). *Dictionary of Nineteenth-Century Journalism in Great Britain and Ireland*. Academia Press and the British Library, 2009.

Brake, Laurel. "London Letter: Researching the Historical Press, Now and Here." *Victorian Periodicals Review*, vol. 48 no. 2, 2015a, pp. 245–53, doi:10.1353/vpr.2015.0028.

---. "Nineteenth-Century Newspaper Press Directories: The National Gallery of the British Press." *Victorian Periodicals Review*, vol. 48 no. 4, 2015b, pp. 569–90, doi:10.1353/vpr.2015.0055.

Branch, Edgar M. "Mark Twain: Newspaper Reading and the Writer's Creativity." *Nineteenth-Century Fiction*, vol. 37 no. 4, 1983, pp. 576–603, doi:10.2307/3044684.

Bremer-Laamanen, Majlis. "Connecting to the Past – Newspaper Digitisation in the Nordic Countries." *International Newspaper Librarianship for the 21st Century*. Ed. Hartmut Walravens. K.G. Saur, 2006. 45–50.

---. "The Present Past: The History of Newspaper Digitisation in Finland." *Newspapers of the World Online: U. S. and International Perspectives: Proceedings of Conferences in Salt Lake City and Seoul, 2006*. Ed. Hartmut Walravens. De Gruyter, 2009. 43–48.

British Library Newspapers Part I: 1800–1900. Database. Gale: A Cengage Company, 2013. https://www.gale.com/intl/c/british-library-newspapers-part-i. Accessed 29 November 2019.

British Library Newspapers Part II: 1800–1900. Database. Gale: A Cengage Company, 2018. https://www.gale.com/intl/c/british-library-newspapers-part-ii. Accessed 29 November 2019.

The British Library. Sustaining the Value: The British Library Digital Preservation Strategy 2017-2020. British Library, 2017. https://www.bl.uk/britishlibrary/~/media/bl/global/digital preservation/bl_digitalpreservationstrategy_2017-2020.pdf. Accessed 28 November 2019.

The British Newspaper Archive. Database. Findmypast Newspaper Archive Ltd., 2019. https://www.britishnewspaperarchive.co.uk. Accessed 29 November 2019.

Brown, Lucy. *Victorian News and Newspapers*. Clarendon Press, 1985.

Bryce, Catriona. "Trove – A Brief History." National Library of Australia, 6 November 2014. Web. https://www.nla.gov.au/blogs/trove/2014/11/06/trove-a-brief-history. Accessed 1 December 2019.

---. "The Teams Behind Trove – Part 1." National Library of Australia, 17 November 2014. Web. https://www.nla.gov.au/blogs/trove/2014/11/19/the-teams-behind-trove-part-1. Accessed 2 December 2019.

---. " The Teams Behind Trove – Part 2." National Library of Australia, 28 November 2014. Web. https://www.nla.gov.au/blogs/trove/2014/11/28/the-teams-behind-trove-part-2. Accessed 1 December 2019.

---. "Trove and Social Media Today: Does a Click-Through Mean What We Think It Means?" National Library of Australia, 23 February 2016. Web. https://www.nla.gov.au/our-publications/staff-papers/trove-and-social-media-today-does-a-click-through-mean-what-we-think. Accessed 2 December 2019.

Budd, Jennifer (ed.). *Proceedings of the IFLA Symposium: Managing the Preservation of Periodicals and Newspapers*. K. G. Saur, 2002.

"Building with Trove." Trove. Web. https://help.nla.gov.au/trove/building-with-trove. Accessed 4 December 2019.

Bull, Sarah. "A Purveyor of Garbage?: Charles Carrington and the Marketing of Sexual Science in Late-Victorian Britain." *Victorian Review*, vol. 38, no. 1, 2012, pp. 55–76, doi:10.1353/vcr.2012.0028.

Bunting, Kirsty. "'Thank God for the Public Press, Which Sheds Its Strong White Light on All the Dark Corners of the Earth!': Ada Nield Chew's Journey from Factory Girl to Author-Activist." *Victorians: A Journal of Culture and Literature,* vol. 132, 2017, pp. 134–47, doi:10.1353/vct.2017.0013.

Butterfield, Andrew, Gerard Ekembe Ngondi and Anne Kerr (eds.). *A Dictionary of Computer Science*. 7th Ed. Oxford University Press, 2016.

Centre for Research Libraries. *The "State of the Art": A Comparative Analysis of Newspaper Digitization to Date*. Centre for Research Libraries, 2015. https://www.crl.edu/sites/default/files/d6/attachments/events/ICON_Report-State_of_Digitization_final.pdf. Accessed 27 November 2019.

Chronicling America: Historic American Newspapers. Database. https://chroniclingamerica.loc.gov. Library of Congress, 2017. Accessed 29 November 2019.

Cogdill, Sharon E. "'As it was not ungrammatical, though of a chatty tendency, it seemed to please': Lady Violet Greville, 'Aristocratic Lady Journalist' of the 1890s." *Victorians: A Journal of Culture and Literature*, vol. 132, 2017, pp. 175–88, doi:10.1353/vct.2017.0016.

"Collaborative Digitisation." *National Library Services*. National Library of New Zealand, 2018. Web. https://natlib.govt.nz/librarians/national-library-services/collaborative-digitisation. Accessed 29 November 2019.

"Content Selection." *Guidelines & Resources*. Library of Congress, 8 May 2016. Web. http://www.loc.gov/ndnp/guidelines/selection.html. Accessed 6 December 2019.

Cooper, Elena. "Copyright in Periodicals during the Nineteenth Century: Balancing the Rights of Contributors and Publishers." *Victorian Periodicals Review*, vol. 51 no. 4, 2018, pp. 661–78, doi:10.1353/vpr.2018.0048.

"Copyright and Privacy." *About this Site*. National Library of New Zealand. Web. https://natlib.govt.nz/about-this-site/copyright-and-privacy. Accessed 29 November 2019.

Cordell, Ryan. "'Q i-jtb the Raven': Taking Dirty OCR Seriously." *Book History*, vol. 20, 2017, pp. 188–225, doi:10.1353/bh.2017.0006.

Coulling, Sidney M. B. "Matthew Arnold and *The Daily Telegraph*." *The Review of English Studies*, vol. 12, no. 46, 1961, pp. 173–79, doi:10.1093/res/XII.46.173.

Cranfield, Geoffrey Alan. *The Press and Society: From Caxton to Northcliffe*. Routledge, 1978.

Cronin, Anne M. "Rags and Refuse: The Newspaper, Empire, and Nineteenth Century Commodity Culture." *Cultural Studies*, vol. 20, no. 6, 2006, pp. 574–98, doi:10.1080/09502380600973952.

Curiel Defossé, Guadalupe and Ricardo Javier Jiménez Rivera. "Diez años de la Hemeroteca Nacional Digital de México (HNDM). Breve reseña de una larga gestión de preservación y acceso, 2002–2012." *Textos, pixeles y bits: Reflexiones sobre la publicación digital*. Ed. Isabel Galina Russell. Universidad Nacional Autónoma de México, 2015. 65–149.

"Data Quality Committee." *Project*. Europeana Pro, 2016. Web. https://pro.europeana.eu/project/data-quality-committee. Accessed 29 November 2019.

Delpher. Database. National Library of the Netherlands, 2018. https://www.delpher.nl/. Accessed 29 November 2019.

Delpher. Overzicht beschikbare titels. Delpher, 2017. https://www.kb.nl/sites/default/files/ docs/Beschikbare_ kranten_alfabetisch.pdf. Accessed 2 December 2019.

Demoor, Marysa. "When the King Becomes your Personal Enemy: W. T. Stead, King Leopold II, and the Congo Free State." *19: Interdisciplinary Studies in the Long Nineteenth Century*, vol. 16, 2013, n.pag., doi:10.16995/ ntn.662.

DiCenzo, Maria. "Pressing the Public: Nineteenth-Century Feminist Periodicals and 'the Press'." *Nineteenth Century Gender Studies*, vol. 6, no. 2, 2010, n.pag., http://www.ncgsjournal.com/issue62/dicenzo.htm. Accessed 27 November 2019.

Dickens, Charles. *The Letters of Charles Dickens*. Pilgrim Edition. General Eds. Madeline House, Graham Storey and Kathleen Tillotson. 12 vols. Clarendon Press, 1965–2002.

"Digging into Data Challenge." *Trans-Atlantic Platform: Social Sciences and Humanities*. Digging into Data, 2017. Web. https://diggingintodata.org. Accessed 29 November 2019.

"DigitalNZ API." *Open Data*. National Library of New Zealand. Web. https://natlib.govt.nz/about-us/open-data/ digitalnz-api. Accessed 29 November 2019.

"Digitisation Workflow Process Overview." *Trove Help Centre*. National Library of Australia. Web. https://help. nla.gov.au/trove/for-digitisation-partners/digitisation-workflow-process-overview. Accessed 2 December 2019.

Djerf-Pierre, Monika and Lennart Weibull. "From Idealist-Entrepreneur to Corporate Executive: Provincial Newspaper Editors' and Publishers' Ways-of-Thinking from the Mid-1800s to the Present." *Journalism Studies*, vol. 12, no. 3, pp. 294–310, doi:10.1080/1461670X.2010.490655.

Droge, Abigail. "'Always Called Jack': A Brief History of the Transferable Skill." *Victorian Periodicals Review*, vol. 50 no. 1, 2017, pp. 39–65, doi:10.1353/vpr.2017.0003.

"Early American Newspapers, 1690–1922." *America's Historical Newspapers*. Readex, 2013. Web. https://www.readex.com/content/early-american-newspapers-1690-1922. Accessed 29 November 2019.

Easley, Alexis. "Chance Encounters, Rediscovery, and Loss: Researching Victorian Women Journalists in the Digital Age." *Victorian Periodicals Review*, vol. 49 no. 4, 2016a, pp. 694–717, doi:10.1353/vpr.2016.0045.

---. "Imagining the Mass-Market Woman Reader: The News of the World, 1843–77." *The 'News of the World' and the British Press*, 1843–2011. Eds. Laurel Brake, Chandrika Kaul and Mark W. Turner. Palgrave Macmillan, 2016b. 81–99.

Edson C. Tandoc Jr., Zheng Wei Lim and Richard Ling. "Defining 'Fake News'." *Digital Journalism*, vol. 6, no. 2, pp. 137–53, doi:10.1080/21670811.2017.1360143.

"Europeana Data Model." *Resources*. Europeana Pro, 2013. Web. https://pro.europeana.eu/resources/ standardization-tools/edm-documentation. Accessed 29 November 2019.

Europeana Newspapers. Database. Europeana Collections, 2018. https://www.europeana.eu/portal/en/ collections/newspapers. Accessed 29 November 2019.

Europeana Newspapers. *D5.3 Final public release with updated online resource for documentation*. Europeana Newspapers, 2015. http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D5.3_Final_ release_ENMAP_1.0.pdf. Accessed 27 November 2019.

"Europeana Newspapers is making Historic Newspaper Pages Searchable." *About*. Europeana Newspapers, 2018. Web. www.europeana-newspapers.eu. Accessed 29 November 2019.

Evans, Harold. *Editing and Design: A Five-Volume Manual of English, Typography and Layout*. *Book 5: Newspaper Design*. Heinemann, 1973.

Feather, John and Paul Sturges (eds.). *International Encyclopedia of Information and Library Science*. Routledge, 2003.

Ferguson, J. A., A. G. Foster and H. M. Green. *The Howes and their Press*. Sunnybrook Press, 1936.

Fieldberg, Allison. "1834: 'Melancholy and Affecting Accident: Eleven Persons Drowned'." *Victorian Review*, vol. 40 no. 2, 2014, pp. 13–17, doi:10.1353/vcr.2014.0045.

Franklin, Bob (ed.). *Pulling Newspapers Apart: Analysing Print Journalism*. Routledge, 2008.

Fritze, Ronald H., Brian E. Coutts and Louis Andrew Vyhnanek (eds.). *Reference Sources in History: An Introductory Guide*. Modern Language Association of America, 2003.

Fyfe, Paul. "An Archaeology of Victorian Newspapers." *Victorian Periodicals Review*, vol. 49, no. 4, 2016, pp. 546–77, doi:10.1353/vpr.2016.0039.

---. "Access, Computational Analysis, and Fair Use in the Digitized Nineteenth-Century Press." *Victorian Periodicals Review*, vol. 51, no. 4, 2018, pp. 716–37, doi:10.1353/vpr.2018.0051.

Gabriele, Sandra. "Transfiguring the Newspaper: From Paper to Microfilm to Database." *Amodern 2: Network Archaeology*, 2014, n.pag., doi:http://amodern.net/article/transfiguring-the-newspaper/.

Gale Digital Scholar Lab. Database. Gale: A Cengage Company, 2019. https://www.gale.com/intl/primary-sources/digital-scholar-lab. Accessed 29 November 2019.

"Gale Historical Newspapers." *Primary Sources*. Gale: A Cengage Company, 2019. Web. https://www.gale.com/intl/primary-sources/historical-newspapers. Accessed 29 November 2019.

Garside, Peter. "Scott as a Political Journalist." *The Review of English Studies*, vol. 37, no. 148, 1986, pp. 503–17, doi:10.1093/res/XXXVII.148.503.

Gaskell, Beth. "Multi-Title Newspaper Volumes – Historic Practice vs. Modern Process." *The Newsroom Blog*. British Library, 16 January 2019. Web. https://blogs.bl.uk/thenewsroom/2019/01/multi-title-newspaper-volumes-historic-practice-vs-modern-process.html. Accessed 20 December 2019.

Gelmi, Caroline. "'The Pleasures of Merely Circulating': Sappho and Early American Newspaper Poetry." *Nineteenth-Century Literature*, vol. 69, no. 2, 2014, pp. 151–74. JSTOR, www.jstor.org/stable/10.1525/ncl.2014.69.2.151.

Gillies, Stewart. "History of British Library Newspapers." *British Library*, 2008. Web. http://www.bl.uk/reshelp/findhelprestype/news/historicalblnews/. Accessed 28 January 2016.

Gilmore, Dehn. "'These verbal puzzles': Wilkie Collins, Newspaper Enigmas, and the Victorian Reader as Solver." *Victorian Literature and Culture*, vol. 44, no. 2, 2016, pp. 297–314, doi:10.1017/S1060150315000637.

Gooding, Paul. *Historic Newspapers in the Digital Age: "Search All About It!"* Routledge, 2017.

Gray, F. Elizabeth. "Alice Meynell: Production, Reproduction, and the World of Work in the 1890s." *Victorians: A Journal of Culture and Literature*, vol. 132, 2017, pp. 148–61, doi:10.1353/vct.2017.0014.

"Guide to Selected Collections." Collections. National Library of Australia. Web. https://www.nla.gov.au/our-collections/guide-to-selected-library-collections. Accessed 29 November 2019.

Hagon, Paul. "Trove Crowdsourcing Behaviour." National Library of Australia, 12 February 2013. Web. https://www.nla.gov.au/our-publications/staff-papers/trove-crowdsourcing-behaviour. Accessed 2 December 2019.

Hampton, Mark. *Visions of the Press in Britain, 1850–1950*. University of Illinois Press, 2004.

Harvey, Ross. "Selection of Newspapers for Digitization and Preservation: A User Perspective." *International Newspaper Librarianship for the 21st Century*. Ed. Hartmut Walravens. K. G. Saur, 2006. 103–13.

Hauswedell, Tessa, Julianne Nyhan, Melodee Beals, Melissa Terras and Emily Bell. "Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives of Historical Newspapers." *Archival Science: International Journal on Recorded Information*, 2020, doi: 10.1007/s10502-020-09332-1.

Hellema, Marian and René Voorburg. *Searching & Harvesting Delpher Newspapers.* Koninklijke Bibliotheek, 2015.

Hellema, Marian. *APIs to the Historical Newspaper Collection*. Koninklijke Bibliotheek, 2015.

Hemeroteca Nacional Digital de México. Database. Universidad Nacional Autónoma de México, 2018. http://www.hndm.unam.mx. Accessed 29 November 2019.

"Hemeroteca Nacional Digital de México." Hemeroteca Nacional Digital de México, 2015. Web. http://www.hndm.unam.mx/index.php/es/acerca-de-hndm. Accessed 29 November 2019.

"Hemeroteca Nacional Digital de México (HNDM)." *Fondo Reservado*. Hemeroteca Nacional Digital de México, 2018. Web. http://www.hnm.unam.mx/index.php/hemeroteca-nacional-de-mexico/colecciones/fondo-reservado/hemeroteca-nacional-digital-mexico. Accessed 29 November 2019.

Hickie, Julia and Mark Raadgever. "Easy as ABC – A Triumph of Re-Usable Metadata." National Library of Australia, 11 February 2015. Web. https://www.nla.gov.au/our-publications/staff-papers/easy-as-abc. Accessed 2 December 2019.

"Historic Newspapers Live on Through Groundbreaking Project." Europeana Pro, 2015. Web. https://pro.europeana.eu/post/historic-newspapers-live-on-through-groundbreaking-project. Accessed 29 November 2019.

"History of Readex: From Microprint to Digital." *About*. Readex, 2012. Web. https://www.readex.com/history-readex. Accessed 29 November 2019.

Hobbs, Andrew. "The Deleterious Dominance of *The Times* in Nineteenth-Century Scholarship." *Journal of Victorian Culture*, vol. 18, no. 4, 2013, pp. 472–97, doi:10.1080/13555502.2013.854519.

---. *A Fleet Street in Every Town: The Provincial Press in England, 1855–1900*. Open Book Publishers, 2018.

Hogan, Paul. "Trove Crowdsourcing Behaviour." National Library of Australia, 2013. Web. https://www.nla.gov.au/sites/default/files/trove-crowdsourcing-behaviour.pdf. Accessed 1 December 2019.

Hughes, Linda K. "Periodical Poetry, Editorial Policy, and W. E. Henley's Scots and National Observer." *Victorian Periodicals Review*, vol. 49, no. 2, 2016, pp. 202–27, doi:https://doi.org/10.1353/vpr.2016.0014.

Hunt, Alyson. "Special Section: Women of the Press in the 1890s." *Victorians: A Journal of Culture and Literature*, vol. 132, 2017, pp. 130–33, doi:10.1353/vct.2017.0012.

IFLA. "Microfilming for Digitisation and Optical Character Recognition" *IFLA, 2002*. https://www.ifla.org/publications/microfilming-for-digitisaton-and-optical-character-recognition. Accessed 1 December 2019.

Ilva, Jyrki. "Kansallisen kulttuuriperinnön digitointi – loppuuko vauhti ennen alkua?" *Agricolan tietosanomat*, vol. 2, 2004, doi:https://www.doria.fi/handle/10024/152558.

Jackson, Alfred. "The Stage and the Authorities, 1700–1714 (As Revealed in the Newspapers)." *The Review of English Studies*, vol. 14, no. 53, 1938, pp. 53–62. JSTOR, www.jstor.org/stable/510202.

Jacobs, Edward. "Edward Lloyd's Sunday Newspapers and the Cultural Politics of Crime News, c. 1840–43." *Victorian Periodicals Review*, vol. 50 no. 3, 2017, pp. 619–49, doi:10.1353/vpr.2017.0043.

Janssen, Flore. "'What you write down is going to the press': Margaret Harkness's Accounts of the 1889 London Dockworkers' Strike." *Victorians: A Journal of Culture and Literature*, vol. 132, 2017, pp. 162–74, doi:10.1353/vct.2017.0015.

JGraph. *draw.io*. Software. Available from https://github.com/jgraph. Accessed 2 December 2019.

Jones, Aled. *Powers of the Press: Newspapers, Power and the Public in Nineteenth-Century England*. Ashgate, 1996.

Jump, J. D. "Weekly Reviewing in the Eighteen-Fifties." *The Review of English Studies*, vol. 24, no. 93, 1948, pp. 42–57, doi:10.1093/res/III.11.244.

Keeble, Richard and Ian Reeves (eds.). *The Newspapers Handbook*. 5th Ed. Routledge, 2014.

Kelly, Tighearnan. "Trove Through Time." National Library of Australia, 11 February 2016. Web. https://www.nla.gov.au/blogs/trove/2016/01/12/trove-through-time. Accessed 2 December 2019.

Kettunen, Kimmo, Tuula Pääkkönen, Tuula and Mika Koistinen. "Between Diachrony and Synchrony: Evaluation of Lexical Quality of a Digitized Historical Finnish Newspaper and Journal Collection with Morphological Analyzers." *Baltic HLT 2016: The 7th Conference. Human Language Technologies – the Baltic Perspective*. Riga, Latvia. 6-7 October 2016. Paper Presentation.

King, Andrew, Alexis Easley and John Morton (eds.). *The Routledge Handbook to Nineteenth-Century British Periodicals and Newspapers*. Routledge, 2016.

King, Andrew. *The London Journal, 1845–83: Periodicals, Production and Gender*. Routledge, 2017.

King, Ed. "British Newspapers 1800–1860." *British Library Newspapers*, 2007. Web. https://www.gale.com/intl/essays/ed-king-british-newspapers-1800-1860. Accessed 29 November 2019.

---. "Digitisation of British Newspapers 1800–1900." *British Library Newspapers*, 2007. Web. https://www.gale.com/intl/essays/ed-king-digitisation-of-british-newspapers-1800-1900. Accessed 29 November 2019.

---. "Digital Historical Newspapers Online: Prospects and Challenges." *Newspapers Collection Management: Printed and Digital Challenges*. Ed. Hartmut Walravens. IFLA Publications, 2008. 59–78.

King, Edmund. "Digitisation of Newspapers at the British Library." *The Serials Librarian: From the Printed Page to the Digital Age*, vol. 49, no. 1/2, 2005, pp. 165–81, doi:10.1300/J123v49n01_07.

Krabbenhoeft, Nick, Katherine Skinner, Matt Schultz and Frederick Zarndt. "Chronicles in Preservation: Preserving Digital News and Newspapers." *Preservation, Digital Technology & Culture*, vol. 42, no. 4, 2013, pp. 2013–29, doi:10.1515/pdtc-2013-0029.

Krobb, Florian. *Selbstdarstellungen: Untersuchungen zur deutsch-jüdischen Erzählliteratur im neunzehnten Jahrhundert*. Königshausen & Neumann, 2000.

Kwiat, Joseph J. "The Newspaper Experience: Crane, Norris, and Dreiser." *Nineteenth-Century Fiction*, vol. 8, no. 2, 1953, pp. 99–117, doi:10.2307/3044435.

Landon, Carol. "Wordsworth, Coleridge, and the Morning Post: An Early Version of 'The Seven Sisters'." *The Review of English Studies*, vol. 11, no. 44, 1960, pp. 392–402, doi:10.1093/res/XI.44.392.

Lanz, Daniel, Frederick Zarndt, Stefan Boddie, Tracy Powell and Vishal Salgotra. "The New Papers Past: An International Collaboration between New Zealand, India, Germany, and the United States." *OCLC Systems & Services: International Digital Library Perspectives*, vol. 25, no. 4, pp. 287–94, doi:10.1108/10650750911001860.

Lee, Bronwyn. *Use of METS and ALTO in the Australian Newspapers Digitisation Program (ANDP) at the National Library of Australia (NLA)*. National Library of Australia, 2012.

Lepidus, Henry. "The History of Mexican Journalism." *The University of Missouri Bulletin*, vol. 29, no. 4, 1928, pp. 2–87.

Library of Congress. "United States Newspaper Program." *Newspaper & Current Periodical Reading Room: Serial & Government Publications Division*. Library of Congress, 23 August 2010. Web. http://www.loc.gov/rr/news/usnp/usnpp.html. Accessed 6 December 2019.

---. "MIX: NISO Metadata for Images in XML Schema, Technical Metadata for Digital Still Images Standard Official Website." *Standards*. Library of Congress, 23 November 2015. Web. www.loc.gov/standards/mix/. Accessed 4 December 2019.

---. "ALTO: Technical Metadata for Layout and Text Objects." *Standards*. Library of Congress, 26 May 2016. Web. https://www.loc.gov/standards/alto/. Accessed 4 December 2019.

---. "MODS: Metadata Object Description Schema Official Website." *Standards*. Library of Congress, 14 September 2018. Web. https://www.loc.gov/standards/mods/. Accessed 4 December 2019.

---. *Enriching the Library Experience: The FY2019–2023 Strategic Plan of the Library of Congress*. Library of Congress, 2018. https://www.loc.gov/strategic-plan/. Accessed 29 November 2019.

---. *chronam*. Software. Available from https://github.com/LibraryOfCongress/chronam. Accessed 29 November 2019.

---. *The FY2019–2023 Digital Strategic Plan of the Library of Congress. Library of Congress*, 2019. https://www.loc.gov/static/portals/digital-strategy/documents/Library-of-Congress-Digital-Strategy-v1.1.2.pdf. Accessed 29 November 2019.

---. "About Chronicling America." *Chronicling America*. Library of Congress, 2019. Web. https://chroniclingamerica.loc.gov/about/. Accessed 29 November 2019.

---. "About the NDNP [The National Digital Newspaper Program]." *National Digital Newspaper Program*. Library of Congress, 2019. Web. www.loc.gov/ndnp/. Accessed 29 November 2019.

---. "METS: Metadata Encoding & Transmission Standard Official Website." *Standards*. Library of Congress, 17 January 2019. Web. www.loc.gov/standards/mets/. Accessed 4 December 2019.

---. "PREMIS: Preservation Metadata Maintenance Activity." *Standards*. Library of Congress, 23 July 2019. Web. www.loc.gov/standards/premis/. Accessed 4 December 2019.

---. "Guidelines & Resources." *National Digital Newspaper Program*. Library of Congress, 11 August 2019. Web. http://www.loc.gov/ndnp/guidelines/. Accessed 6 December 2019.

Liddle, Dallas. "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' *The Times* using *The Times Digital Archive*." *Journal of Victorian Culture*, vol. 17, no. 2, 2012, pp. 230–37, doi:10.1080/13555502.2012.683151.

Lieder, Hans-Jörg. *Europeana Newspapers Project and Beyond*. Europeana Newspapers, 2014. www.europeana-collections-1914-1918.eu/wp-content/uploads/2014/02/Lieder-Unlocking-Sources-2014-01-31.pdf. Accessed 29 November 2019.

Lockwood, Thomas. "Did Fielding Write for The Craftsman?" *The Review of English Studies*, vol. 59, no. 238, 2008, pp. 86–117, doi:10.1093/res/hgl158.

MacQueen, Donald S. "Developing Methods for Very-Large-Scale Searches in ProQuest Historical Newspapers Collection and Infotrac the Times Digital Archive: The Case of Two Million Versus Two Millions." *Journal of English Linguistics*, vol. 32, no. 2, 2004, pp. 124–43, doi:10.1177/0075424204265944.

Mak, Bonnie. "Archaeology of a Digitization." *Journal of the American Society for Information Science and Technology*, vol. 65, no. 8, 2014, pp. 1515–26, doi:10.1002/asi.23061.

Maurer, Oscar. "Anonymity vs. Signature in Victorian Reviewing." *Studies in English*, vol. 27, no. 1, 1948, pp. 1–27. JSTOR, www.jstor.org/stable/23309508.

Milligan, Ian. "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010." *The Canadian Historical Review*, vol. 94, no. 4, 2013, pp. 540–69, doi:10.3138/chr.694.

Moreux, Jean-Philippe. "Data Mining Historical Newspaper Metadata: Old News Teaches History." *2016 IFLA International News Media Conference*. Hamburg, Germany. 20-22 April 2016. Paper Presentation.

Murray, Ray L. "Toward a Metadata Standard for Digitized Historical Newspapers." *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. Ed. Mary Marlino, Tamara Sumner and Frank Shipman. ACM New York, 2005. 330–31. doi:1065385.1065459.

Mussell, Jim. "Teaching Nineteenth-Century Periodicals Using Digital Resources: Myths and Methods." *Victorian Periodicals Review*, vol. 45 no. 2, 2012, pp. 201–209, doi:10.1353/vpr.2012.0024.

---. "Elemental Forms: The Newspaper as Popular Genre in the Nineteenth Century." *Media History*, vol. 20, no. 1, 2014, pp. 4–20, doi:10.1080/13688804.2014.880264.

National Library of Australia. *Content Analysis for the Australian Newspapers Digitisation Program*. National Library of Australia, 2010.

---. *Specifications for OCR Services*. National Library of Australia, 2013.

---. *Guidelines for Identifying Possible Title Changes (Draft)*. National Library of Australia, 2014.

---. *Specifications for Scanning Services*. National Library of Australia, 2016.

Neudecker, Clemens and Apostolos Antonacopoulos. "Making Europe's Historical Newspapers Searchable." *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. IEEE, 2016. 405–10. doi:10.1109/DAS.2016.83.

"News Licensing." News Corp UK & Ireland Ltd., 2019. Web. https://newslicensing.co.uk/en/page/show_home_page.html. Accessed 29 November 2019.

The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771–1874). Database. National Library of Finland, 2011. http://urn.fi/urn:nbn:fi:lb-2015051201. Accessed 29 November 2019.

The Newspaper and Periodical OCR Corpus of the National Library of Finland (1875–1920). Database. National Library of Finland, 2017. https://korp.csc.fi/download/Digilib/1875_1920/. Accessed 29 November 2019.

"Newspapers." Chronicling America, 2018. Web. https://chroniclingamerica.loc.gov/newspapers.txt. Accessed 29 November 2019.

"Newspapers API – API Methods." *Data*. Europeana Pro, 2015. Web. https://pro.europeana.eu/data/newspapers-methods. Accessed 29 November 2019.

Nicholson, Bob. "'You Kick the Bucket; We Do the Rest!': Jokes and the Culture of Reprinting in the Transatlantic Press." *Journal of Victorian Culture*, vol. 17, no. 3, 2012, pp. 273–86, doi:10.1080/13555502.2012.702664.

---. "Review of Europeana Newspapers (review no. 1894)." *Reviews in History*, 2015, doi:10.14296/RiH/2014/1894.

Nowell-Smith, Simon. *International Copyright Law and the Publisher in the Reign of Queen Victoria*. Clarendon, 1968.

"Oceanic Exchanges – A Digging into Data Project." Tracing Global Information Networks in Historical Newspaper Repositories, 1840–1914. GitHub, 2017. Web. https://github.com/OceanicExchanges. Accessed 29 November 2019.

"Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914." Oceanic Exchanges, 2018. Web. https://oceanicexchanges.org. Accessed 29 November 2019.

O'Malley, Tom. "Mitchell's Newspaper Press Directory and the Late Victorian and Early Twentieth-Century Press." *Victorian Periodicals Review*, vol. 48 no. 4, 2015, pp. 591–606, doi:10.1353/vpr.2015.0057.

Onnela, Tapio. "Historiallinen sanomalehtikirjasto 1771–1860." *Agricolan tietosanomat*, vol. 2, 2001. doi:https://agricolaverkko.fi/vintti/julkaisut/tietosanomat-1998-2005/numero2-01/sanomalehtikirjasto.html.

Pääkkönen, Tuula, Jukka Kervinen, Asko Nivala, Kimmo Kettunen and Eetu Mäkelä. "Exporting Finnish Digitized Historical Newspaper Contents for Offline Use." *D-Lib Magazine*, vol. 22, no. 7/8, 2016, n.pag., doi:10.1045/july2016-paakkonen.

Padilla, Thomas, Laurie Allen, Hannah Frost, Sarah Potvin, Elizabeth Russey Roke and Stewart Varner. *Collections as Data Facets*. Collections as Data Framework, 2017–2018. doi:10.5281/zenodo.3066240.

---. *Santa Barbara Statement on Collections as Data*. Collections as Data Framework, 2018. doi:10.5281/zenodo.3066209.svg.

---. *Always Already Computational: Collections as Data*. Collections as Data Framework, 2019. doi:10.5281/zenodo.3152935.

Papers Past. Database. National Library of New Zealand, 2018. https://paperspast.natlib.govt.nz. Accessed 29 November 2019.

"Papers Past Metadata." *Open Data*. National Library of New Zealand. Web. https://natlib.govt.nz/about-us/open-data/papers-past-metadata. Accessed 29 November 2019.

"Preservation Policy." *Strategy and Policy*. National Library of New Zealand, 2004. Web. https://natlib.govt.nz/about-us/strategy-and-policy/preservation-policy. Accessed 29 November 2019.

Pollard, Graham, and Mayne Reid. "Novels in Newspapers: Some Unpublished Letters of Captain Mayne Reid." *The Review of English Studies*, vol. 18, no. 69, 1942, pp. 72–85. JSTOR, www.jstor.org/stable/509879.

Powell, Arthur. *The Law Specially Affecting Printers, Publishers, and Newspaper Proprietors*. Stevens and Sons, 1887.

"Preserve." *Australian Newspaper Plan – Annual Report 2011/2012*. National Library of Australia, 2014. Web. https://www.nla.gov.au/australian-newspaper-plan/about/preserve. Access 29 November 2019.

"ProQuest Historical Newspapers™." *Products & Services*. ProQuest, 2018. Web. https://www.proquest.com/products-services/pq-hist-news.html#resources. Accessed 29 November 2019.

Pusapti, Teja Varma. "Blowing the Cover: News of Emily Crawford's Foreign Correspondence in the 1890s Periodical Press." *Nineteenth-Century Gender Studies*, vol. 10, no. 2, 2014, n.pag., http://www.ncgsjournal.com/issue102/pusapati.htm. Accessed 27 November 2019.

Richardson, John E. *Analysing Newspapers: An Approach from Critical Discourse Analysis*. Red Globe Press, 2007.

Rosengarten, Herbert J. "Charlotte Brontë's Shirley and the *Leeds Mercury*." *Studies in English Literature, 1500–1900*, vol. 16, no. 4, 1976, pp. 591–600. JSTOR, www.jstor.org/stable/450276.

Rubery, Matthew. *The Novelty of Newspapers: Victorian Fiction After the Invention of the News*. Oxford University Press, 2009.

Sagendorf, Cathy and David Moore. "Module 33: Newspapers." *Conser Cataloguing Manual*. Library of Congress, 2006. https://www.loc.gov/aba/pcc/conser/pdf/ccm/CCM-Module-33.pdf. Accessed 27 November 2019.

Sala, George Augustus. *Things I Have Seen and People I Have Known*. 2 vols. Cassell and Company Ltd., 1894.

---. *The Life and Adventures of George Augustus Sala*. 2 vols. C. Scribner's Sons, 1895.

Sanz, Pascal. "Development of Electronic Periodicals at the Bibliotheque National de France: Digitisation of French Daily Newspapers from Mid-19th Century to 1944." *International Newspaper Librarianship for the 21st Century*. Ed. in Hartmut Walravens. Müchen: K. G. Saur, 2006. 11–17.

Schulze, Francesca. "Anwendungsprofile und Best Practice Guides." *METS/MODS*. DDBinfo für Daten. 14 November 2019. https://wiki.deutsche-digitale-bibliothek.de/display/DFD/Anwendungsprofile+und+Best+Practice+Guides#AnwendungsprofileundBestPracticeGuides-GesamtaufnahmeeinerZeitung. Accessed 20 December 2019.

Scott, Frank W. "XXI. Newspapers, 1775–1860." *The Cambridge History of English and American Literature: An Encyclopedia in Eighteen Volumes.* Vol. XVI: American. Early National Literature: Part II. Later National Literature: Part I. Eds. W. P. Trent, J. Erskine, S. P. Sherman and C. Van Doren. G. P. Putman's Sons, 1907–21. www.bartleby.com/cambridge/. Accessed 27 November 2019.

"Selection Criteria." *Database of Digitised Daily Newspapers*. KB National Library of the Netherlands. Web. https://www.kb.nl/en/organisation/research-expertise/digitization-projects-in-the-kb/databank-of-digital-daily-newspapers/selected-titles-and-selection-procedure/selection-criteria. Accessed 28 November 2019.

Shattock, Joanne and Michael Wolff. *The Victorian Periodical Press: Samplings and Soundings*. University of Toronto Press, 1982.

Shattock, Joanne. "Michael Wolff, 1927–2016: A Tribute." *Victorian Periodicals Review*, vol. 50 no. 1, 2017a, pp. 3–8, doi:10.1353/vpr.2017.0001.

Shattock, Joanne (ed.). *Journalism and the Periodical Press in Nineteenth-Century Britain*. Cambridge University Press, 2017b.

Shaw, Jane. "10 Billion Words: The British Library Newspapers 1800–1900 Project. Some Guidelines for Large-Scale Digitisation." *World Library and Information Congress: 71st IFLA General Conference and Council "Libraries – A voyage of discovery."* Oslo, Norway. 14–18 August 2005. https://archive.ifla.org/IV/ifla71/papers/154e-Shaw.pdf. Accessed 29 November 2019.

Shaw, Jane. *British Newspapers 1620–1900: Final Report*. JISC and the British Library, 2009. http://www.webarchive.org.uk/wayback/archive/20140614080134/http://www.jisc.ac.uk/media/documents/programmes/digitisation/blfinal.pdf. Accessed 28 November 2019.

Sherratt, Tim. "From Portals to Platforms: Building New Frameworks for User Engagement." *National Library of Australia*, 5 November 2013. Web. https://www.nla.gov.au/our-publications/staff-papers/from-portal-to-platform. Accessed 2 December 2019.

---. *Trove API Console*. troveconsole.herokuapp.com. Accessed 2 December 2019.

Scissors and Paste Database. Database. Ed. M. H. Beals, 2016. http://scissorsandpaste.net/. Accessed 23 March 2020.

Simons, Gary. "The Squab and the Idler: A Cosmopolitan – Colonial Dialogue in the Calcutta Star between William Thackeray and James Hume." *Victorian Literature and Culture*, vol. 42, no. 3, 2014, pp. 387–406, doi:10.1017/S1060150314000059.

Skinner, Katherine and Matt Schultz. *Guidelines for Digital Newspaper Preservation Readiness*. Educopia Institute, 2014.

Smith, David A. *passim*. Software. Available from https://github.com/dasmiq/passim. Accessed 29 November 2019.

Stamm, Michael. "The Space for News: Ether and Paper in the Business of Media." *Media History*, vol. 21, no. 1, 2015, pp. 55–73, doi:10.1080/13688804.2014.950949.

Sterling, Christopher H. *Enyclopedia of Journalism*. 6 vols. SAGE Publications, 2009.

"Strategic Plan 2018–2020." *About NSLA*. National and State Libraries Australia, 2018. Web. https://www.nsla.org.au/about/strategic-plan. Accessed 29 November 2019.

Suomen Kansalliskirjaston digitoidut sanomalehdet. Database. The National Library of Finland, 2019. https://digi.kansalliskirjasto.fi/etusivu. Accessed 29 November 2019.

Tanner, Simon, Trevor Muñoz and Pich Hemy Ros. "Measuring Mass Text Digitization Quality and Usefulness: Lessons Learned from Assessing the OCR Accuracy of the British Library's 19th Century Online Newspaper Archive." *D-Lib Magazine*, vol. 15, no. 7/8, 2009, n.pag., doi:http://www.dlib.org/dlib/july09/munoz/07munoz.html.

Tener, Robert H. "Breaking the Code of Anonymity: The Case of the Spectator, 1861–1897." *The Yearbook of English Studies*, vol. 16, 1986, pp. 63–73. JSTOR, www.jstor.org/stable/3507766.

Text Reuse in Finnish Newspapers and Journals, 1771–1920. Database. Eds. Aleksi Vesanto, Filip Ginter, Hannu Salmi, Asko Nivala, Reetta Sippola, Heli Rantala and Petri Paju, 2017. http://comhis.fi/clusters. Accessed 4 December 2019.

Thérenty, Marie-Ève. "The 'Fooliton': A French Media Invention." *Victorian Review*, vol. 38 no. 2, 2012, pp. 35–38, doi:10.1353/vcr.2012.0012.

Tilley, Elizabeth. "Periodicals in Ireland." *The Routledge Handbook to Nineteenth-Century British Periodicals and Newspapers*. Eds. Andrew King, Alexis Easley and John Morton. Routledge, 2016. 208–20.

The Times Digital Archive. Database. Gale: A Cengage Company, 2018. https://www.gale.com/intl/c/the-times-digital-archive. Accessed 29 November 2019.

"The Times Digital Archive." *Primary Sources*. Gale: A Cengage Company, 2019. Web. https://www.gale.com/intl/c/the-times-digital-archive. Accessed 29 November 2019.

"Titles." *Newspapers Home*. Trove, 2019. Web. https://trove.nla.gov.au/newspaper/rss/titles. Accessed 29 November 2019.

Todd, Chris. "Metadata Mayhem: Cataloguing Electronic Resources in the National Library of New Zealand." *The Electronic Library*, vol. 21, no. 3, pp. 214–22, 2003, doi:10.1108/02640470310480452.

Turner, Mark W. "Companions, Supplements, and the Proliferation of Print in the 1830s." *Victorian Periodicals Review*, vol. 43 no. 2, 2010, pp. 119–32, doi:10.1353/vpr.0.0116.

Trove. Database. National Library of Australia, 2018. https://trove.nla.gov.au. Accessed 29 November 2019.

Vanden Bossche, Chris R. "The Value of Literature: Representations of Print Culture in the Copyright Debate of 1837–1842." *Victorian Studies*, vol. 38, no. 1, 1994, pp. 41–68. JSTOR, www.jstor.org/stable/4618881.

Vesanto, Aleksi, Asko Nivala, Heli Rantala, Tapio Salakoski, Hannu Salmi and Filip Ginter. "Applying BLAST to Text Reuse Detection in Finnish Newspapers and Journals, 1771–1910." *Proceedings of the 21st Nordic Conference of Computational Linguistics*. Linkoping University Electronic Press, 2017. 54–58. doi:http://www.ep.liu.se/ecp/133/010/ecp17133010.pdf.

Vesanto, Aleksi, Asko Nivala, Tapio Salakoski, Hannu Salmi and Filip Ginter. "A System for Identifying and Exploring Text Repetition in Large Historical Document Corpora." *Proceedings of the 21st Nordic Conference of Computational Linguistics*. Linkoping University Electronic Press, 2017.330–33. doi:http://www.ep.liu.se/ecp/131/049/ecp17131049.pdf

Viola, Lorella and Jaap Verheul. "Mining Ethnicity: Discourse-Driven Topic Modelling of Immigrant Discourses in the USA, 1898–1920." *Digital Scholarship in the Humanities*, fqz068, 2019, doi:https://doi.org/10.1093/llc/fqz068.

Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines. Ed. Ryan Cordell and David Smith, 2017. http://viraltexts.org/. Accessed 23 March 2020.

Vizetelly, Henry. *Glances Back Through Seventy Years: Autobiographical and Other Reminiscences*. 2 vols. K. Paul, Trench, Trübner & Co., Ltd., 1893.

Walravens, Hartmut. *Newspapers in Central and Eastern Europe: Papers Presented at an IFLA Conference held in Berlin, August 2003*. De Gruyter, 2017.

Ward, Candace. "'An Engine of Immense Power': The Jamaica Watchman and Crossings in Nineteenth-Century Colonial Print Culture." *Victorian Periodicals Review*, vol. 51 no. 3, 2018, pp. 483–503, doi:10.1353/vpr.2018.0033.

The Waterloo Directory of English Newspapers and Periodicals: 1800–1900. Database. Ed. John S. North, 2009. www.victorianperiodicals.com/series3/index.asp. Accessed 28 November 2019.

The Wellesley Index to Victorian Periodicals, 1824–1900. Database. ProQuest, 2006–2019. wellesley.chadwyck.co.uk. Accessed 28 November 2019.

Westmancoat, John. *Newspapers*. British Library Publishing, 1985.

"What is a Newspaper?" *Resources*. National Library of Australia. Web. https://www.nla.gov.au/content/what-is-a-newspaper. Accessed 28 November 2019.

"Wikipedia:List of online newspaper archives." Wikipedia, 16 December 2019. Web. https://en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives. Accessed 20 December 2019.

Willems, Marieke and Rossitza Atanassova. "Europeana Newspapers: Searching Digitized Historical Newspapers from 23 European Countries." *Insights*, vo. 28, no. 1, 2015, pp. 51–56, doi:10.1629/uksg.218.

Williams, Kevin. *Read All About It!: A History of the British Newspaper*. Routledge, 2009.

Woodruff, James P. and H. B. de Groot. "'This Day Is Published': Advertisements in Periodicals as a Means of Dating Eighteenth and Nineteenth Century Publications." *Victorian Periodicals Newsletter*, vol. 11, no. 3, 1978, pp. 104–108. JSTOR, www.jstor.org/stable/20085195.

World Intellectual Property Organisation. *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)*. Geneva, 1978.

Wu, Duncan. "The Journalism of William Hazlitt (1737–1820) in Boston (1784–5): A Critical and Bibliographical Survey." *The Review of English Studies*, vol. 57, no. 229, 2006, pp. 221–46. JSTOR, www.jstor.org/stable/4095473.

Yates, Edmund. *Edmund Yates: His Recollections and Experiences*. 2 vols. R. Bentley & Son, 1884.

ZEFYS. Database. SBB, 2011–2017. http://zefys.staatsbibliothek-berlin.de. Accessed 20 December 2019.

ZEFYS. "Liste der digitalisierten Zeitungen." ZEFYS, 18 May 2017. Web. zefys.staatsbibliothek-berlin.de/index.php?id=list. Accessed 20 December 2019.

"ZEFYS." *SBB-Lab*. Staatsbibliothek zu Berlin. Web. https://lab.sbb.berlin/5393/?lang=en. Accessed 20 December 2019.

Index

M

N

O

P

Q

R

Between 2017 and 2019, Oceanic Exchanges (www.oceanicexchanges.org) brought together leading efforts in computational periodicals research from six countries—Finland, Germany, Mexico, the Netherlands, the United Kingdom, and the United States—to examine patterns of information flow across national and linguistic boundaries. *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges* brings together, in one place, the histories and structures of ten digitised newspapers databases from around the world.

The maps of this *Atlas* explore the metadata of these collections in detail, providing a selection of language variants, the technical definition employed in the categorisation process, and notes on its usage across collections and in the wider world of press history. With contributions from leading academics and key digitisation partners, it is the first report of its kind, and lays the foundations for cross-database analysis and a deeper understanding of digitised newspaper archives' creation and structure. A dynamic version of the report and dataset, which is open to new contributions, can be accessed at www.digitisednewspapers.net.

Citation: Beals, M. H. and Emily Bell, with contributions by Ryan Cordell, Paul Fyfe, Isabel Galina Russell, Tessa Hauswedell, Clemens Neudecker, Julianne Nyhan, Mila Oiva, Sebastian Padó, Miriam Peña Pimentel, Lara Rose, Hannu Salmi, Melissa Terras, and Lorella Viola. *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges*. Loughborough: 2020. DOI:10.6084/m9.figshare.11560059.

Contact
School of Social Sciences and Humanities
Loughborough University
Leicestershire
LE11 3TU, UK

Dr M. H. Beals
E: m.h.beals@lboro.ac.uk
www.digitisednewspapers.net

@lborouniversity
/lborouniversity
/lborouniversity
/lborossh

77731/C&PS/April 2020