

# Global News Index: Quick Reference Tables

## Variables Summary Table

Variable Name	Data Type	Description
<b>Metadata &amp; Content Variables</b>		
aid	String	Unique article identification number
publication_date	Timestamp	Date article was published
ingest_date	Timestamp	Date Cline Center processed the article
source_name	String	Cline Center corpus the article belongs to
url	String	The URL of the article
title	String	† Title of the article
title_length	Numeric	A character count of the article title
* content	String	† The normalized text of the article
content_length	Numeric	A character count of the article content
word_count	String	The number of words in an article
other_metadata	String	Any extra metadata provided with the article
publisher	String	The article publisher
<b>Named Entity Recognition Variables</b>		
extracted_people	§ String	† Personal names extracted from article content
* extracted_people_text	§ String	Untokenized version of 'extracted_people'; only returns exact matches
extracted_locations	§ String	† Named locations extracted from article content
* extracted_locations_text	§ String	Untokenized version of 'extracted_locations'; only returns exact matches
extracted_organizations	§ String	† Organizational names extracted from article content
* extracted_organizations_text	§ String	Untokenized version of 'extracted_organizations'; only returns exact matches
country	§ String	Nation-state names extracted from article content
<b>* hidden</b>	<b>§ multivalued</b>	<b>† tokenized</b>
This variable can be queried, but will not be displayed in query results.	This variable is returned as a list.	This variable is tokenized, so you can search for individual words/phrases within the field.
		<b>‡ comma-tokenized</b>
		This field is tokenized by comma, so you can search for individual place names separated by comma.

Geolocation Variables		
geolocation_original	§ String	‡ Corresponds to 'extracted_locations', repeated for each place estimated by the algorithm
geolocation_locations	§ String	‡ The detailed placename for each estimated geolocation
* geolocation_locations_text	§ String	Untokenized version of 'geolocation_locations'; only returns exact matches
geolocation	§ String	Latitude and longitude for each estimated location
geolocation_featureids	§ Numeric	Unique geographic identifier for each estimated location
geolocation_probabilities	§ Numeric	A score for each estimate that ranges from 0 (lowest score) to 1 (highest score).
Sentiment Variables		
anew_arousal	Numeric	ANEW Dictionary - Arousal
anew_dominance	Numeric	ANEW Dictionary - Dominance
anew_valence	Numeric	ANEW Dictionary - Valence
bing_liu_neg	Numeric	Bing Liu Negative Sentiment
bing_liu_pos	Numeric	Bing Liu Positive Sentiment
dal_activation	Numeric	DAL - Activation
dal_imagery	Numeric	DAL - Imagery
dal_pleasantness	Numeric	DAL - Pleasantness
inquirer_neg	Numeric	General Inquirer – Negative Sentiment
inquirer_pos	Numeric	General Inquirer – Positive Sentiment
lexicoder_neg	Numeric	Lexicoder-based Negative Sentiment
lexicoder_pos	Numeric	Lexicoder-based Positive Sentiment
mf_moralitygeneral	Numeric	Enhanced Moral Foundations – Morality General
mf_authorityvice	Numeric	Enhanced Moral Foundations – Authority Vice
mf_authorityvirtue	Numeric	Enhanced Moral Foundations – Authority Virtue
mf_fairnessvice	Numeric	Enhanced Moral Foundations – Fairness Vice
mf_fairnessvirtue	Numeric	Enhanced Moral Foundations – Fairness Virtue
mf_harmvice	Numeric	Enhanced Moral Foundations – Harm Vice
mf_harmvirtue	Numeric	Enhanced Moral Foundations – Harm Virtue
mf_ingroupvice	Numeric	Enhanced Moral Foundations – In-Group Vice
mf_ingroupvirtue	Numeric	Enhanced Moral Foundations – In-Group Virtue
mf_purityvice	Numeric	Enhanced Moral Foundations – Purity Vice
mf_purityvirtue	Numeric	Enhanced Moral Foundations – Purity Virtue
pronouns	Numeric	Count of the collective in-group pronouns
Classifier Variables		
status	String	Tracks whether the article is relevant for civil conflict analysis or is in a foreign language
* <b>hidden</b>	§ <b>multivalued</b>	† <b>tokenized</b>
This variable can be queried, but will not be displayed in query results.	This variable is returned as a list.	‡ <b>comma-tokenized</b> This field is tokenized by comma, so you can search for individual place names separated by comma.

## Corpora Summary Table

Source	Article	Time Span	Provider
<b>NYT Corpus</b>			
New York Times (NYT)	5,711,504	01/01/1945 - 01/31/2006	ProQuest
<b>WSJ Corpus</b>			
Wall Street Journal (WSJ)	2,246,861	01/02/1945 - 01/31/2006	ProQuest
<b>SWB Corpus</b>			
BBC Monitoring: Summary of World Broadcasts (SWB)	4,730,860	01/01/1979 - 06/14/2015	LexisNexis
<b>FBIS Corpus</b>			
Foreign Broadcast Information Service (FBIS)	3,275,879	01/13/1995 - 08/18/2004	UIUC Library
<b>LexisNexis Corpus</b>			
Agence France Presse (AFP)	310,537	02/01/2011 - 11/30/2012	LexisNexis
Associated Press Online (APO)	272,153	02/01/2011 - 11/30/2012	LexisNexis
The Associated Press State & Local Wire (APSLW)	957,566	02/01/2011 - 11/30/2012	LexisNexis
Associated Press Worldstream (APW)	67,847	02/01/2011 - 11/30/2012	LexisNexis
Deutsche Presse-Agentur (DPA)	97,282	02/01/2011 - 11/30/2012	LexisNexis
Interfax News Agency (INA)	488,384	02/01/2011 - 11/30/2012	LexisNexis
The Associated Press (AP)	380,564	02/01/2011 - 11/30/2012	LexisNexis
United Press International (UPI)	97,088	02/01/2011 - 11/30/2012	LexisNexis
Xinhua General News Service (XGNS)	439,008	02/01/2011 - 11/30/2012	LexisNexis
<b>BulkLexisNexis Corpus</b>			
Lexis Nexis Web Services Kit (WSK) and Lexis Nexis Bulk Content API (BLN) <sup>†</sup>			
The Associated Press (AP)	8,060,837	01/01/1977 - Present	BLN
Agence France Presse (AFP)	4,264,932	05/05/1991 - 08/18/2019	WSK
The Associated Press Election Results	26,521	11/08/1994 - Present	BLN
Associated Press Financial Wire	1,015,820	02/17/2005 - Present	BLN
Associated Press International	3,289,451	10/01/1993 - Present	BLN
Associated Press News Briefs	30	Recent 2 weeks	BLN
Associated Press Online	2,686,743	07/08/1997 - Present	BLN
The Associated Press State & Local Wire (same day)	9,551,955	07/21/1998 - Present	BLN
BBC Monitoring: International Reports	3,796,367	01/01/1979 - Present	BLN
Deutsche Presse-Agentur (DPA)	1,528,698	07/03/1994 - 08/18/2019	WSK
<sup>†</sup> Article counts for BLN sources are subject to change			

Inter Press Service	167,766	04/02/1984 - Present	BLN
Financial Times (London)	3,185,414	01/02/1982 - Present	BLN
New York Times (NYT)	4,221,531	06/01/1980 – 06/27/2022	BLN
The New York Times Blogs	160,393	09/07/1998 - Present	BLN
The Washington Daybook	442,550	01/03/2000 - Present	BLN
The Washington Post (WP)	3,563,889	01/01/1977 - 08/18/2019	WSK
United Press International (UPI)	2,666,693	09/26/1980 – Present	BLN
Xinhua General News Service (XGNS)	4,198,518	01/01/1977 – Present	BLN
Xinhua Economic News Service	881,555	10/08/1999 - Present	BLN
<b>Web Corpus<sup>†</sup></b>			
Web Articles (Web)	> 110,000,000 <sup>^</sup>	02/05/2006 - Present*	Voyager WebCrawler

<sup>†</sup> Article counts are subject to change

<sup>^</sup> This article count is lower than what may appear on Archer. This is due to the fact that there are over 26,000,000 'empty' articles currently in the Web Corpus. Empty articles are articles that contain no content and have a status field = 'empty'. When we calculate the true number of news articles that are in the Web Corpus, we take into consideration the 'empty' articles since these articles may be of no use.

\* Corpus is updated daily; Gap in coverage from October 2011 - June 2012