

Statement of Requirements

Global News Media Archive Department of Defense Analysis

Naval Postgraduate School

1.0 Background/Introduction:

This project is part of a research effort sponsored by Marine Forces Cyber Command, entitled “Media Analytics for Enhanced Indications and Warnings.” Our research project seeks to examine differences in the language and tone of news stories, as they are reported in different countries, so as to generate enhanced predictive analytics for assessing future hotspots of conflict. This research includes a requirement for a global historical data archive, providing the full text of news stories, as they were reported across a wide variety of sources and languages. This data will also be used for pedagogical purposes in the classroom, giving students new opportunities for thesis research, ranging from the development of “big data” analytic skills, to the development of substantive analytic products offering new metrics for the assessment of military conflict through global media. The fundamental goal of this project is to further Naval and Joint research and educational goals, by providing enhanced tools and capabilities for the Information Warfare community.

2.0 Minimal Technical Requirements/Salient Characteristics:

- Minimum Time Period: 1/1/2013 - 12/31/2014
- Minimum Number of Total Records: 600,000,000
 - The minimum number of records shall be from the minimum time period. Additional records outside of the minimum time period are allowable.
- Minimum Number of Original Sources: 50,000
 - No minimum number of records from any single source.
- Minimum Number of Source Countries: 100
 - Minimum 1,000 records per each country.
- Minimum Number of Source Languages: 40
 - Minimum 1,000 records per each language.
- Each record in the archive should contain a news story gathered from a news website, and should include, at a minimum, the original source URL, date, country of origin, and full text of the original news story in its original language. Digital transcriptions of physical news media or broadcast news media are also allowable and should include, at a minimum, the original source information, date, country of origin, and full text of the original news story in its original language. There is no requirement regarding the type of topic of the news stories.
- The Contractor shall provide a protected URL, where they have posted the dataset online, to allow the Government to download and store the data on its own machines. The preferred format of data is JSON; data formats of XML, CSV, or similar are acceptable.

NPS Technical POC: To be provided upon award.

NPS POC for Acceptance of Deliverables: To be provided upon award.

NPS Financial POC: To be provided upon award.